

# 監視システムのための画像処理による 移動人物の検出と認識に関する研究

中島 慶人

博士(情報学)

総合研究大学院大学  
複合科学研究科  
情報学専攻

平成 18 年度  
(2006)

2006 年 9 月

本論文は総合研究大学院大学複合科学研究科情報学専攻に  
博士(情報学)授与の要件として提出した博士論文である。

#### 審査委員会

|    |          |                    |
|----|----------|--------------------|
| 主査 | 上野晴樹教授   | 国立情報学研究所/総合研究大学院大学 |
|    | 佐藤真一教授   | 国立情報学研究所/総合研究大学院大学 |
|    | 古山宣洋助教授  | 国立情報学研究所/総合研究大学院大学 |
|    | 後藤田洋伸助教授 | 国立情報学研究所/総合研究大学院大学 |
|    | 小野欽司教授   | 国立情報学研究所           |
|    | 白井良明教授   | 立命館大学              |

**A Study on Moving Persons Detection and Recognition  
by Image Processing for Surveillance System**

Chikahito Nakajima

DOCTOR OF  
PHILOSOPHY

Department of Informatics,  
School of Multidisciplinary Sciences,  
The Graduate University for Advanced Studies (SOKENDAI)

September 2006

A dissertation submitted to the Department of Informatics,  
School of Multidisciplinary Sciences,  
The Graduate University for Advanced Studies (SOKENDAI)  
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Advisory Committee

|         |                               |   |
|---------|-------------------------------|---|
| Advisor | Prof. Haruki Ueno             | National Institute<br>of Informatics (NII),<br>SOKENDAI |
|         | Prof. Satoh Shin'ichi         | NII, SOKENDAI   |
|         | Assoc.Prof. Nobuhiro Furuyama | NII, SOKENDAI   |
|         | Assoc. Prof. Hironobu Gotoda  | NII, SOKENDAI   |
|         | Prof. Kinji Ono               | NII   |
|         | Prof. Yoshiaki Shirai         | Ritumeikan University                                   |

# 内容梗概

電力設備や石油プラントなどで設備の稼動状態の把握や、保守保全および警備の観点から、監視カメラを用いた画像監視システムが数多く導入されている。さらに近年、ネットワーク技術の進展や監視カメラの低コスト化などが背景となり、監視カメラの設置台数が増える傾向にある。これらの多くの監視カメラ画像を監視員が見続けることが不可能なため、目視監視を補助あるいは代行する画像処理が求められている。特に、監視員が目視で行っている以下の機能を持つ画像処理の開発が望まれている。

- ・ 監視領域に入る人物を選択的に実時間で検出する機能
- ・ 検出人物の動作に応じて警報を出すための姿勢の認識機能
- ・ 検出人物の個別認識機能

電気事業で利用している画像監視システムに対し、最も要望の高いニーズが侵入監視の自動化である。また、監視対象になる頻度が高いがその全身画像による認識はほとんど着手されていないという技術的な理由で、人物の検出と認識を本論文の対象とし、上記3機能の実現方式を目指す。特に、人物の検出と認識は、その形状などが不安定でパラメータ設定が困難なため、機械学習が不可欠である。この機械学習として2分類問題用の識別器であるサポートベクターマシン（SVM）を使い、監視領域に入る人物を検出し認識する画像処理を考案する。ただし、機械学習を適用しただけでは実時間処理や安定した認識ができない。そこで本論文では、人の視覚野の動き処理と認識処理の関係を参考に、動きによる人物領域の絞り込みを先に行い、その後SVMによる人物の検出・認識を行う段階的な処理で、人物の実時間検出と認識を可能にする方式を示す。

第1に、SVMを使い監視画像から人物を検出する方式は既に提案されているが、従来のSVMによる人物検出方式は、画像内から人物を選択的に検出できるものの、実時間(30画像/秒)で処理ができなという弱点がある。そこで本論文では、SVMの人物検出方式を使い実時間で人物を検出するための新しい前処理手法として、固定カメラ用にライン型前処理を、旋回カメラ用に側抑制型前処理を提案する。ライン型前処理は、監視領域内に監視ラインを設定し、監視ライン上で動きの速い部分に優先順位をつけ、優先順位の上位を人物検出の探査範囲とする方式である。一方、側抑制型前処理は、移動中の人物が周囲の背景と異なる動きをすることに着目し、周囲と異なる動き部分を人物検出の探査範囲とする方式である。評価実験では、考案した前処理と従来の機械学習による人物検出方式を使い、実時間で人物検出が可能であることを示す。

第2に、検出した人物の動作に応じて警報を出すには、人物の姿勢を認識する機能が必要となる。全身画像に身体モデルを当てはめる研究成果が多く報告されているが、人物の様々な姿勢を定義することが困難なため、人物検出と同様

な枠組みで姿勢認識を行う画像監視システムは存在していない。そこで本論文では、代表的な姿勢が定義できる発電所運転時の運転員を撮影する監視カメラを用い、SVMによる姿勢認識を研究する。運転員の動きと運転姿勢の間には相関があるので、個々の運転姿勢を個別に学習したSVMを運転員の動きに合わせて動的に切替る動的切替方式を前処理として提案する。評価実験では、典型的な4種類の姿勢を学習することで、監視カメラで撮影した1時間30分の実験画像に対し、運転員の向きや姿勢のばらつきに影響されることなく95%の安定した姿勢認識率が得られたことを示す。なお、本成果により画像内の作業員の立ち位置と運転姿勢が分かるため、標準的な運転操作を逸脱した時などに警報を出す画像監視システムの作成が可能となる。

第3に、上述方式で検出した全身画像から人物を特定する認識機能が重要である。しかし、人物の服装や形状が常に変化するために安定した認識特徴が得られないことから、人物認識に全身画像は利用されていない。そこで、本論文では研究室内に出入する既知人物8人の16日間の全身画像を使い、SVMによる人物認識を研究した。8人を認識する必要からSVMを段階的に切替る多段戦略と、画像から得られる4種類の画像特徴を用いた認識実験結果を示す。実験の結果、監視カメラで一度撮影した同一日であれば約88~98%の割合で個人を特定できたことを示す。また、100物体が写る7,200画像による定量的な認識検証実験を行い、同様の認識率が得られたことを示す。SVMの多クラス化では、1対他方式と2クラス対方式を比較し、識別人数や学習画像が豊富な場合は学習時間がかからず追加学習が容易な2クラス対方式による多段戦略が実務面で有利であることを示す。また、SVMの多段戦略よりも、利用する画像特徴に認識精度が依存することを示す。本実験結果は限定的なものだが、監視カメラで撮影した全身画像から、一度撮影した人物を個別に検出する機能の開発に貢献する成果である。

以上、本論文では監視カメラに写る人物の全身画像を使い、第1に人物の実時間検出、第2に検出した人物の姿勢認識、第3に検出した人物の認識を実現する方式を提案する。提案方式は、視覚野の処理を参考に、はじめに画像内の動きで処理範囲を絞り込み、次に機械学習で検出/認識を行う手順となっている。学習画像を変更することで人物以外の移動体にも拡張が容易な方式である。さらに、実用の観点から特殊な装置を使うことなく市販の計算機と監視カメラで実現できる方式を提案している。

本論文は、監視員の目視を補助あるいは一部を代行する高機能な画像監視システムの実現に寄与する具体的な研究成果である。

# ABSTRACT

Many surveillance cameras are used in facilities that must be kept secure such as electric power plants or oil plants, for security, maintenance, and observing machine conditions. The number of surveillance cameras in use has recently increased in proportion to an increasing number of crimes. Human operators are now physically unable to constantly monitor these cameras. Advanced automated surveillance systems that can augment or replace human operators must therefore be developed. Such systems need to be able to

- detect target objects such as human intruders in video-rates (30 frames per second : fps),
- recognize the posture of detected humans, and
- identify the intruders.

We focused on humans entering the surveillance area as targets. Our goal was to develop methods for intruder detection, posture recognition, and person identification. We used machine-learning techniques such as Support Vector Machine (SVM) because empirical decision rules cannot be devised for these tasks. However, real-time detection and high recognition rates cannot be achieved by using only SVM. To propose another suitable method, we took our inspiration from the human visual system. In the human brain, the retinas and the inside of the visual cortex are connected by two routes. One route is for object recognition and the other, whose processing speed is twice as fast, is for motion analysis. We based our method on this two-route structure and connections. First, using motion analysis, our method detects candidate areas for moving targets. Then, using SVM object recognition, it classifies the candidates into target objects and non-target objects.

Machine-learning techniques such as SVM are used widely for intruder detection, but existing methods have processing speeds of only 5 fps and cannot detect intruders in real-time, which is necessary for higher accuracy. To develop a high-speed method, we designed two preprocessing schemes to the SVM-based intruder detection. Each corresponds to the motion analysis route in the human visual system: a line type preprocessing for stationary cameras and a side-control-type preprocessing for rotating cameras. The line-type preprocessing does motion detection on monitoring lines and gives priority to movements on the lines. The side-control-type preprocessing, on the other hand, detects areas moving relative to the background and gives priority to those whose moving speed and direction differ from their immediate surroundings. Our experiments showed that our methods made it possible to detect intruders in real-time from stationary and rotating cameras.

So far, posture recognition has not been achieved using SVM-based methods. This is because the large number of possible human postures makes it difficult to define the posture classes necessary to train an SVM. To solve this problem, we used typical operator's postures that have been defined in power plant operations. For the posture recognition task we trained SVMs with four typical postures and developed a strategy to select one SVM based on the detected motion of the operator. Motion is a good indicator of posture because motion and posture are highly correlated. Using this method, the posture recognition rate in our 90-minute experiment was 95 percent. In an advanced surveillance system, this method can be used, for example, to recognize potentially dangerous or inappropriate movements of operators.

In an advanced surveillance system, detected people should also be identified. Though many image-processing techniques have been proposed, many challenging tasks remain because people change their physical appearance and clothing. Getting steady features from whole body images is therefore difficult. To address this problem, we prepared about 1,000 images of eight people taken over sixteen days in our laboratory. We conducted the person recognition training using four different image characteristics. Since SVMs are two-class problem machines, we compared multi-class SVM strategies: the one-vs-all approach and pairwise approaches. From experimental results, we describe that the pairwise approaches are effective for fast training and easy system improvement when a lot of recognition classes and a lot of training images are obtained. For same-day training pictures, the recognition rates in our experiment were between 88 and 98 percent. When we increased the number of objects to 100 and the number of images to 7,200, the method still performed at the same recognition rates. These high recognition rates indicate the suitability of the method for more sophisticated surveillance applications.

This thesis describes new methods inspired by the human visual system for real-time intruder detection, posture recognition, and person identification by combining motion analysis and machine learning. These methods apply not only to human targets but can be used for surveillance of other objects such as cars, bicycles, or animals.

## 謝 辞

東京電機大学在学中に学部と大学院の指導教官としてご指導いただき、その後も 18 年間の(財)電力中央研究所での研究生活を見守りいただくと共に、さらに総合研究大学院大学において再度ご指導いただきました上野晴樹教授のご支援なくして、本博士論文はできあがらなかったものであり、心より感謝いたします。

総合研究大学院大学での研究にあたり、画像処理研究の進め方のご相談に伺うたびに、適切なお助言とご指導をいただきました佐藤真一教授に感謝いたします。ご相談させていただいたあとは、気持ちが晴れて不思議とやる気が沸いてくるのをいつもながら感じておりました。画像処理の投稿論文作成にあたり、比較する既存処理を決めかねているときに、適切な比較対象と実験方法をご指導いただきました白井良明教授に感謝いたします。認知科学の観点から、錯視などの工学的な活用方法をご教授いただくと共に、研究の方向性に関してご指導いただきました古山宣洋助教授に感謝いたします。中間発表等でのご指導や平素より暖かい励ましの言葉をいただきました後藤田洋伸助教授に感謝いたします。本論文の作成にご指導いただきました小野欽司教授に感謝いたします。研究室ゼミなどにおいて、機械学習の最新情報等をお教えいただきました松井知子助教授に感謝いたします。人と機械の相互適応の重要性をお教えいただきました山田誠二教授に感謝いたします。

総合研究大学院大学上野研究室のメンバー、Ampornaramveth Vuthichai 博士、張涛博士、Pattara Kiatisevi 博士、MD Hasanuzzaman 博士、Alexander Kovacs 氏、テイラー三代子様には研究や事務だけではなくメンタル面からもサポートしていただいたことに感謝いたします。

総合研究大学院大学での研究にご理解をいただきました(財)電力中央研究所システム技術研究所の谷口治人所長、栗原雅幸前副所長（現在、電力技術研究所副所長）、鈴木正上席研究員に感謝いたします。(財)電力中央研究所での現在の直属の上司であるセクターリーダー松井正一上席研究員には、投稿論文の所内査読をはじめ本論文の作成などで日常業務に多大なる便宜を図っていただきましたことに深く感謝いたします。画像監視システムの研究遂行にあたり研究指導していただきました(財)電力中央研究所の篠原靖志上席研究員、二方厚志上席研究員、小野田崇上席研究員に感謝いたします。共著で学術論文を作成するなど、これまでの研究遂行にともに汗を流していただきました(財)電力中央研究所の堤富士雄上席研究員、伊藤憲彦主任研究員に感謝いたします。永年共に画像処理の研究を続け、様々な作業や情報交換を行っていただいている(財)電力中央研究所の石野隆一主任研究員、村田博士主任研究員に感謝いたします。(財)電力中央研究所で同じグループに所属する佐賀井重雄上席研究員、千田恭子主

任研究員には、他大学での社会人学生としての貴重な経験を日頃よりお教えいただき、メンタル面からご支援いただいたことに感謝いたします。

画像監視システムの最新状況や監視員の要望収集に、(社)電気協同研究会主催による国内主要電力会社 11 社ならびにメーカー 10 社との「電力用画像監視システム設計技術専門委員会」(平成 14 年 10 月～平成 16 年 9 月)の 21 回の作業会が役立っている。その作業会で活発に議論をさせていただいた参加者全員に感謝するとともに、幹事役として密な意見交換をしていただきました九州電力株式会社の山岡林蔵氏、今村弘氏、野村晃彦氏、関西電力株式会社の瀬戸口正志氏、東京電力株式会社の川角浩亮氏に感謝いたします。本研究の背景になっている監視システムの実務上の課題と市場のニーズを常日頃からお教えいただいている画像監視システム作成・販売・設置会社のユー・エス・イー株式会社 加藤真良代表取締役とケイアンドケイキャピタル株式会社 梶浦一高代表取締役(兼、船井キャピタル株式会社投資管理部長)に感謝いたします。

第 3 章の研究にあたり、監視カメラの画像提供や監視カメラの被験者等を快く引き受けていただくとともに活発に意見交換をさせていただきました九州電力株式会社総合研究所の松尾健一氏、猪毛尾壮一郎氏、福田正巳氏、中部電力株式会社電力技術研究所の宮澤久永氏、藤野雅彦氏、説田武文氏に感謝いたします。

第 4 章の研究にあたり、発電所の運転方法の解説や監視画像をご提供いただきました株式会社 BWR 運転訓練センターのインストラクター鈴木弘一氏に感謝いたします。運転姿勢の定義などをお教えいただきました(財)電力中央研究所社会経済研究所ヒューマンファクター研究センターの佐相邦英上席研究員に感謝いたします。

第 5 章の研究にあたり、機械学習が視覚機能の解明に中心的な働きを果たすこと、ならびに画像処理の工学応用に直接役立つことをご教授いただくとともに、人物認識に関して議論していただきましたマサチューセッツ工科大学の Tomaso Poggio 教授に感謝いたします。研究遂行にあたり、ともに画像データを解析していただいた現在 University College London の Massimiliano Pontil 博士、Honda Research Institute の Bernd Heisele 博士、フランスのビジネススクール INSEAD の Theodoros Evgeniou 助教授に感謝いたします。

最後に、これまでの研究と本論文の作成は、いずれも家族の支援の上で成し遂げた成果であり、父・賢太郎、母・久子、義父・岡部剛一、義母・岡部智恵子、息子・史翔、何よりも妻・洋江の多大なる援助にこの場を借りてお礼を述べたい。

# 目次

|                             |           |
|-----------------------------|-----------|
| 内容梗概                        | i         |
| Abstract                    | iii       |
| 謝辞                          | v         |
| 目次                          | vii       |
| 図目次                         | x         |
| 表目次                         | xiii      |
| <br>                        |           |
| <b>第1章 序論</b>               | <b>1</b>  |
| 1. 1 はじめに                   | 1         |
| 1. 2 画像監視システム               | 2         |
| 1. 2. 1 画像監視システムの導入先        | 2         |
| 1. 2. 2 画像監視システムの構成と機能      | 4         |
| 1. 2. 3 監視カメラ               | 4         |
| 1. 2. 4 各種センサと画像送信          | 6         |
| 1. 2. 5 画像蓄積と画像監視           | 7         |
| 1. 3 画像処理の歴史的観点からの考察        | 8         |
| 1. 3. 1 画像処理の技術史            | 8         |
| 1. 3. 2 視覚のシーン理解            | 11        |
| 1. 4 要求される画像処理の概要           | 12        |
| 1. 4. 1 変化領域の検出             | 13        |
| 1. 4. 2 姿勢認識と人物認識           | 15        |
| 1. 4. 3 機械学習を用いた画像監視の処理時間   | 16        |
| 1. 5 研究の位置付けとアプローチ          | 17        |
| 1. 5. 1 社会的な要求と技術的な課題       | 17        |
| 1. 5. 2 参考とする視覚処理           | 18        |
| 1. 5. 3 研究の位置付け             | 19        |
| 1. 5. 4 人物の高速な検出            | 20        |
| 1. 5. 5 姿勢の認識               | 20        |
| 1. 5. 6 人物の認識               | 21        |
| 1. 6 論文の構成                  | 21        |
| 1. 7 本章のまとめ                 | 22        |
| <br>                        |           |
| <b>第2章 画像処理による侵入監視の関連研究</b> | <b>23</b> |
| 2. 1 はじめに                   | 23        |
| 2. 2 移動中の人物検出               | 24        |
| 2. 2. 1 画像差分による移動中の人物検出     | 24        |
| 2. 2. 2 動きや色などによる人物検出       | 27        |
| 2. 2. 3 パターン認識による人物検出       | 38        |
| 2. 3 人物の認識と姿勢の認識            | 41        |

|                                   |                       |     |
|-----------------------------------|-----------------------|-----|
| 2. 3. 1                           | 人物の認識                 | 41  |
| 2. 3. 2                           | 姿勢の認識                 | 42  |
| 2. 4                              | 画像監視の処理時間             | 44  |
| 2. 5                              | 検出と認識に利用する機械学習        | 45  |
| 2. 5. 1                           | ベイズ識別                 | 45  |
| 2. 5. 2                           | k-最近傍決定則              | 49  |
| 2. 5. 3                           | サポートベクターマシンの概要        | 50  |
| 2. 5. 4                           | SVMによる多クラス分類          | 54  |
| 2. 5. 5                           | AdaBoostの概要           | 56  |
| 2. 6                              | 視覚処理                  | 58  |
| 2. 6. 1                           | 視覚野モデル                | 58  |
| 2. 6. 2                           | 動き処理経路でのMT野の機能        | 59  |
| 2. 6. 3                           | 認識処理経路でのV4・IT野の機能     | 60  |
| 2. 6. 4                           | 画像処理との関連              | 60  |
| 2. 7                              | 本章のまとめ                | 63  |
| <b>第3章 固定カメラと旋回カメラによる人物検出の高速化</b> |                       | 64  |
| 3. 1                              | はじめに                  | 64  |
| 3. 2                              | 固定カメラによる人物検出の高速化      | 65  |
| 3. 2. 1                           | 監視環境と監視対象             | 65  |
| 3. 2. 2                           | ライン型前処理による歩行者の高速な検出処理 | 65  |
| 3. 2. 3                           | 実験結果                  | 70  |
| 3. 2. 4                           | 考察                    | 78  |
| 3. 2. 5                           | 本節のまとめ                | 79  |
| 3. 3                              | 旋回カメラによる人物検出の高速化      | 80  |
| 3. 3. 1                           | 監視環境と監視対象             | 80  |
| 3. 3. 2                           | 既存検出方式の枠組みと既存の前処理     | 81  |
| 3. 3. 3                           | 側抑制型前処理による歩行者の高速な検出処理 | 85  |
| 3. 3. 4                           | 実験結果                  | 90  |
| 3. 3. 5                           | 考察                    | 98  |
| 3. 3. 6                           | 本節のまとめ                | 101 |
| 3. 4                              | 本章のまとめ                | 102 |
| <b>第4章 作業中の人物の姿勢認識</b>            |                       | 103 |
| 4. 1                              | はじめに                  | 103 |
| 4. 2                              | 監視環境と監視対象             | 104 |
| 4. 3                              | 姿勢認識システムの概要           | 106 |
| 4. 3. 1                           | 人物領域検出                | 106 |
| 4. 3. 2                           | 状態判定                  | 107 |
| 4. 3. 3                           | 動的な切替による姿勢認識          | 109 |
| 4. 4                              | 認識実験                  | 111 |
| 4. 4. 1                           | 3姿勢の認識ルールの検証          | 111 |
| 4. 4. 2                           | 動線検出と姿勢認識実験           | 113 |

|                         |                  |     |
|-------------------------|------------------|-----|
| 4. 4. 3                 | 2万枚の画像による姿勢認識    | 114 |
| 4. 5                    | 考察               | 117 |
| 4. 5. 1                 | 下半身画像による姿勢認識     | 117 |
| 4. 5. 2                 | 提案方式の限界と精度向上方法   | 118 |
| 4. 6                    | 本章のまとめ           | 119 |
| <b>第5章 複数の既知人物の個人特定</b> |                  | 121 |
| 5. 1                    | はじめに             | 121 |
| 5. 2                    | 監視環境と監視対象        | 121 |
| 5. 3                    | 認識システムの概要        | 122 |
| 5. 3. 1                 | 前処理              | 122 |
| 5. 3. 2                 | SVMの多段戦略         | 125 |
| 5. 4                    | 既知人物の認識と向きへの認識実験 | 126 |
| 5. 4. 1                 | 実験1：人物の認識と向きへの認識 | 127 |
| 5. 4. 2                 | 実験2：16日間の人物認識    | 129 |
| 5. 5                    | 考察               | 132 |
| 5. 5. 1                 | SVMの多段戦略による認識    | 132 |
| 5. 5. 2                 | 物体の認識特性          | 132 |
| 5. 5. 3                 | 向きへの認識特性         | 134 |
| 5. 5. 4                 | 提案方式の限界と対策       | 138 |
| 5. 6                    | 本章のまとめ           | 139 |
| <b>第6章 結 論</b>          |                  | 141 |
| 6. 1                    | 各章の総括と結論         | 141 |
| 6. 1. 1                 | 画像監視システムの必要性     | 141 |
| 6. 1. 2                 | 関連する画像処理研究       | 142 |
| 6. 1. 3                 | 人物の実時間検出         | 142 |
| 6. 1. 4                 | 検出人物の姿勢認識        | 143 |
| 6. 1. 5                 | 検出人物の認識          | 143 |
| 6. 1. 6                 | 視覚処理との関連         | 144 |
| 6. 2                    | 今後の課題と展望         | 145 |
| 6. 2. 1                 | 今後の課題            | 145 |
| 6. 2. 2                 | 監視システムの今後        | 147 |
| <b>参考文献</b>             |                  | 148 |
| <b>研究業績</b>             |                  | 156 |

# 目次

|        |   |    |
|--------|---|----|
| 図 1-1  | 画像監視システムの利用先                              | 2  |
| 図 1-2  | 設備監視用の画像監視システムの一般的な構成                     | 3  |
| 図 1-3  | 監視カメラの基本構成                                | 5  |
| 図 1-4  | 監視カメラの撮像素子                                | 5  |
| 図 1-5  | IP ネットワークを利用した画像送信の概要                     | 7  |
| 図 2-1  | 人物検出や認識の画像処理概要                            | 23 |
| 図 2-2  | 変化領域の検出例                                  | 27 |
| 図 2-3  | 時空間画像による動き検出                              | 29 |
| 図 2-4  | 動きベクトルの検出結果                               | 30 |
| 図 2-5  | 特徴点の検出例と特徴点の対応付けの例                        | 31 |
| 図 2-6  | 動き検出用に画像内に配置する小マスク                        | 32 |
| 図 2-7  | 3次元形状の推定                                  | 35 |
| 図 2-8  | 奥行きによる移動体の検出結果例                           | 36 |
| 図 2-9  | MHI を用いた移動体領域の検出結果例                       | 36 |
| 図 2-10 | 動きと色による領域の追跡例                             | 37 |
| 図 2-11 | 学習用人物画像例                                  | 38 |
| 図 2-12 | 画像から特徴を抽出するための Haar ウェブレット                | 38 |
| 図 2-13 | 人物画像と背景画像の学習イメージ                          | 38 |
| 図 2-14 | 歩行者の検出例                                   | 39 |
| 図 2-15 | 検出に利用する歩行者の画像特徴例                          | 39 |
| 図 2-16 | カスケードによる処理の高速化                            | 40 |
| 図 2-17 | $W^4$ での人物特徴の検出例                          | 41 |
| 図 2-18 | 人物と物体の認識例                                 | 42 |
| 図 2-19 | 人体モデルの当てはめ例                               | 43 |
| 図 2-20 | 人物の姿勢認識例                                  | 43 |
| 図 2-21 | 携帯ビデオカメラで撮影した画像からの実時間検知例                  | 44 |
| 図 2-22 | 教師付き学習のモデル                                | 45 |
| 図 2-23 | 門の入退出者数を計測するための領域                         | 47 |
| 図 2-24 | 通過時間の長さ と 人数 の関係 $P(\text{長さ} \text{人数})$ | 48 |
| 図 2-25 | 通過時間の長さ と 人数 の関係 $P(\text{人数} \text{長さ})$ | 48 |
| 図 2-26 | SVM による分類結果                               | 53 |
| 図 2-27 | SVM による多クラス分類方式                           | 55 |
| 図 2-28 | AdaBoost のアルゴリズム                          | 56 |
| 図 2-29 | AdaBoost による歩行者検知例                        | 57 |
| 図 2-30 | 人の網膜                                      | 58 |
| 図 2-31 | 視覚野の構造と機能                                 | 59 |
| 図 2-32 | MT 野の Mo 型細胞が反応する動きの組み合わせ例                | 60 |
| 図 2-33 | オプティカルフローと移動,回転の関係                        | 61 |
| 図 2-34 | 注視位置の変化                                   | 62 |
| 図 3-1  | ライン型前処理による監視処理の概要                         | 66 |
| 図 3-2  | 監視ラインの設定例                                 | 66 |
| 図 3-3  | 人物の通過に伴う輝度値の時間変化例                         | 67 |
| 図 3-4  | 時空間画像の例                                   | 68 |
| 図 3-5  | 学習に利用した人物画像例                              | 72 |
| 図 3-6  | 提案方式での人物の検出例                              | 72 |

|        |  |     |
|--------|--|-----|
| 図 3-7  | 提案方式による侵入者検知結果                           | 73  |
| 図 3-8  | 学習に利用した人物画像例 (多人数)                       | 74  |
| 図 3-9  | アスファルト上での監視結果例                           | 74  |
| 図 3-10 | 樹木の前での監視結果例                              | 74  |
| 図 3-11 | アスファルトを背景とした監視結果                         | 76  |
| 図 3-12 | 植物を背景とした監視結果                             | 76  |
| 図 3-13 | 別シーンの学習結果を利用した監視結果例                      | 77  |
| 図 3-14 | 充電部接近監視結果                                | 78  |
| 図 3-15 | 監視ライン近傍全体に侵入者検知を行った結果                    | 78  |
| 図 3-16 | 対象とする監視場所の例                              | 81  |
| 図 3-17 | 侵入者検出方式の一般的な枠組み                          | 82  |
| 図 3-18 | 動き補正後のフレーム間差分画像                          | 83  |
| 図 3-19 | 固定/旋回カメラからの MHI による移動体検候補の検出例            | 84  |
| 図 3-20 | 縦横方向の動き検出マスク                             | 86  |
| 図 3-21 | 提案検出方式のパラメータの関係                          | 87  |
| 図 3-22 | ブロックサイズと探索範囲に依存する処理速度                    | 88  |
| 図 3-23 | ブロックの配置                                  | 88  |
| 図 3-24 | 記録時間を変えたときの MHI と検出結果                    | 89  |
| 図 3-25 | 固定カメラでの侵入者候補領域の検出結果例                     | 91  |
| 図 3-26 | 固定カメラによる侵入者候補領域の検出数の推移                   | 92  |
| 図 3-27 | 旋回カメラからの侵入者候補領域の検出結果例                    | 93  |
| 図 3-28 | 旋回方向と逆方向に移動するシーンでの侵入者候補領域の検出数推移          | 94  |
| 図 3-29 | 侵入者候補領域の検出処理速度                           | 95  |
| 図 3-30 | 学習結果後の人物検出結果例                            | 97  |
| 図 3-31 | 仮想の旋回監視カメラ画像による検出結果例                     | 99  |
| 図 3-32 | 上下にランダムに旋回したカメラで撮影した<br>画像からの移動体候補の検出結果例 | 100 |
| 図 3-33 | 携帯ビデオカメラで撮影した画像からの移動体候補の検出結果例            | 100 |
|        |  |     |
| 図 4-1  | 対象としている画像例と画像の消点                         | 104 |
| 図 4-2  | 提案方式での姿勢認識処理手順                           | 105 |
| 図 4-3  | 人物領域の検出結果例                               | 106 |
| 図 4-4  | 検出した人物領域サイズを正規化した結果                      | 108 |
| 図 4-5  | 訓練員の移動状態と静止状態の分離                         | 108 |
| 図 4-6  | 電話と他の姿勢の分離                               | 109 |
| 図 4-7  | 学習数に伴う認識結果の推移                            | 111 |
| 図 4-8  | 特徴による認識結果の比較                             | 112 |
| 図 4-9  | 提案方式による姿勢認識結果例                           | 113 |
| 図 4-10 | 認識結果を表示するために準備した姿勢 CG                    | 114 |
| 図 4-11 | 学習画像と同一人物の姿勢の認識結果                        | 115 |
| 図 4-12 | 別人物の姿勢の認識結果                              | 116 |
| 図 4-13 | 下半身画像による 3 姿勢の学習結果                       | 118 |
| 図 4-14 | 姿勢認識結果を検出位置に合わせて表示した結果                   | 119 |
|        |  |     |
| 図 5-1  | 人物認識処理の概要                                | 122 |
| 図 5-2  | 移動人物の検出例                                 | 123 |
| 図 5-3  | 局所自己相関の局所パターン                            | 124 |
| 図 5-4  | SVM の多段戦略による多クラス分類                       | 126 |
| 図 5-5  | 4 人のシルエット検出例                             | 127 |
| 図 5-6  | 4 方向のシルエット検出例                            | 127 |
| 図 5-7  | 服装による人物と向き認識の実験結果例                       | 128 |

|        |                              |     |
|--------|------------------------------|-----|
| 図 5-8  | 4 種類の特徴を使った人物認識の実験結果         | 129 |
| 図 5-9  | 4 種類の特徴を使った向き認識の実験結果         | 129 |
| 図 5-10 | 8 人の 16 日間分の撮影画像を利用した認識実験結果  | 130 |
| 図 5-11 | 実験に利用した 8 人の画像例              | 131 |
| 図 5-12 | 形状不変な物体領域を検出した画像例            | 133 |
| 図 5-13 | 形状が変形しない物体認識の結果例             | 133 |
| 図 5-14 | データベース COIL の画像例             | 135 |
| 図 5-15 | 学習サンプル数と物体認識率 (使用画像: 7200 枚) | 135 |
| 図 5-16 | 形状不変物体の向きの認識結果例              | 135 |
| 図 5-17 | SVM による前面・背面の分類結果            | 136 |
| 図 5-18 | SVM による前面・左側面の分類結果           | 136 |
| 図 5-19 | SVM による前面・右側面の分類結果           | 136 |
| 図 5-20 | SVM による背面・左側面の分類結果           | 137 |
| 図 5-21 | SVM による背面・右側面の分類結果           | 137 |
| 図 5-22 | SVM による左側面・右側面の分類結果          | 137 |
| 図 5-23 | 2 クラス比較の順番で結果が変わる領域          | 138 |
| 図 6-1  | 動き処理と認識処理の比率                 | 144 |

# 表目次

|        |                                   |     |
|--------|-----------------------------------|-----|
| 表 1-1  | 画像出力形式                            | 6   |
| 表 1-2  | 画像圧縮方式                            | 6   |
| 表 1-3  | 画像監視システムに組み込まれている主なセンサ例           | 7   |
| 表 1-4  | 画像監視システムに必要な侵入者等の検出・認識・追跡機能       | 13  |
| 表 1-5  | 電気事業で活用されている画像監視システム例             | 16  |
| 表 1-6  | 監視箇所の特性と誤検知を引き起す環境変化              | 18  |
|        |                                   |     |
| 表 2-1  | 計測領域内の動き変化時間の長さと通過人数の頻度           | 47  |
| 表 2-2  | 機械学習に利用した画像例                      | 57  |
|        |                                   |     |
| 表 3-1  | 実験に利用する画像例                        | 71  |
| 表 3-2  | 設定パラメータ                           | 72  |
| 表 3-3  | 人物 551 画像、背景 2,838 画像の SVM 学習結果   | 73  |
| 表 3-4  | 提案方式の反応結果 (1,900 画像)              | 73  |
| 表 3-5  | 1,900 画像に対する実験結果の要約               | 73  |
| 表 3-6  | 人物 2,410 画像、背景 9,281 画像の SVM 学習結果 | 75  |
| 表 3-7  | アスファルト背景時の監視結果 (14,400 画像)        | 75  |
| 表 3-8  | 樹木が背景時の監視結果 (14,400 画像)           | 75  |
| 表 3-9  | アスファルトを背景とした実験結果の要約               | 76  |
| 表 3-10 | 樹木を背景とした実験結果の要約                   | 76  |
| 表 3-11 | 電力設備への接近監視結果 (6,000 画像)           | 78  |
| 表 3-12 | 電力設備への接近監視実験結果の要約                 | 78  |
| 表 3-13 | 固定カメラでの侵入者候補領域検出結果                | 91  |
| 表 3-14 | 各実験シーンの録画時間                       | 92  |
| 表 3-15 | 前処理での侵入者候補領域の検出率                  | 95  |
| 表 3-16 | 前処理での侵入者候補領域の過剰検出率                | 95  |
| 表 3-17 | 侵入者検知処理との連携結果                     | 96  |
| 表 3-18 | 機械学習に利用した画像例                      | 97  |
| 表 3-19 | マスクサイズに対する処理速度と検出率                | 98  |
| 表 3-20 | 仮想画像による実験結果                       | 99  |
|        |                                   |     |
| 表 4-1  | 訓練員の状態と姿勢                         | 105 |
| 表 4-2  | 正規化した姿勢の学習用画像例                    | 109 |
| 表 4-3  | 学習画像と同一人物の姿勢認識結果                  | 116 |
| 表 4-4  | 学習画像と別人物の姿勢認識結果                   | 116 |
|        |                                   |     |
| 表 5-1  | 15 日間の画像を学習し残り 1 日分を認識した結果        | 130 |

# 第1章

## 序論

### 1.1 はじめに

電力施設や石油プラントなどで設備の稼動状態の把握や、保守保全および警備の観点から、監視カメラを用いた監視システムが数多く導入されている。また、各企業の自発的な設置ばかりでなく、社会的な要求により監視システムの導入も増えている。例えば、2001年のアメリカ同時多発テロを契機に、2004年に改正された国際条約（SOLAS条約：海上人命安全条約）に港湾施設の保安強化義務が盛り込まれたため、国内の国際港湾施設に監視カメラが数多く設置されだしている。さらに近年、ネットワーク技術の進展と監視カメラの低コスト化や、日常生活での犯罪件数の増加などが背景となり、特殊な場所だけでなく日常生活の安全を見守るために監視カメラの台数が増える傾向にある。

現在、数多く設置されている監視カメラの画像を監視員が目視監視することは、問題発生時の経済損失に見合う効果がないと行われておらず、監視カメラの画像のほとんどが問題発生後の事後調査用として使われている程度である。また、監視員が配置できる状況でも、業務効率化の観点から一人の監視員が多数の監視カメラの画像をマルチ画面で監視することが強いられている。そのため、目視監視を補助あるいは代行する機能を備えた画像処理を実現できれば、事後調査や監視員の補助のみならず、これまで費用対効果の観点から目視監視できなかった場所を監視でき、問題発生の予防/予測などに監視システムを活用できる。

また、目視監視に近い機能を備えた監視システムは重要施設ばかりでなく、我々の日常生活の安全を見守ることが可能となる。特に、国内の社会的な問題、例えば、夫婦の共働きと核家族化により子供を見守る環境がなくなり子供が犯罪に巻き込まれる事件が増えている問題や、高齢化社会を迎え、離れて暮らす高齢者の日常生活を見守ることができない問題などに対し、監視カメラを用いた監視システムがそれらの対策の1つになり得る。今後、目視監視と同様な機能を備えた監視システムが実現できれば、社会を支えるコア・システムの一つとなることが予想される。以降、本論文では監視カメラを用いる監視システムを画像監視システムと呼ぶことにする。

本章では1.2節で既存の画像監視システムの利用先と、画像監視システムの構成と機能を示す。また、画像監視システムの中で自動化が唯一できていない部分が、監視員による目視監視であることを示す。1.3節では監視員による目視監視機能を画像処理で実現することが、シーン理解の難しい課題であることを技術史の観点から示す。1.4節で監視員の目視機能を検出・認識・追跡としてとらえ、それらを実現する既存画像処理の概要と課題を示す。1.5節では電気事業で利用している画像監視システムを参考に、現在の画像監視システムに対する社会的なニーズと技術課題を示す。さらに、それらの諸課題の中から目視監視を補

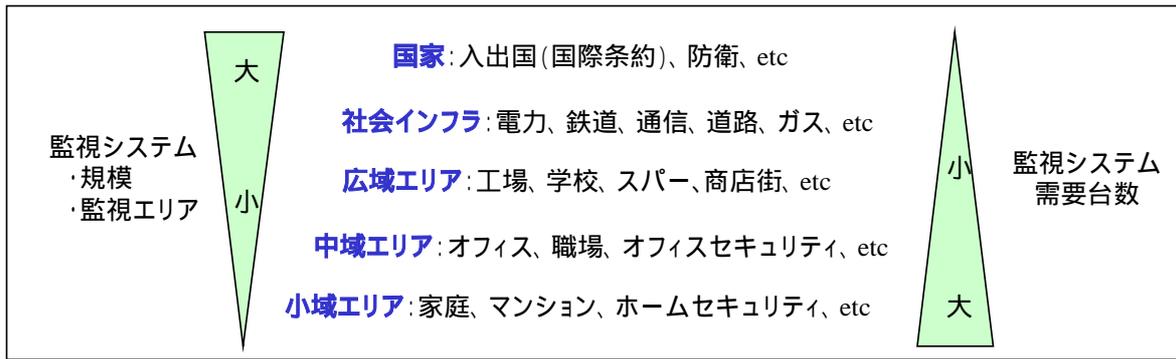


図 1-1 画像監視システムの利用先

助あるいは代行する画像処理の実現に重要となる2つの課題「検出」と「認識」を取り上げ、3章以降で示す研究の位置付けと研究のアプローチを述べる。また、画像処理以外の各種センサを利用しても、最終的な人物の検出と認識は目視に相当する画像処理が必要であることを示す。1.6節で本論文の全体構成を示し、1.7節で本章のまとめを示す。

## 1.2 画像監視システム

### 1.2.1 画像監視システムの導入先

監視カメラが市販されだした初期のころから画像監視システムを導入している業界として、電力業界がある。電力設備などでは無人の発電所や変電所などに、監視カメラを設置することを業界団体の規定(発電機規定、電力保安通信規定等)として義務づけているため、国内の主要な監視個所だけでも数万台以上の監視カメラが設置されている。それらの画像監視システムの中には、画像処理による侵入者の自動検知機能を導入している事例も多い。しかし、木々の揺れや太陽光の変化あるいは影や動物などで誤報が頻発し、本業務に支障をきたすため、問題発生時の現場の確認、あるいは問題発生後の事後調査に記録画像を利用する形態が多い。その他の業界、例えば石油プラントなどの製造業やスーパーなどの小売業などでも、監視カメラが数多く導入されているが、画像の利用形態はほぼ同様である。

1990年代後半から監視カメラ画像の使われ方が受身的な事後調査だけでなく、予防対策に向けた積極的な使われ方に代わってきている。例えば、英国政府はテロ予防対策として1998年頃からロンドン市内に数多くの監視カメラを設置しはじめ、現在、ロンドン市内だけで50万台、全土で250万台の監視カメラが設置されていると言われている[21]。2005年7月7日のロンドン同時爆破テロ事件では、監視カメラの映像2500本を解析し5日後に実行犯を特定している[21]。なお、ロンドンの監視カメラには、犯罪者の顔と画像内の人物を照合するシステムが導入されている。しかし、現状は目視監視に及ばず、監視カメラ画面を切り替えながら多数の監視カメラ画像を監視員が目視している状態である。

日本でも2001年に警視庁が新宿の歌舞伎町に50台の街頭防犯カメラを設置し、防犯対

策に利用することが TV や新聞のニュースとして取上げられていたが、現在はすでに各地の繁華街や商店街に監視カメラが設置されている状態である。また、国内では高速道路や一般道に監視カメラが数多く設置されており、道路管理や混雑状況の提供、ならびに事件解決に有効活用されている。

近年、公共の監視ばかりでなく、ホームセキュリティ事業に民間企業が参入しはじめている。一例として、2001年に設立された関西地域でのホームセキュリティの専門会社（関電セキュリティ・オブ・サイエンス株式会社）を参考にすると、その新規契約数が年々増えており、2005年だけでも約2,200件の家庭と新規契約が結ばれている。2005年には九州でも同様のホームセキュリティ会社（九電ホームセキュリティ株式会社）が設立されている。従来、家庭内の防犯用に赤外センサが主に利用されているが、近年、監視カメラの設置も増えはじめている。ただし、費用対効果の観点から、家庭内に設置した監視カメラの画像を警備員が全て目視確認することは不可能である。そのため、より積極的に家庭内の安全安心を見守る画像監視システムが必要とされはじめている。また、家庭内の画像監視はプライバシーの問題があるため、目視ではなく画像処理による自動監視が必須であり、問題発生時に限り契約者や警備員へ目視要求を出す画像監視システムが望まれている。

図 1-1 に画像監視システムの利用先を大分類し、システムの規模や監視エリアサイズならびに、今後予想される画像監視システムの需要台数を示す。すでに公共インフラや職場などを監視する画像監視システムが一般化しており、現在、前述のホームセキュリティに安価な画像監視システムが導入されはじめようとしている。それらの膨大なカメラ画像を目視監視することが不可能なため、監視カメラ画像を自動的に監視し、侵入者等の検出時

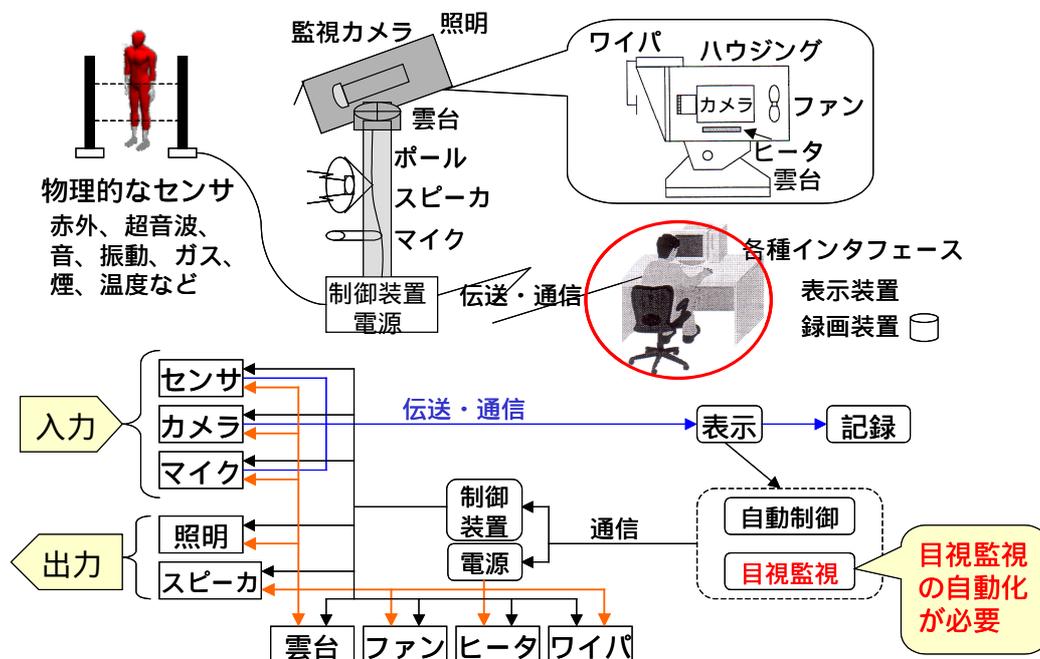


図 1-2 設備監視用の画像監視システムの一般的な構成

に警報を出すシステムが社会的に求められている。そのため、民間企業がビジネスチャンスとして画像監視システムの研究開発に取り組んでおり、昨今その一部が商品化されはじめている。しかし、いずれの商品も目視の補助あるいは代行として十分な機能が得られていない状況である。

画像監視システムは図 1-1 に示すように様々なレベルがある。その中で、電気事業などの社会インフラで利用されている画像監視システムを対象に、その構成と機能を下記に示す。

### 1.2.2 画像監視システムの構成と機能

画像監視システムは、監視カメラ、警報装置、録画装置、情報配信装置、画像処理など多数の装置や処理で構成されている複合システムである。ここでは電力設備監視用などに導入されている画像監視システムの構成と機能を概説し、既存の画像監視システムの中で自動化できていない部分が監視員による目視であることを示す。

図 1-2 に代表的な画像監視システムの構成を示す。画像監視システムは、監視カメラ、雲台、ハウジング（ハウジング内温度を一定に保つファンやヒータ、ワイパ）、警告用スピーカ、現場音収集用マイク、監視カメラを取り付けるポール、各種センサなどの物理的な機器と、目視確認に用いるマンマシーン・インタフェースや画像/制御情報などの通信手段から構成されている。撮影映像は、従来、同軸ケーブルで監視員まで届けられていたが、近年、IP ネットワーク網を使い遠隔地の監視員へ画像を送信するケースが増えている。

### 1.2.3 監視カメラ

画像監視システムで光を画像に変換する監視カメラは、主に図 1-3 のようにレンズ、撮像素子、信号処理回路と出力回路の構成になっている。

#### (1) レンズ関連

設備監視などに導入されている監視カメラは、オートアイリス機能やオートフォーカス機能が付いた通常のレンズが広く利用されている。広域を監視する監視カメラには、監視方向とフォーカス等を事前にセットできるプリセット機能が付いているものも多い。近年、180 度の画角を撮影する魚眼レンズや、双曲面ミラとの組み合わせで 360 度の全周画像を 1 ショットで撮影する装置なども画像監視に使われだしている。しかし、設備監視では依然として通常レンズの監視カメラが主に利用されている。

#### (2) 撮像素子

現在広く使われている監視カメラの撮像素子は、主に CCD 型 (Charge Coupled Device) と CMOS 型 (Complementary Metal Oxide Semiconductor) に大別できる[120]。いずれも 2 次元に配置した受光素子 (フォトダイオード) に光を投影し、受光素子ごとの受光量を電気信号に変換するデバイスである。

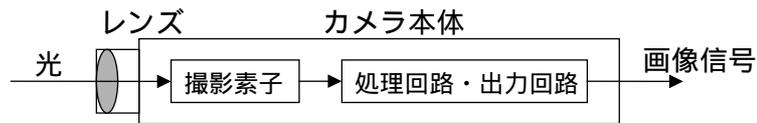


図 1-3 監視カメラの基本構成

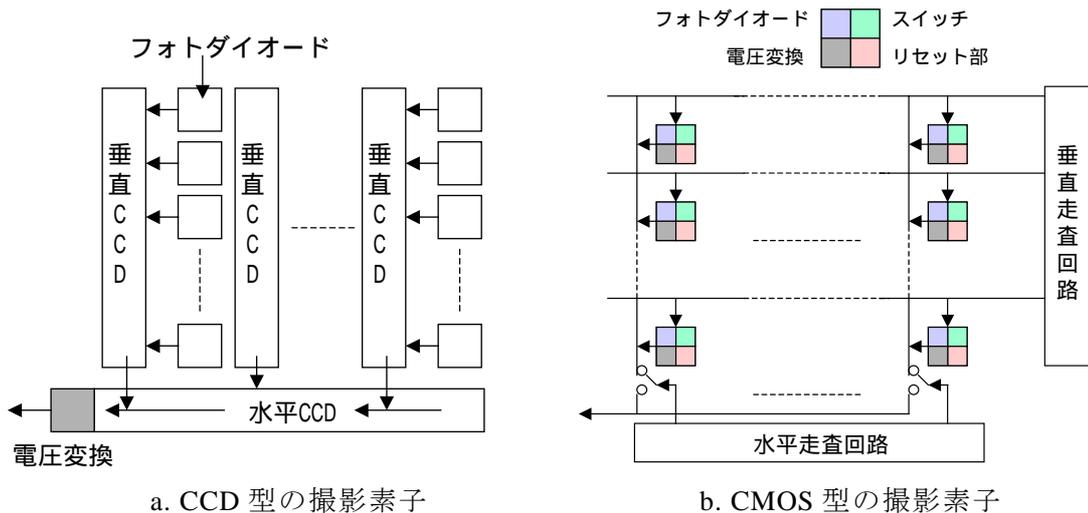


図 1-4 監視カメラの撮像素子

CCD 型はフォトダイオードで光電変換した電子を、垂直 CCD（信号電化を蓄積・転送するためのレジスタに相当するデバイス）と水平 CCD で出力部に転送し、出力端子の直前で電子を信号電圧に変換（増幅）する方式である（図 1-4a 参照）。CMOS 型はフォトダイオードで光電変換した電子を個別に信号電圧に変換（増幅）し、画素選択のスイッチングで信号を出力する方式である（図 1-4b 参照）。従来、画質の面で優れている CCD カメラが監視カメラの撮像素子として広く利用されてきたが、昨今の CMOS 型撮像素子の研究成果により両者の画質差が解消され、CCD 型に比べ小型化が可能で省電力な CMOS 型カメラが普及しはじめている。

その他、光電子増幅管と CCD 撮像素子の組み合わせで星明り（ $10^{-3}$ ルクス程度）でも撮影できる光電子増倍方式や、シャッタ時間を長くして撮影する電子増倍方式などのカメラがある。赤外投光器で投光した赤外線反射を CCD カメラで撮影する方式や、一般的な撮像素子ではなく赤外受光素子（例えば、 $7.5 \mu\text{m} \sim 14 \mu\text{m}$  程度の波長を吸収する素子）を使い物体の表面温度を見るサーマルカメラなども利用されている。近年、サーマルカメラを利用した人体検出装置が市販の乗用車に搭載されはじめている[122]。さらに、従来のカメラよりも解像度を上げた HDTV カメラ（High Definition Television: 高精細度 TV）なども普及しはじめている。

表 1-1 画像出力形式

|         | NTSC            | PAL                  | SECAM                |
|---------|-----------------|----------------------|----------------------|
| 走査線数    | 525 本           | 625 本                | 625 本                |
| フレーム数/秒 | 30              | 25                   | 25                   |
| 垂直同期周波数 | 60Hz            | 50Hz                 | 50Hz                 |
| 採用国     | アメリカ、カナダ、<br>日本 | ドイツ、イギリス、<br>オーストラリア | フランス、ロシア、<br>サウジアラビア |

表 1-2 画像圧縮方式

| 圧縮方式                    | 圧縮アルゴリズム                                | ビットレート                |
|-------------------------|---|-----------------------|
| <b>Motion-JPEG</b>      | <b>DCT (Discrete Cosine Transform)</b>  | <b>64kbps～10Mbps</b>  |
| <b>MPEG-1</b>           | DCT、動き補償                                | <b>1Mbps～1.5Mbps</b>  |
| <b>MPEG-2</b>           | DCT、動き補償                                | <b>4Mbps～80Mbps</b>   |
| <b>MPEG-4</b>           | DCT、動き補償、オブジェクト別圧縮                      | <b>4kbps～38.4Mbps</b> |
| <b>Motion-JPEG2000</b>  | <b>DWT (Discrete Wavelet Transform)</b> | あらゆるレートに対応            |
| <b>H.264/MPEG-4 AVC</b> | DCT、動き補償、フレーム間予測                        | あらゆるレートに対応            |

### (3) 画像出力

既設監視カメラの多くが、NTSC (National Television Standard Committee) 形式のコンポジット信号を出力している。NTSC 形式は、1 秒間に 29.97 枚のインターレース (飛び越し走査) 画像を出力する。各画像は 480 本の垂直解像度、525 本の走査線で構成されている。世界的に見ると、ドイツやイギリスなどでは NTSC 形式を改良した PAL (Phase Alternation by Line) 形式が利用され、フランスやロシアなどでは SECAM (Séquentiel Couleur à Mémoire) 形式が利用されている。各形式の概要を表 1-1 に示す。

近年のネットワークカメラや Web カメラなどは表 1-1 の出力形式ではなく、撮像素子の出力を MPEG 方式や JPEG 方式で圧縮したデータとして直接出力し、パソコンで表示する方式も増えている。表 1-2 に代表的な画像圧縮方式を示し、各方式の画像圧縮アルゴリズムと送信時のビットレートを示す。

#### 1.2.4 各種センサと画像送信

広域監視が必要な画像監視システムには、赤外センサなどの各種センサと監視カメラの連携が図られている。一例として、火力発電所の広い構内を監視する画像監視システムには、赤外センサ (赤外線アクティブセンサ：発光器から発射した赤外線を受光器で検出) が反応した方向に監視カメラを自動的に旋回させ映像を録画する機能がついているものが多い。赤外センサ以外にも、表 1-3 に示すようなセンサが画像監視システムの中で利用されている。いずれのセンサを利用しても、センサが動作した場合、その原因調査は監視カメラ画像を用いた目視に頼っている状態である。

表 1-3 画像監視システムに組み込まれている主なセンサ例

| 種類   | 方式          | 検出方式  |
|------|-------------|---|
| 温度   | 赤外線パッシブセンサ  | 人物が発する熱を赤外線で検出                                |
|      | 赤外線アクティブセンサ | ・発光器から発射した赤外線を受光器で検出<br>・赤外投光器から放射する赤外線の反射を検出 |
| 電波等  | 超音波センサ      | 送波器から放射する超音波の変化を検出                            |
|      | マイクロ波センサ    | 送波器から放射するマイクロ波の変化を検出                          |
| 振動・音 | 振動センサ、音センサ等 | 電気的あるいは光の変化等で振動や音を検出                          |
| 光    | レーザーセンサ     | 足元にレーザ光を照射し、レーザ光の当たる位置により奥行き変化を検出             |
|      | 光ファイバセンサ    | フェンスに光ファイバを沿わせ、フェンスの変形をファイバ内の光の変化で検出          |
| その他  | 煙、ガスセンサ     | 電気抵抗変化等により煙やガスなどを検出                           |

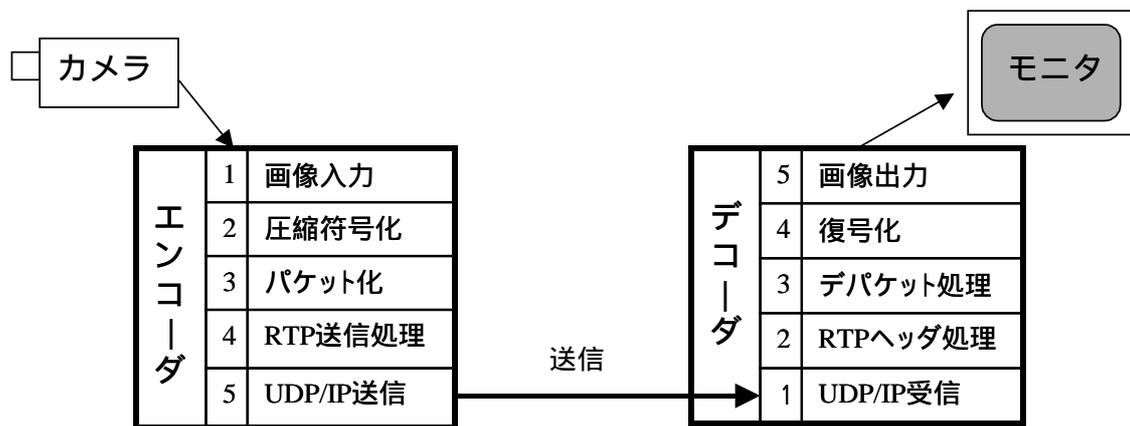


図 1-5 IP ネットワークを利用した画像送信の概要

従来、監視カメラの画像は同軸ケーブルで警備室に直接送信されていたが、最近は光ファイバによるネットワーク網の整備が進み、IP ネットワークを使う画像送信が増えている。図 1-5 に IP ネットワークを利用した画像送信処理の概要を示す。図 1-5 に示すように送信側のエンコーダでは、画像を MPEG 形式などで圧縮しパケット化した後、時刻などを記録したリアルタイム・データ転送プロトコル RTP (Real-time Transport Protocol) に加工し、下位ネットワーク層の UDP/IP プロトコルでデータを送信している。一方、受信側のデコーダでは、逆の手順で画像を再生している。

### 1.2.5 画像蓄積と画像監視

監視カメラから送られてくる画像は、従来、コマ落としによる長時間録画が可能なタイムラプスビデオでテープに録画していた。しかし近年、ハードディスクの大容量化と低価格化により、監視カメラで撮影した画像を MPEG 等で圧縮し、ハードディスクに長時間録

画する方式に変わりつつある。

一方、目視監視の補助として、監視カメラで撮影した画像から背景差分やフレーム間差分で画像内変化を検出する機能が多くの画像監視システムに導入されている。しかし、他のセンサ同様、変化検出後の原因調査は目視確認が前提となっており、目視監視を代行するレベルに達していない。むしろ、既存の画像監視システムに組み込まれている画像処理は、太陽光の変化や木々の揺れ、動物などによる誤報が多い状態である。

以上のように、本節では監視員による目視部分を除いて、現在の画像監視システムはほぼ自動化できていることを示した。今後、電力設備等に設置されている膨大な数の監視カメラの画像を監視するには、監視員の目視に相当する画像処理の開発が不可欠である。しかし、目視監視に相当する機能を画像処理で実現することが現在でもチャレンジングな課題であることを、次節で画像処理の歴史的な観点から説明する。

### 1.3 画像処理の歴史的観点からの考察

画像処理の技術史を、画像監視システムの視点から年代別に考察する。その後、目視監視を実現している視覚のシーン理解について概説する。

#### 1.3.1 画像処理の技術史

##### (1) 1960年代以前

紙とインクによる画像処理技術として、1440年代頃に J.Gutenberg により発明されたプレス式の活版印刷技術や、日本の江戸時代に(1765年鈴木春信らの多色刷り技術の開発で)盛行した浮世絵などがあげられる[1]。その後、半世紀後の1839年には L.J.M.Daguerre がいわゆる銀板写真の開発に成功し、約1世紀後の1925年には現在の小型スチールカメラの原型となったカメラ「ライカ A」が販売されている[1]。同年の1925年に J.L.Baird による機械走査式テレビジョンの公開実験が行われ[2]、2年後の1927年には高柳健次郎が走査線40本、1秒あたり14コマの電子走査式ブラウン管受信機を開発している[1]。1933年には V.K.Zworykin が電子走査式の撮像管アイコノスコープ(光電変換機能と電子的な走査機能の両方をそなえた真空管)の開発に成功し、1945年に A.Rose らがより感度が高く鮮明な画像を送れるイメージオルコシス管の開発に成功し本格的な放送時代への道を開いた[1][2]。その後のテレビジョンの研究により、アナログ電子回路による画像のエッジ強調や雑音除去などの画像処理基礎技術が開発され、それらの技術はその後の計算機の出現により現在のデジタル画像処理へと受け継がれている。

1946年に米国の弾道研究所が計画したデジタル計算機 ENIAC(Electronic Numerical Integrator And Calculator)が稼動し、1951年には米国政府の国政調査に UNIVAC(UNIVersal Automatic Computer)が利用されるなど、1940年代後半から1950年代にかけて計算機が使われはじめた[2][3]。当時の計算機では大きな2次元配列の画像を扱うデジタル画像処理ができなかった。そのような状況でも、米国では1950年から文字読取りの研究を開始し、1955年には数字の読取り装置(OCR)が発売される[4]など、計算機を画像処理に活用する研究が

始まっている。

## ( 2 ) 1960 年代

1960年代に入り、IBM360 やミニコンピュータ DEC/PSP-8 により計算機で画像が扱えるようになりだした[2]。この頃、現在の数値解析や画像処理の基礎となる各種のアルゴリズムが考案されている。その一例として、1965年に Cooley と Tukey により考案された高速フーリエ変換のアルゴリズム(FFT : Fast Fourier Transform)がある。現在、FFT は衛星画像や航空写真の解析、医療画像の解析などの様々な画像処理で不可欠なものとなっている。このようなアルゴリズムの検討の中、1960年代後半に米国マサチューセッツ工科大学の人工知能研究所などが中心となり、一般的なシーンや3次元物体の認識が始まった。特に、積木の状態を理解する Block World の問題が人工知能研究の代表的な一分野とされ、現在の画像処理によるシーンの理解、物体認識、ロボットビジョンなどの先駆的な研究となった。1970年には人工知能の専門誌として Artificial Intelligence (AI)ジャーナルが創刊され、その年の3号に積木シーンの理解に関する研究成果[5]が AI ジャーナル内の初の画像処理論文として掲載されている。このように、初期の頃よりシーン理解がチャレンジングな課題として取上げられていた。

## ( 3 ) 1970 年代

1970年代に入り海洋観測衛星や気象衛星などの衛星画像解析(リモートセンシング)や、医療用の X 線 CT 装置、文字認識装置、工業用ロボットの画像処理装置等、様々な研究が加速された[2]。また、1960年代後半に始まったパターン認識による顔認識等の研究[6]も1970年代に成果が報告されている。この頃、画像処理の研究者が増えたことで、現在の米国電気電子学会(IEEE : The Institute of Electrical and Electronics Engineers)の Transaction on Pattern Recognition and Machine Intelligence の前々進にあたる Computer Graphics and Image Processing が 1972年4月に創刊され、翌年1973年には第1回パターン認識国際会議(International Conference on Pattern Recognition)が開催されている。1960-70年代に、計算機の普及とともに現在の画像処理やパターン認識の基礎をなす多くのアルゴリズムが開発されている。1979年からそれらの画像処理アルゴリズムを整理し、パッケージ化(SPIDER: Subroutine Package for Image Data Enhancement and Recognition)する作業が始まっている[7]。1970年代は、計算機的能力と費用や処理速度の面から、画像処理の研究は主に研究機関で基礎研究が行われていた時代である。例えば国内では、1971年から10年間続けられた工業技術院のプロジェクト「パターン情報処理システムの研究開発」などが上げられる。なおこの頃、カメラの撮影デバイスには撮像管が使われていたが、1970年代初頭からカメラの撮像素子として CCD を使う研究が始まっている[120]。

## ( 4 ) 1980 年代

1980年代は画像処理専用プロセッサを搭載したワークステーションが広く普及し、画像処理を計算機であつかえるようになった。プリント基板検査装置などの専用システムが組めるようになり、費用対効果が成り立つ工業製品の検査用に画像処理システムが投入され

はじめた。特に、限定された環境下で特定の検査を行うシーン理解は目視機能を超え始めた。そのため、より変化に富んだ図面認識[118]などの課題を扱う研究が盛んになり始めた。

## ( 5 ) 1990 年代

1990 年代に入り、パーソナルコンピュータの性能向上と、高性能で安価なデジタルカメラやビデオカメラなどの普及により、個人で画像処理プログラムを開発できる環境が整った。その一方で、インターネットを經由して様々な画像を閲覧できるようになった時代である。特に、1993 年に M. Andreessen が Web ブラウザ「NCSA Mosaic」を発表したことで、World Wide Web が広く普及しはじめた。この 1990 年代は、画像処理とインターネットが結びつき画像転送用に圧縮/復元技術が活発に議論され、JPEG (Joint Photographic Experts Group)による静止画像の圧縮や MPEG (Motion Picture Experts Group) による動画像圧縮方式の世界的標準作りが進んだ。

また、自律移動型ロボットの目として画像処理が積極的に使われるようになり始めた。例えば、1997 年に自律移動型ロボット同士がサッカーゲームを競う第 1 回国際会議 (ロボカップ) の開催や、ロボットとの共存共栄を目指した研究[8]などが始まっている。顔認識では 1993 年に米国国防省防衛高等研究計画局(DARPA : The Defense Advanced Research Projects Agency)が主催した FERET(FacE REcognition Technology)[9]と呼ばれる顔認証アルゴリズム・コンテストにより共通の評価データベース基盤を持ったことで、急速にアルゴリズムが発展した。近年は、動画像圧縮 MPEG7 の標準化に顔の検出や認識処理の導入が検討され始めている[10]など、顔検出/認識が一般的な技術となっている。

画像監視を見てみると、1990 年代後半にパーソナルコンピュータあるいはワークステーションの高速化により、画像処理による実時間監視が費用対効果の面から実現可能となり始めた。そのような背景の元、1998 年に DARPA がビデオ監視プロジェクト(Video Surveillance and Monitoring)を開始し、ヨーロッパや日本などでも画像監視に特化した会議が開催され始め、画像処理の国際会議のトピックスに画像監視が取上げられるようになった。特に、1998 年に画像監視に特化した IEEE の第 1 回画像監視ワークショップ(IEEE Workshop on Visual Surveillance)が開催されている。2000 年には共通のテスト画像を使い侵入者の検出と追跡のパフォーマンスを競うワークショップへと発展している。

## ( 6 ) 2000 年代

2001 年 9 月 11 日の米国での同時多発テロ事件により、セキュリティへの社会的な感心が高まり、画像処理による監視システムとともに、顔、指紋、虹彩、静脈などを使う画像処理による個人認証技術が、空港、主要設備の入退出管理、銀行の ATM やカード照合等に急速に利用されだした。一例として、アメリカ同時多発テロを契機に改正された国際条約 (SOLAS 条約 : 海上人命安全条約) に港湾施設の保安強化義務が盛り込まれ、世界の国際港湾施設に監視カメラが数多く設置されはじめている。

現在、監視カメラや Web カメラの低価格化などにより、数え切れない数の監視カメラが街頭や主要施設、あるいは商店街に設置されているが、近年でも年間 100 万台の勢いで増加している[11]とされている。また、インターネットを通して監視カメラ画像をライブで

閲覧できるサイトが増えている。一例として、国土交通省では港湾や河川の画像をリアルタイムで配信している 100 以上のサイトを集めたホームページを災害情報ページ[12]として公開している。また、警備会社がホームセキュリティと称して個人住宅の警備に力を入れており、公共の場だけでなく家庭内に監視カメラが入り始めている。一方、駅、スーパー、商店街あるいは職場やマンションのエレベータ等、日常生活のいたるところに監視カメラがあり、個人情報保護の観点からプライバシーの問題が議論される時代となっている。

これらの監視カメラ画像の殆どは、問題発生後の現場状況の確認、あるいは記録画像を事後調査に利用する程度に留まっている。現在、大量に蓄積された画像から目的の画像を検索する研究[13]や歩行者検知の研究[62]など、目視監視にかわる画像監視システムの実現を目指し積極的に研究開発が進められている。しかし、今のところ目視と同等あるいはそれ以上のシーン理解を行う画像監視システムを実現できていない状態である。

### 1.3.2 視覚のシーン理解

1959年頃から D.H.Hubel と T.N.Wiesel らによる脳の 1 次視覚野の構造と機能の生理的な研究[117]が開始され、脳細胞が行う画像処理の解明が始まった。(なお、Hubel と Wiesel はそれらの研究成果で 1981 年にノーベル医学・生理学賞を受賞している。) 生理的な解明結果に刺激を受けて、1970 年代に D.Marr らが脳機能の研究と画像処理を具体的に結びつけた議論を始めている。その結果、物体認識やシーン理解の問題は 2 次元画像から 3 次元シーンを理解する逆光学であり、問題は解が一意に決まらない不良設定問題 (ill-posed problem) であることを T.Poggio らと共に示している[14]-[16]。現在、指紋あるいは顔認識などの対象を限定した画像処理の実用化が進んでいる。しかし、より一般的なシーン理解は、対象を限定することが難しい不良設定問題に相当するため実用化段階に到達していない[2]。そのようなシーン理解の一つに、監視カメラ画像の自動監視があげられる。

不良設定問題を視覚がどのように処理しているかを解明できれば、シーン理解の研究に大きく貢献するばかりでなく、画像監視システムの機能開発の参考になることが予想される。特に、国内では脳関連の文部省重点領域研究の代表者が集まり 1993 年頃に「脳の世紀推進会議」が発足し、世界的な協力のもと視覚処理も含めた脳機能の解明が進められ、多くの発見がなされている。下記に脳の視覚処理を概観する。

網膜や大脳での機能分化とその構造は、DNA の設計図に従って整然と構成されている。網膜上で DNA に従い作成されるロドプシンと呼ばれるたんぱく質が、特定の光の波長 (400nm~700nm の内、長波長、中波長、短波長) に反応し、光を信号に変えることから処理がはじまり、それらの信号が大脳の視覚野へ伝達される[17]-[19]。視覚野は階層構造をもち、機能領域ごとに特定の部位に集まって存在している。また、その構造と機能は人やサルでもほぼ同じである。網膜から視覚野へ送られる信号の伝達経路は、大きく 2 系統 (形や色を解析する系統、動きを解析する系統) に分かれ、この 2 系統の処理で様々な環境下で安定に機能する画像処理を実現している。これまでの脳科学と計算理論の研究から、視覚処理の多くのことが分かってきた。しかし、残念ながらその処理の実現方法はまだ研究段階であり、これまでの研究成果を簡単に画像処理へ導入できる状態ではない。

脳のニューロンは 100 スパイク/秒程度で発火しうる[20]ので、計算機にたとえると

100Hz 程度の処理速度である。今日の格安ノートパソコンでもその CPU 速度が数 GHz であり、明らかに単体の処理速度はパソコンのほうが上である。しかし、100 億個以上も存在するといわれている脳細胞の構造化と機能分化により、パソコンによる画像処理以上のシーン理解を可能にしている。今後、特に脳の視覚野の構造と機能の解明が、シーン理解の研究に大きく貢献するものと考えられる。

同様の考え方でマサチューセッツ工科大学内に 1992 年に設立された Center for Biological & Computational Learning では、脳の主要な機能の働きが学習にあるとして、計算論的な機械学習の観点から脳機能の解明を進めている。間違いなく学習は脳の重要な働きの一つであり、シーン理解での物体検出や認識、環境変化への適応など、これまでの画像監視システムが不可能であったことを可能にする技術である。一方、脳には動きの解析など、学習以外で説明可能な機能も多く存在するが、学習と動きの解析を結びつけたシーン理解の研究はこれからの研究課題である。

#### 1.4 要求される画像処理の概要

画像監視システムのハードウェア（監視カメラや録画装置等）は既存技術レベルで充分実用に耐え、日常の監視業務の中で利用されている。しかし、画像監視システムの一構成要素である監視員による目視監視（侵入者や車両等の検出や認識機能）は自動化ができていない部分である。

現在、監視員が行っている目視監視は、侵入者の監視、作業の安全監視、積雪の推移監視、河川の水位監視、設備メータの読取などその監視内容は多岐に渡る。その中でも、積雪量や河川の水位読取り、あるいは各種メータの読取[123]など、画像内の位置を限定できる画像処理は自動化が可能な状況である。しかし、環境が大きく変化する屋外で侵入者を検出・認識・追跡する実用的な画像処理は研究段階であり、現在のところ侵入者等の監視は目視に依存している状態である。

目視監視では画像内から人物や車両等を選択的に検出し、検出した人物などの姿勢や動きから動作を認識するとともに、全身画像から許可されている人物あるいは既知人物であるかを判定している。さらに、監視領域内での移動位置を観察し、危険領域への接近を監視するなど、侵入者を監視するだけでなく作業員の安全を見守る役割なども果たしている。このような目視監視を実現するには、少なくとも人物を選択的に検出し、その姿勢などから動作を把握するとともに、服装や全身画像から社員か侵入者であるかを認識する画像処理が必要である。さらに、移動軌跡から危険領域への接近を判断する機能等が必要である。表 1-4 に目視監視を代行する画像監視システムに要求される機能を示す。

監視画像内から侵入者等を選択的に検出/認識できれば、その後の追跡が容易となり追跡誤り等も低減できる。そこで、目視を代行する画像監視システムの開発に向け、以下では検出と認識の実現方式の代表的な関連研究を概観し現状の課題を示す。さらに、検出や認識に不可欠な機械学習の概要を説明する。なお、関連研究の詳細は 2 章で示す。

表 1-4 画像監視システムに必要な侵入者等の検出・認識・追跡機能

| 主要機能 | 画像監視システムでの処理の主な内容  |
|------|--|
| 検出   | <ul style="list-style-type: none"> <li>・ 画像内の変化領域の検出</li> <li>・ 移動体領域候補の検出</li> </ul>  |
| 認識   | <ul style="list-style-type: none"> <li>・ 特定監視対象の選択的な検出<br/>(人物、車両、動物、自転車などを選択して検出)</li> <li>・ 検出物体の姿勢や向きなどの認識</li> <li>・ 検出物体の個別認識<br/>(顔認証、全身画像による認証、歩き方による認証など)</li> </ul> |
| 追跡   | <ul style="list-style-type: none"> <li>・ 検出/認識対象の移動軌跡の検出</li> </ul>  |

### 1.4.1 変化領域の検出

侵入者などが監視カメラのアンクルに入ると画像内に変化が発生する。その変化を背景画像との差分で検出する方式や、動きや色の変化で検出する方式、あるいは人の形状に似ている部分を検出する方式などが提案されている。

#### (1) 画像差分による領域検出

既存の画像監視システムは背景差分やフレーム間差分で画像内の移動体検出を実現しているものが多い。これらの画像差分による移動体検出では、画像の小領域の時間的な平均とその標準偏差などを計算し、変化が一定値（閾値）を超えた部分を移動領域として検出している。例えば、背景差分やフレーム間差分の結果から、輝度値の標準偏差の倍数を動的な閾値にする方式[29][30]や、画像ヒストグラムの級内分散を最小にする分割を行い背景と物体を分離する方式[32]なども広く利用されている。カルマンフィルタによる背景画像の更新[33]や、画像の固有値を変化検出に利用する方式[34]、あるいは画像内の各画素の輝度値を1年以上の長期間に渡り計測し、各画素の輝度値の統計分布を使い変化領域を特定する方式[30]などが提案されている。

また、照明変化への対応を目的に画像の輝度値を正規化した「正規化距離」による変化検出[35]や、正規化距離と画像の固有値を組み合わせて使う方式[36]、隣接する画素の輝度値の明度増分を用いる「増分符号化相関」による方式[37]など数多くの移動体検出方式が提案されている。さらに、画像差分だけでは雑音等の影響で移動体領域を正確に検出できないため、画像からエッジや線分を検出し差分とともに活用する方式[38][39]や、差分で大まかに移動体領域を決定した後に動的輪郭モデルで輝度境界を成長させ移動体を検出する方式[40]なども提案されている。隣接する領域の相関の強さ（画像変化の共起性）で木々の揺れなどを除外する方式[41]や、歩行者の直線的な動きと背景の動きを分離する方式[42]なども提案されている。

いずれも固定カメラが前提であり、旋回中のカメラで撮影した画像や背景が変化する場合などに適用できない問題がある。また、検出した変化領域が何であるかを判定する処理

が不可欠である。特に、特定監視対象だけを選択的に検出するには、変化領域の認識あるいは分類処理が必要となる。一般に、画像差分処理による変化領域検出は、認識や分類の前処理として利用することが多い。

## (2) 動きや色などによる領域検出

画像内の動き特徴により移動体を検出する多くの方式が提案されている。例えば、画像内の動き変化の大きさをグレー画像化し、その輝度勾配と濃度変化で領域を分割し移動体領域を決定する方式[43]や、画像内の移動をアフィン変換などのフィッティングモデルで表現し、そのモデルの係数で画像内を領域分割する方式[44][45]などがある。移動カメラで移動体を検出する方法として、周囲の画素との相対的な移動量から3次元形状を推定し、周りとの差が際立つ部分を検出する方式[46][47]も提案されている。また、画像内の動き成分を空間に投影して背景と移動体の動きを分離する方式[48][49]などがある。しかし、いずれの方式も実時間で動作させるには、オプティカルフローなどの動き特徴の計算を含め処理全体の高速化が必要である。また、(1)の「画像差分による領域検出」と同様に、監視対象だけを選択的に検出するには、検出した領域を特定する認識処理が不可欠である。

画像内の動きと輝度値を使う移動体検出方式[50]が1980年に提案されて以来、動きと他の特徴を組み合わせた多くの移動体検出方式が報告されている[45]。例えば、動きと色による領域分割を繰返す方式[51][52]、動きとテクスチャによる領域分割を利用する方式[53]、時間方向のテクスチャ相関と時間及び空間方向の領域相関を使い領域を分割し移動体を検出する方式[54]などが提案されている。扱う特徴を増やすことで、精度の高い移動体検出が可能になるが、弛緩法などによる繰返し計算が必要となる。そのため、これらの従来方式を実時間監視に利用するには特別なハードウェアを用いるか、計算量を激減する新たなアルゴリズムを考案する必要がある。

## (3) 機械学習による選択的な領域検出

歩行者画像と背景画像を機械学習し、画像内から歩行者を検出する研究[55][56]が進められている。特に、汎用性の高い方式としてHaarウェブレットで画像特徴を算出し、サポートベクターマシン(SVM: Support Vector Machine)[57]と呼ばれる機械学習により画像内から人物を検出する方式[58][59]が提案されている。この歩行者検出では、連続画像内の動きの不連続部分に検出処理を行うと検出率が向上する実験結果が得られている[60]。そこで、歩行者の移動で発生するフレーム間変化を画像特徴とし、AdaBoostと呼ばれる機械学習で歩行者を検出する方式[61]が近年提案されている。この歩行者検出では、処理の高速化のために段階的に人物領域を絞り込むカスケード方式[62]が利用されている。

このような機械学習を用いる移動体検出方式は、検出対象ごとに事例画像を収集し、事前に機械学習を行うことで様々な物体を検出対象にできる利点がある。しかし、それらの従来方式は画像全体を探索する必要があるため、実時間で動作しない問題がある。また、既存の成果はいずれも固定監視カメラが前提であり、監視カメラが旋回する場合などに利用できない問題がある。

## 1.4.2 姿勢認識と人物認識

機械学習は前述の人物検出だけでなく、検出人物の姿勢や向き認識、あるいは検出人物の個別認識などができる可能性のある汎用的な技術である。そこで、ここでは機械学習の概要と、検出人物の認識と姿勢認識の研究動向を示す。

### (1) 機械学習による認識

機械学習によるパターン認識を統計的決定理論から見ると、古典的な技法であるベイズ識別と見なせる。ここで、あるパターンの母集団を $\{\omega_1, \omega_2, \dots, \omega_n\}$ とし、観測されたパターンをベクトル $\mathbf{x}$ で表すと、 $\mathbf{x}$ が $\omega_i$ から生じた確率は $P(\omega_i | \mathbf{x})$ である。つまり、 $\mathbf{x}$ が観測された時に確率 $P(\omega_i | \mathbf{x})$ が最大になる $\omega_i$ を見つける必要がある。しかし、この確率は直接求められないのでベイズの定理により

$$P(\omega_i | \mathbf{x}) = \frac{P(\omega_i)P(\mathbf{x} | \omega_i)}{P(\mathbf{x})}$$

と変形し、分子 $P(\omega_i)P(\mathbf{x} | \omega_i)$ の最大値を求める問題とすることが一般的である。この分子は、例えば手書き数字の「0」「1」～「9」のような母集団 $\omega_i$ から多数のサンプル $\mathbf{x}$ を取り出し観測することで実験的に推定できる場合が多い。

画像監視システムを作成する場合、一度設定した監視カメラは数年以上に渡り長期運用されるので、その映像を全て蓄積しておけば十分な母集団 $\omega_i$ とその分布状態を準備できる可能性もある。しかし、実際の画像監視システムの作成/運用では、一箇所の設定を短期間で終了する必要がある、母集団 $\omega_i$ の数やその分布状態は未知である場合がほとんどである。実際、ある監視現場用の画像監視システムの作成では、限られた時間に撮影した事例画像から母集団 $\omega_i$ の数や分布を決定しなければならない問題に頻繁に直面する。このように、母集団 $\omega_i$ が既知で、そのサンプル $\mathbf{x}$ が十分に観測でき、その分布が既知である理想的な状況が得にくいことが画像監視システム作成時の大きな制約となっている。

母集団の確率分布の知識を用いないパターン認識方式として、パターン分布ではなく、パターン同士の境界を分離に利用する機械学習方式が提案されている。その代表として、数理的にも明快でかつVapnik-Chervonenkis理論[57]でその理論的な背景を得たSVM (Support Vector Machine)がある。このSVMは線形の識別器であるが、Mercerカーネルを使うことで非線形に拡張され、近年、文字認識や画像認識で高い認識率を上げているパターン認識方式である。ただし、母集団 $\omega_i$ に関しては事前に決定できないため、SVMなどでも撮影環境と監視目的に合わせた特定の監視対象を利用者が設定し、監視対象のサンプル画像を利用者が与える必要がある。

### (2) 姿勢と人物の認識

監視画像内の人物や車両を検出しその種別を分類する研究としてカーネギーメロン大学のVSAM (System for Video Surveillance and Monitoring)プロジェクト[63]や、マサチューセッツ工科大学AIラボのForest of Sensorsプロジェクト[64]、あるいは人物監視に特化した画像監視システムW<sup>4</sup>[65]などがある。より詳細に画像内の人物の姿勢を調べ、Computer Graphics (CG)で表現する研究成果として、腕、足、腰等の間接を持つ身体モデルを画像内

の人物に当てはめ、動作を正確にCG化する研究が数多く進められている [66][67][68]。しかし、機械学習による人物検出[58][59]と同様の汎用的な枠組みで姿勢を認識する研究成果[69]は多くない。これは、監視カメラの設置位置や監視場所により人物の動作や姿勢の見え方が異なるため、一般化が困難なことに起因している。

また、人物認識では顔/虹彩/指紋/静脈など、個人ごとに明確な特徴を用いた個人認証技術が近年産業界で広く利用されだしている。しかし、監視画像から検出した全身画像から個人を特定する方式は、個人を特定するための安定した特徴が検出困難なこともあり、研究成果が得られていない。近年、監視画像で得られる人物の歩き方を特徴として人物を特定する研究[56]が進められているが、監視画像から検出した全身画像を使い機械学習で人物を特定することは困難な課題である。

### 1.4.3 機械学習を用いた画像監視の処理時間

機械学習を利用した汎用性のある人物検出方式[58][59]は、画像内を探索する必要から1画像の処理に数分程度かかる。そこで高速化のために、歩行者の移動で発生するフレーム間変化を画像特徴とし、カスケード[62]と呼ばれる段階的に探索範囲を絞り込むことで歩行者を検出する方式[61]が近年提案されている。この高速化方式では、固定カメラで撮影した360×240画像に対し、Pentium4の2.8GHzの計算機で4fps (frame per second) の処理速度が得られている。また、画像内の局所変化を3特徴(「空間平均した動きの強さ」、「時間平均した動きの強さ」、「時間的な動きの一様性」)で表現し、歩行者と背景を分離する方式[42]が提案されている。その提案方式ではPentium3, 933 MHzの計算機 (Matrox Meteorキャプチャーカード) を使い320×240画像を7.5 fpsで処理できている。

目視監視に代わる画像監視システムとして、ビデオレート(30 fps)で監視できることが望ましいが、今のところ機械学習を利用した方式では4~7.5 fps程の処理速度である。また、いずれも固定した監視カメラを前提としており、監視カメラの旋回中は歩行者を検出できない課題がある。

表 1-5 電気事業で活用されている画像監視システム例 (文献[121]を参考としている。)

| 利用先     | 監視対象  |
|---------|---|
| 火力発電所監視 | 侵入監視、煤煙監視、取水口監視、港湾監視、貯炭場監視                                    |
| 水力発電所監視 | 侵入監視、屋外変電設備監視、配電盤室監視、水車発電設備監視                                 |
| ダム監視    | 侵入監視、ダムゲート監視  |
| 変電所監視   | 侵入監視、屋外変電設備監視、配電盤室監視、母線監視                                     |
| 送電線監視   | 侵入監視、工事監視、気象監視  |
| 無線中継所監視 | 侵入監視、通信機器監視、空中線設備監視、気象状況監視、周囲環境監視                             |
| その他     | 管理区域への入退所監視、門扉監視、駐車場監視、現場作業監視、作業員監視、漏水や火災の監視、天候状況の監視、メータ読取りなど |

## 1.5 研究の位置付けとアプローチ

画像監視システムに対する社会的な要求と技術課題を示し、本論文で示す研究の位置付けと目標、ならびに目標に対するアプローチを示す。

### 1.5.1 社会的な要求と技術的な課題

現在、電気事業で運用している画像監視システムを例に取上げ、画像監視システムに対する社会的なニーズとそのニーズを実現するために解決しなければならない技術課題を示す。

#### (1) 社会的な要求

電気事業では設備の稼働状態の把握や、電力設備の警備、台風や積雪などの気象変化による現場状況の把握、非常災害時の情報収集など多彩な目的のために画像監視システムを利用している。特に重要な設備の監視には、地上通信の遮断を考慮し通信衛星をバックアップ手段とする画像監視システムも作成されている。2000年に(社)電気学会がとりまとめた資料[121]から、電力事業で用いている代表的な画像監視システム例を表1-5に示す[121]。2004年の(社)電気協同研究会による国内主要電力会社に対する画像監視システムの活用調査でも表1-5と同様の結果を得ている。いずれの調査結果でも、現場の電力設備の状態監視が利用目的のトップであり、次に侵入監視が上がっている。設備監視の対象は部門ごとに異なるが、侵入監視はいずれも人物、車両、動物など、共通した監視対象であり、その中でも人物が最も監視現場でニーズの高い監視対象となっている。

電気事業で侵入者が監視対象になる監視個所は、屋内、電気所構内、電気所構外の3個所に大分類できる。室内や構内では、画像監視システムに組み込まれている赤外センサなどを使い人物の監視領域への立ち入りを監視している。ただし、これらのセンサは草木の揺れや鳥などにも反応するため誤報も多く、センサが反応するたびに監視画像を目視確認しなければならない問題に直面している。また、センサが動作した理由を画面で確認できない場合、現場に直行し原因を調べる必要があり、監視員に負荷を強いている。一方、河川等の電気所構外の監視は、監視領域が遠方におよびセンサ等を設置できないため、監視員の目視監視に依存している状態である。

近年、無線ICタグと赤外センサの組み合わせで、入構証あるいは社員証にICタグを組み込み、ゲートの通過を自動監視することが可能となった。しかし、主要変電所構内などでは携帯電話や他の電波発信源の使用を禁じているため、構内への無線ICタグ基地局の設置は今のところ時期尚早である。たとえ、電気所構内の入口に設置しても、ICタグを持たずに入ろうとした物体が人物なのか動物なのかを判断する目視監視が不可欠な状況である。

#### (2) 技術課題

侵入者の監視場所を室内、電気所構内、構外に分類し、各監視場所の特性と画像処理による誤検知や未検知を引き起す環境要因を表1-6に示す。

表 1-6 監視箇所の特性と誤検知を引き起す環境変化

|         |    | 室内<br>(人物サイズ: 大)   | 電気所構内<br>(人物サイズ: 中)   | 構外<br>(人物サイズ: 中・小)   |
|---------|----|--|---|--|
| 監視箇所の特性 |    | 照明変化や背景変動が少ない。室内や出入口などに監視箇所を特定できる。   | 照明変化や背景変動が大きい、監視箇所を特定できる。また、壁や道路などの人工物や、花壇などのように手入れの行き届いた場所が背景となる。  | 照明変化や背景変動が大きく、監視領域が遠方までおよぶ。また、ダム周辺のように自然の木々や水面のゆれなど変動の激しい場所が背景となる。   |
| 環境変化    | 照明 | <ul style="list-style-type: none"> <li>・窓からの日差し</li> <li>・室内の照明変化</li> </ul> | <ul style="list-style-type: none"> <li>・太陽光、影、反射光の変化による輝度/色変化</li> <li>・背景変動や車両等の移動物体の反射光による輝度/色変化</li> <li>・雨、雪、霧などによる輝度/色変化</li> </ul>  |  |
|         | 背景 | 変形/移動する機器や設備、他の人物  | <ul style="list-style-type: none"> <li>・車などの動く物体とその影</li> <li>・形状が変わる設備とその影</li> <li>・風による木々や草のゆれ</li> <li>・人工水路内の波や水面の映り込み</li> <li>・雨、雪、霧等によるノイズ</li> <li>・他の歩行者など</li> </ul> | <ul style="list-style-type: none"> <li>・車などの動く物体とその影</li> <li>・形状が変わる設備とその影</li> <li>・風による電線等の設備のゆれ</li> <li>・風による自然の木々や草のゆれ</li> <li>・河川の波、水面の映り込みや流木等のゴミ</li> <li>・陽炎等の揺らぎ</li> <li>・雨、雪、霧等によるノイズ</li> </ul> |

照明が一定の屋内であれば、人物の動きで発生する変化を背景差分やフレーム間差分、あるいは相関を利用した差分処理で人物を容易に検出できる。あるいは、画像内の変化領域だけに機械学習の人物判定を行うことで、高速な人物検知ができる。しかし、検出人物が社員か侵入者かを判断し、その動作内容に応じた警報を出すには人物や姿勢の認識が必要になるが、画像処理による人物や姿勢の認識は依然として困難な技術課題である。

一方、電気所構内などの屋外では、背景や明るさなどが常に変化するため安定した人物検出が難しく、現在、実時間で動作する人物検出方式の研究が続けられている状態である。特に、画像から機械学習により人物を選択的に検出する汎用的な方式では、検出処理速度が4~7.5fpsと遅くビデオレートでの処理が不可能な状態である。

さらに、電気所構外の監視では、ダム周辺の河川の監視など監視領域が遠方におよび、その背景も自然の川や森、住宅街など様々で、かつ監視対象がカメラの近傍にいるか遠方にいるかでサイズが大きく変化する。そのため、室内や構内に比べ目視でも監視が困難で、画像処理による人物検知が困難な状態である。

### 1.5.2 参考とする視覚処理

脳細胞は100スパイク/秒程度で発火しうる[20]ので、計算機にたとえると100Hz程度の処理速度である。今日の格安ノートパソコンでもそのCPU速度が数GHzであり、明らかに処理速度はパソコンのほうが上である。しかし、100億個以上の細胞があるといわれている脳では、1個の細胞が100個から1,000個の細胞に影響を及ぼし[20]、それらの細胞の階層構造と機能の分化により、パソコンによる画像処理以上のシーン理解を可能にしている。そこで、本論文では画像監視システム用の画像処理の考案に、以下の視覚野の構成と機能を参考にする。

人の視覚処理は光を受光する網膜レベルから動き処理経路と形状や色の認識経路の2経路に処理が分かれている。動き処理経路は認識経路よりも画像の転送速度や処理速度が速い。また、2経路の各階層間で処理の結びつきがある。なお、2経路間の接続の具体的な役割は今後の解明が必要な状態だが、処理速度の速い動き処理経路の解析結果が認識処理経路で使われていると容易に仮定できる。

動きを高速に捕らえ反射的に行動することは、生命を維持する上でも重要な行動であり、認識よりも動きの処理が速いことが自然である。そこで、本論文では下記に示す画像処理の諸課題を、動きによる前処理と色や形状による認識の後処理を組み合わせる方式を検討する。

### 1.5.3 研究の位置付け

上述のように電力設備や石油プラントなどで設備の稼働状態の把握や、保守保全および警備の観点から、画像監視システムが数多く導入されている。さらに近年、ネットワーク技術の進展や監視カメラの低コスト化などが背景となり、監視カメラの設置台数が増える傾向にある。画像監視システムは各種の装置で構成されているが、唯一自動化できていない部分が監視員による目視部分である。多くの監視カメラの画像を監視員が見続けることが不可能なため、表 1-4 に示すような検出・認識・追跡機能を持つ画像監視システムの開発が望まれている。

本論文では、その中でも屋内や電気所構内の画像監視システムに対し、画像処理により目視を補助あるいは代行する下記の検出機能と認識機能の実現を目標とする。

- ・監視領域に入る移動中の人物を選択的に実時間で検出する機能
- ・検出人物の動作に応じて警報を出すための姿勢認識機能
- ・検出人物の個別認識機能

監視領域に入る移動体として、人物、車両、動物、自転車等があるが、監視対象になる頻度が高く監視ニーズの高い人物を本論文の監視対象として議論する。ただし、検出や認識に機械学習を利用し、学習画像を変更することで人物以外の物体にも拡張が容易な方式を研究する。

検出と認識の課題として以下の3つがあげられる。第1に、汎用的なSVMなどの機械学習を使い環境変化に強い歩行者検出方式が提案されているが、従来方式はビデオレートでの実時間処理ができない。第2に、検出した人物の動作に応じて警報を出すには、人物の姿勢や向きの認識機能が必要となるが、人物の様々な姿勢を定義することが困難なため、上述の人物検出と同様な枠組みによる姿勢認識は行われていない。第3に、検出した人物が既知の人物なのか、一度撮影した人物なのかなどの個人の認識機能を持つ必要がある。しかし、監視カメラで撮影した全身画像は人物の服装などが日ごとに変化するため、安定した認識特徴が得られないことから研究が行われておらず、その実現は一般に困難と考えられている。

上記の課題に対し本論文では、目視監視を代行する画像監視システムの開発に向け、監視カメラに写る人物の全身画像を使い、人物の実時間検出、検出人物の姿勢認識、検出人物の認識を実現する方式を提案する。いずれの方式も、視覚処理を参考に動きによる人物領

域の絞り込みを先に行い、その後、SVMなどの機械学習を用いた検出/認識を行う段階的な処理としている。なお、3章から5章では、監視環境と監視対象を限定した評価可能な具体的な問題設定を行い、対処方式を議論する。また、実用化の観点から、特殊な装置を使うことなく、市販の計算機と監視カメラで実時間動作する方式を議論の前提とする。

#### 1.5.4 人物の高速な検出

従来のSVMなどの機械学習による人物検出方式は、画像内から人物を選択的に検出できるが、画像内を検索するため実時間(30fps)で処理ができない。そこで本論文では、従来の機械学習による人物検出方式を使い実時間で人物を検出するため、固定監視カメラ用と旋回監視カメラ用に2通りの前処理を提案する。固定カメラ用にライン型前処理を提案し、旋回カメラ用に側抑制型前処理を提案する。

##### (1) 固定監視カメラ用のライン型前処理

固定カメラ用のライン型前処理は、監視領域内に監視ラインを設定し、監視ライン上で動きの速い部分に優先順位をつけ、優先順位の上位を人物検出の探査範囲とする方式である。一度人物ではないと判定された部分は、一定時間のあいだ人物検出の探査範囲として選択しない方式とし、処理の高速化を図る。

##### (2) 旋回監視カメラ用の側抑制型前処理

旋回監視カメラ用の側抑制型前処理は、移動中の人物が周囲の背景と異なる動きをすることに着目し、周囲と異なる動き部分を人物検出の探査範囲とする方式である。つまり、周囲と同じ方向に動く部分の選択を抑え、周囲と異なる方向に動く部分を優先して探査範囲に選択する方式となっている。

いずれの前処理も、既存の機械学習による人物検出と組み合わせることで、実時間での人物検出が可能であることを実験結果で示す。

本技術は、従来、汎用的な機械学習で不可能であった実時間での侵入者検知を可能とする技術であり、既存の画像監視システムの高度化に直接役立つ成果である。

#### 1.5.5 姿勢の認識

画像内の人物の動作を把握するには、検出した人物の姿勢を認識する必要があり、全身画像に人物の形状モデルを当てはめる研究成果が数多く報告されている。しかし、様々な姿勢を定義することが困難なため、SVMの人物検出と同様な枠組みで姿勢を認識する研究は行われていない。

そこで、本論文では代表的な作業姿勢が定義できる発電所運転時の姿勢を監視カメラで撮影し、その操作姿勢を人物検出と同じ枠組みで認識する方式を議論する。発電所運転特有の4種類の姿勢を学習し、個々の作業姿勢を学習したSVMを動きの状態に動的に切替えて認識する動的切替方式を示す。提案方式により、人物の向きや姿勢のばらつきなどを吸収した姿勢認識ができることを実験で示す。

提案方式は、通常の運転から逸脱した運転操作時に警報を出す画像監視システムの作成等を可能にする成果である。

### 1.5.6 人物の認識

顔、指紋、虹彩、静脈などを利用した個人認証に比べ、監視画像から検出した人物は、形状や服装がたえず変化し、撮影する高さや角度により見え方が大きく異なるため、人物を特定する安定した特徴が得られにくい。そのため、監視カメラで撮影した全身画像から人物を特定することが難しく、SVMなどの機械学習を用いた全身画像による人物認識は行われていない。

そこで、本論文では研究室内に出入する既知人物8人の16日間の全身画像(約1,120画像)を使い、SVMによる人物認識を検討する。ただし、理論的に議論できるほど問題の特性を定式化できないため、実験結果の観点から検討を行う。実験ではSVMを段階的に切替えて8人を認識する多段戦略と、4種類の画像特徴を用いた認識実験を行い、同一撮影日の全身画像を学習すると95%~98%程度の個人認識率が得られたことを示す。また、100物体が写る7,200画像による認識実験を行い、より定量的な検証結果を示す。特に考察では、SVMの多段戦略よりも、利用する画像特徴に認識精度が依存することを示す。

本結果は限定的なものだが、監視カメラで撮影した全身画像から、一度撮影した人物を個別に特定する機能の開発や、社員服を着た人物と侵入者を区別する機能の開発に貢献する成果である。

## 1.6 論文の構成

画像監視システムを実現するための画像処理の関連研究を2章で示し、監視カメラの画像内から実時間で人物を検出する方式を3章で示す。3章の提案方式で屋内・屋外から安定した人物検出ができるため、4章では検出人物の動作を把握するための姿勢認識方式を示す。また、5章では検出人物の個別認識方式を示す。6章では本論文の全体的なまとめと課題、今後の展望を示す。以下、2章から5章の概要を示す。

2章では画像監視システムの既存研究、ならびに本研究で利用する機械学習を示す。既存研究として、画像内の变化領域の検出、検出物体の認識、監視の処理速度に関連した研究事例を中心に示す。また、2章の後半で、本研究で利用する機械学習の概要を述べ、参考となる脳の視覚処理の概要を示す。

3章では屋外の電力設備の監視問題を設定し、実時間で人物を検出する方式を提案する。固定監視カメラ用に考案したライン型前処理と、旋回カメラ用に考案した側抑制型前処理を示し、従来のSVMなどの機械学習との連携で、実時間で人物検出ができることを実験結果で示す。

実時間で人物を検出する方式を3章で提案するため、4章では検出した人物の姿勢認識方式を示す。姿勢認識に対し認識精度が議論できる原子力発電所の運転員の姿勢認識問題を設定し、監視カメラ画像から検出した人物の全身画像を使い、作業姿勢を認識する方式を示す。評価実験では、監視カメラで撮影した1時間30分の実験画像で95%の姿勢認識率が得られたことを示す。

5章では、監視カメラ画像から検出した人物の全身画像を使い、既知人物の認識方式を

提案する。具体的な問題として、研究室内のコーヒーマシンを利用する研究室メンバーを認識する問題を設定し、監視画像から検出した全身画像を使い個人を特定する方式を議論する。評価実験では、8人の16日間の監視画像を用いた実験結果と、既存の研究用画像データを利用した実験結果を示す。

6章では本論文を総括するとともに、提案方式に関する今後の課題を示す。最後に、各種センサと画像処理の連携による高度な画像監視システムの将来像を示し本論文を締めくくる。

## 1.7 本章のまとめ

本章では、画像監視システムの構成と機能を示し、システムの中で唯一自動化できていない部分が、監視員による目視機能であることを示した。画像処理の歴史的な考察により、監視員による目視機能を画像処理で置換る問題は、シーン理解の研究としてデジタル画像処理の黎明期から議論されている困難な課題であることを示した。その後、電気事業で利用している画像監視システムを参考に、監視現場で望まれている具体的な機能と、その機能を画像処理で実現する上での諸課題を示した。それらの諸課題の中で、社会的要求が強いが未だ実現できていない3課題、①人物の選択的な実時間検出、②検出人物の全身画像による姿勢認識、③検出人物の全身画像による個人認識を本論文の研究対象とすることを示した。最後に研究のアプローチとして、視覚処理を参考に、動きによる人物領域の絞り込みを先に行い、その後 SVM などの機械学習を用いた検出/認識を行う段階的な処理で、3課題の解決を目指すことを示した。最後に本論文の構成を示した。

## 第 2 章

### 画像処理による侵入監視の関連研究

#### 2.1 はじめに

画像監視システムの中で、自動化できていない部分が監視員による目視機能である。このような目視機能を画像処理で実現することが、シーン理解の研究として長年に渡り検討されてきたが、未だに困難な課題である。画像から移動中の人物を検出・認識する観点から従来の代表的な画像処理を概観すると、図2-1に示すような関係が描ける。はじめに、画像処理の各種フィルタにより画像特徴を抽出し、人物検出や認識に適した特徴（点線・動き・色・テクスチャ・3次元距離など）を組み合わせ、目的を達成する画像処理プログラムを作成していると見なせる。

画像監視システム用の画像処理では、人物の検出と認識の精度、ならびに処理速度が重要となる。そこで、本章では2.2節で移動中の人物を検出する観点から代表的な関連研究を紹介し、その技術課題を示す。2.3節では人物と姿勢の認識に関する関連研究とその技術課題を示し、2.4節では既存の人物検出処理速度の現状と課題を示す。さらに、2.5節では本研究で利用する機械学習を、その理論的な限界を中心に説明する。最後に、2.6節で目視を補助あるいは代行する画像処理の設計に参考となる視覚野の構造と機能を紹介する。

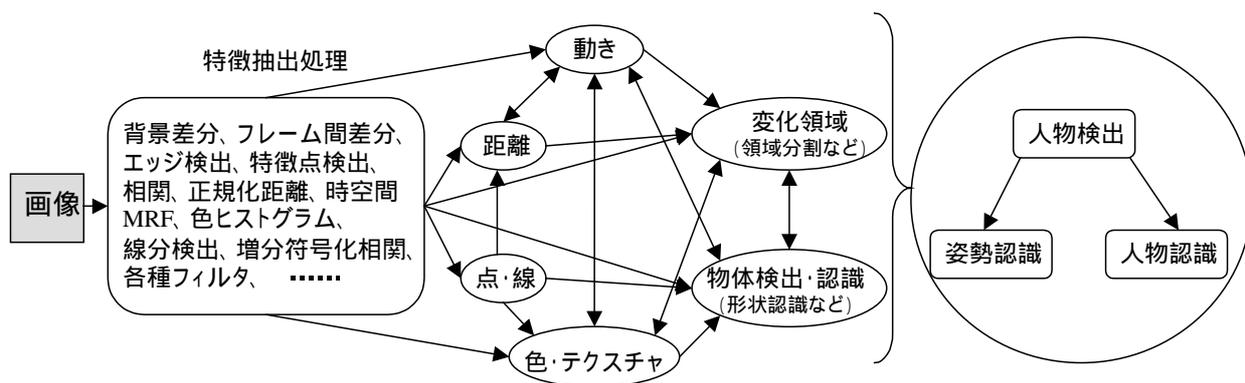


図 2-1 人物検出や認識の画像処理概要

## 2.2 移動中の人物検出

監視カメラ画像から歩行者を検出する代表的な3通りの従来方式（①背景差分やフレーム間差分などの画像差分を利用する方式、②画像内の色や動きを利用する方式、③パターン認識を使い人物形状だけを検出する方式）を示す。

### 2.2.1 画像差分による移動中の人物検出

#### (1) 背景差分とフレーム間差分

静止背景を人物が移動する場合、背景差分やフレーム間差分で検出した画像内の変化領域が歩行者などの移動体候補領域である。例えば、最も一般的な背景差分とフレーム間差分による変化領域の検出は次のようになる。

時刻 $t$ の画像 $I_t$ 内の座標 $(x,y)$ での画素値 $I_t(x,y)$ と、各座標の画素値を一定時間平均した画像 $\mu$ 、例えば時刻 $t-1$ まで平均した画像を $\mu_{t-1}(x,y)$ とすると、背景差分による変化領域 $B_t(x,y)$ を

$$B_t(x,y) = \begin{cases} 1 & \text{if } |\mu_{t-1}(x,y) - I_t(x,y)| \geq T_t \\ 0 & \text{else} \end{cases}$$

として求めることができる。ここで、 $T_t$ は閾値である。フレーム間差分は連続した画像 $I_{t-1}$ と $I_t$ を使い

$$E_t(x,y) = \begin{cases} 1 & \text{if } |I_t(x,y) - I_{t-1}(x,y)| \geq T_t \\ 0 & \text{else} \end{cases}$$

として変化領域 $E_t(x,y)$ を検出できる。閾値 $T_t$ の一例として、背景画像などの標準偏差 $\sigma_t(x,y)$ を使い、

$$T_t(x,y) = \alpha \sigma_{t-1}(x,y)$$

としたり、時間的な経過を反映するために重み係数 $\alpha$ を調整して

$$T_t(x,y) = \alpha T_t(x,y) + (1 - \alpha) T_{t-1}(x,y)$$

とするなどの設定が用いられている。あるいは、画像内の輝度変化に応じ重み係数 $\alpha$ や閾値 $T$ の値を条件分岐で変化させる方式なども利用されている。

照明変化のない室内で動く物体が人物だけであれば、差分処理で検出した変化領域が人物領域となる。ただし、差分処理では人物の影部分も検出する問題があり、影を削除する研究が現在も続けられている。

#### (2) 閾値の自動設定

画像差分では閾値の決定が難しい問題となる。閾値の自動的決定方法として、画像内のヒストグラムを背景と歩行者の2クラスに分割することを想定し、

$$\sigma_B^2(k) = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2$$

を最大にする $k$ を閾値に選ぶ「大津の閾値選定方法」[31]を変化領域検出に用いる方式[32]

などが広く利用されている。ここで、 $\omega_0, \omega_1, \mu_0, \mu_1, \mu_T$ は、画素数 $N$ 、画像の階調数 $L$ 、 $n$ をヒストグラムのレベル $i$ の画素数とし、その出現確率を $p_i=n_i/N$ としたときに次式で表現される値である。

$$\begin{aligned}\omega_0 &= \sum_{i=0}^{k-1} P_i, & \omega_1 &= \sum_{i=0}^{L-1} P_i \\ \mu_T &= \sum_{i=0}^{L-1} iP_i, & \mu_0 &= \sum_{i=0}^{k-1} \frac{iP_i}{\omega_0}, & \mu_1 &= \sum_{i=0}^{L-1} \frac{iP_i}{\omega_1}\end{aligned}$$

この閾値 $k$ は最小二乗近似の意味で画像の最適な2値化になっている[2]。あるいは背景画像の上位 $M$ 番目までの固有値を使い画像内の雑音変化を無視し移動体だけを検出する

$$D_t(x, y) = \begin{cases} 1 & \text{if } |I_t(x, y) - \Phi(I_t(x, y) - \mu(x, y))| \geq T_t \\ 0 & \text{else} \end{cases}$$

なども提案[34]されている。ここで、 $\phi$ は背景画像を多数利用して作成した上位 $M$ までの固有値ベクトルの行列であり、 $\mu$ は画像平均である（詳細は文献[34]を参照）。あるいは1年以上の長期間に渡る輝度値の統計的な分布を使用した方式[30]などが提案されている。

### (3) 照明変化への対応

差分処理は照度変化に弱いため、照明変化への対応を目的に画像の輝度値を正規化した「正規化距離」による変化検出方式が提案されている[35]。これは、ある時刻 $t$ での画像の着目領域の各画素値を基底ベクトル $v_1, v_2, \dots, v_n$ とする空間を作り、 $V_t$ を着目領域内の各画素値で構成されるベクトルと見なし、 $V_t$ の単位球上への射影を $P_t$ とする。もし次画像 $t+1$ で画像内の構造的変化がなければ着目領域のベクトル $V_{t+1}$ の単位球上で射影 $P_{t+1}$ は前の $P_t$ の方向に近いが、物体が入り込み着目領域内に構造的な変化が起きると $P_{t+1}$ と $P_t$ の方向が大きくずれることになる。そこで、単位球に射影した $P_{t+1}$ と $P_t$ の距離 $d$ を、

$$d(V_t, V_{t+1}) = \left| \frac{V_t}{|V_t|} - \frac{V_{t+1}}{|V_{t+1}|} \right|$$

として求め「正規化距離」と呼ぶ。この正規化距離が閾値以上であれば侵入物体によるものと判断し、閾値以下であれば照度変化によるものと判断する方式である。

「正規化距離」を使わなくとも、一般的に知られている相関係数を利用して照明変化に対応した変化検出ができる。例えば、連続する2画像の同一領域 $f$ と $g$ の画像相関（相関係数） $r$ は

$$r = \frac{\sigma_{fg}}{\sigma_f \sigma_g}$$

と表現できる。ただし、2画像の領域 $f$ と $g$ の輝度値の平均 $\mu_f, \mu_g$ と標準偏差 $\sigma_f, \sigma_g$ を、領域サイズを $M \times N$ とし、

$$\sigma_f = \left[ \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \{f(x, y)\}^2 - \mu_f^2 \right]^{1/2}$$

$$\sigma_g = \left[ \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \{g(x, y)\}^2 - \mu_g^2 \right]^{1/2}$$

$$\mu_f = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)$$

$$\mu_g = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x, y)$$

として求め、 $\sigma_{fg}$  を

$$\sigma_{fg} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)g(x, y) - \mu_f \mu_g$$

とする。画像が濃淡階調で線形変換を起こしている2画像間で、画像相関は常に普遍的な値となるため、構造変化が無く全体的に照度に変化した状態では2画像の相関 $r$ が高くなる。この相関係数値の一致度により、画像内の構造的変化を判断できる。

さらに、この相関をよりロバストにしたものが、画像内の隣接画素の明度増分だけを利用する「増分符号化相関法」[37]であり、

$$r = \frac{1}{(M-1) \cdot (N-1)} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \{b_t(x, y) \cdot b_{t-1}(x, y) + (1 - b_t(x, y)) \cdot (1 - b_{t-1}(x, y))\}$$

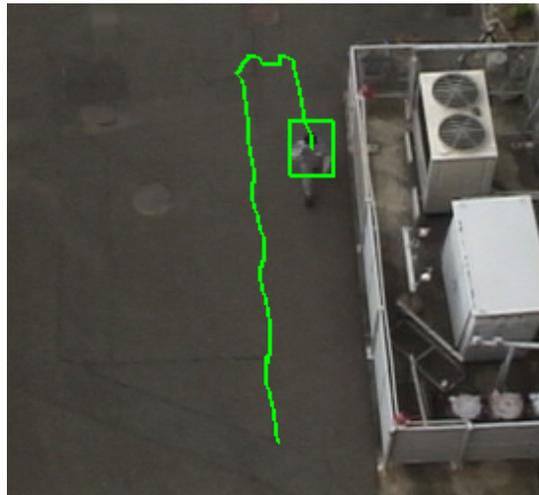
と定義される。ここで $b_t$ は画素の明度増分の記録であり、

$$b_t(x, y) = \begin{cases} 1 & \text{if } I_t(x+1, y) \geq I_t(x, y) \\ 0 & \text{else} \end{cases}$$

である。画像内の構造的な変化が発生し、隣接画素同士の値の大小が入れ替わった部分だけを検出するため、画像相関よりもロバスト性が増す場合がある。

さらに、差分では雑音等の影響で移動体領域を正確に検出できないため、背景差分画像 $B$ と新規画像 $I$ のそれぞれのエッジ画像 $B'$ ,  $I'$ を作成し、 $B'$ と $I'$ の論理関積により移動体エッジを検出する方式[38]なども利用されている。差分で大まかに移動体領域を決定したあとに、物体境界の分離や結合などの変化に追従する動的輪郭モデルで輝度境界を成長させ移動体を検出する方式[40]なども提案されている。画像内の小領域の固有値を使い、隣接する領域の相関の強さ（画像変化の共起性）で木々の揺れなどを除外する方式[41]や、歩行者の直線的な動きと背景の動きを空間と時間軸上の特徴として分離する方式[42]なども提案されている。

照明変化や影の影響を廃除するために様々なバリエーションが提案されているが、一般的に画素値の時間的な変化を利用する方式が多い。一例として、背景差分とフレーム間差分を利用して変化領域を検出し、その変化領域の移動軌跡を表示した例を図2-2に示す。図



画像差分を利用して検出した変化領域と、  
検出した変化領域の移動軌跡を表示した結果

図 2-2 変化領域の検出例

2-2に示すように人物だけが移動している画像では、差分処理で人物検出が可能である。ただし、画像差分による方式はいずれも静止したカメラが前提となっている。そのため、移動中のカメラで撮影した画像は背景が連続的に変化するため適用できない問題がある。また、背景が動き続けているような場合も同様である。なお、画像内に人物以外の動きが含まれる場合、検出した領域が人物による変化かその他の物体による変化なのかを判定する後処理が不可欠である。

## 2.2.2 動きや色などによる人物検出

画像内の動きや色、テクスチャなどを用いた変化領域の検出方式が数多く提案されている。以下では、動きを利用した変化領域の検出と、動きと色などの他の特徴を組み合わせた変化領域検出方式を示す。ただし画像差分と同様、検出した領域が人物であるかどうかを判断する後処理がいずれも必要な方式である。

### (1) 動きの検出

画像内の動きから移動体などの変化領域を検出する方式が広く利用されている。画像内の移動体を検出するために、始めにオプティカルフローとして画像内の動きベクトルの分布を求めることが一般的である。オプティカルフローを求める方式は、以下のブロックマッチング法と勾配法の2つの方式が広く利用されている。

ブロックマッチングは $p \times q$ 画素のブロックのテンプレートをそれよりも大きい $P \times Q$ 画素上の探索範囲内 $((P-p+1) \times (Q-q+1))$ で動かし、類似度が最大になる部分を探す方式である。類似度の判定には、画素の2乗差(SSD: Sum of Square Differences)や前述の相関計数などが利用されている。類似度にどのような関数を使うか、どのような計算打ち切り基準を作るかなど、多くの研究成果がデジタル画像処理技術の黎明期に報告されている。

一方、1980年代にHornらにより時空間勾配法が定式化[16][79]され、現在、ブロックマッチング法とともに広く利用されている。勾配法は以下のように表現できる[71]。時刻  $t$  での画素  $(x,y)$  の輝度値を  $I(x,y,t)$  とし、微小時間  $\Delta t$  の間に縦横方向に物体が  $\Delta x, \Delta y$  移動したとすると、 $I(x,y,t)$  とその微小時間後の  $I(x+\Delta x, y+\Delta y, t+\Delta t)$  の関係は

$$I(x,y,t) = I(x+\Delta x, y+\Delta y, t+\Delta t)$$

となる。右辺をテイラー展開し高次の項を切り捨てると

$$I(x,y,t) = I(x,y,t) + \Delta x \frac{\partial I}{\partial x} + \Delta y \frac{\partial I}{\partial y} + \Delta t \frac{\partial I}{\partial t}$$

となる。この式を整理すると、

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

となり、速度ベクトル  $dx/dt, dy/dt$  を  $u, v$  とし、 $\partial I/\partial x, \partial I/\partial y, \partial I/\partial t$  を  $I_x, I_y, I_t$  とおくと、

$$I_x \cdot u + I_y \cdot v + I_t = 0$$

と表現できる。この式はオプティカルフローの拘束式と呼ばれ、画像の局所的な観測から得られるものである。誤った対応を少なくするため、画面が全体に渡って滑らかに変化すると仮定し、画像全体の速度変化  $E_s$  を小さくしながら拘束式の誤差  $E_c$  が小さくなるように、

$$E = E_s + \lambda E_c$$

の最小化問題として  $u$  と  $v$  を求める。ここで、 $\lambda$  は重みパラメータであり、 $E_s$  と  $E_c$  はそれぞれ下記の式で表現できる。

$$E_s = \sum_x \sum_y ((u_x^2 + u_y^2) + (v_x^2 + v_y^2))$$

$$E_c = \sum_x \sum_y (I_x u + I_y v + I_t)^2$$

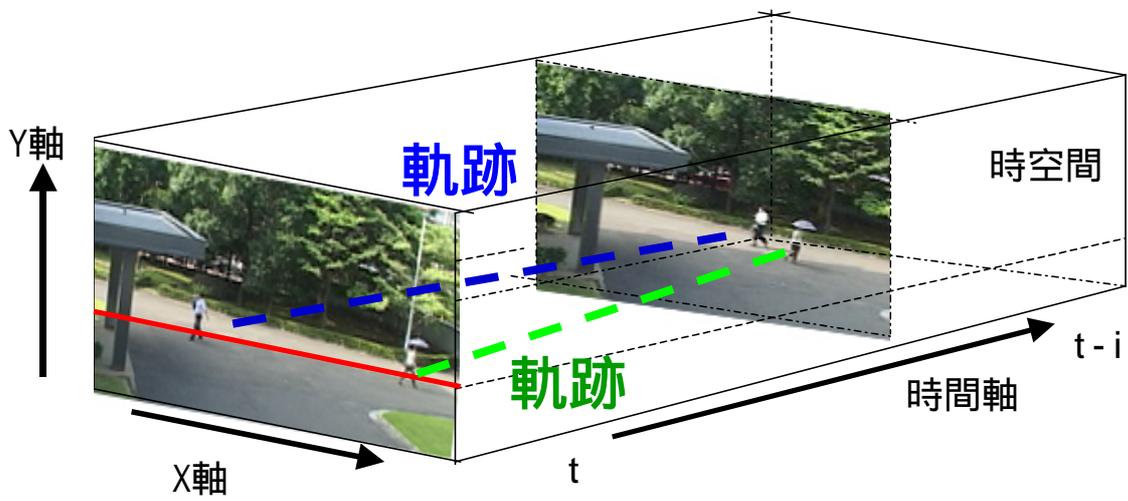
$E$  の最小化は変分法を用いて、繰り返し計算で最適解を算出できる。

より正確に動きを求める方法として、時間軸での変化を記録した時空間画像から動きを検出する方法がある。例として、図2-3aに連続画像の1画像を示し、図2-3bに連続画像を3次元的に並べ  $y$  軸に垂直に切断した様子を示す。図2-3cは切断面 ( $x-t$  軸断面) を画面の上側に表示した結果である。図2-3cより静止物体は  $x-t$  軸断面で垂直に上方向に伸びているが、歩行者の断面は斜めの軌跡を示している。図2-3cでは、歩行者の速度が速い方が傾きの大きな斜め線 (軌跡) となっている。

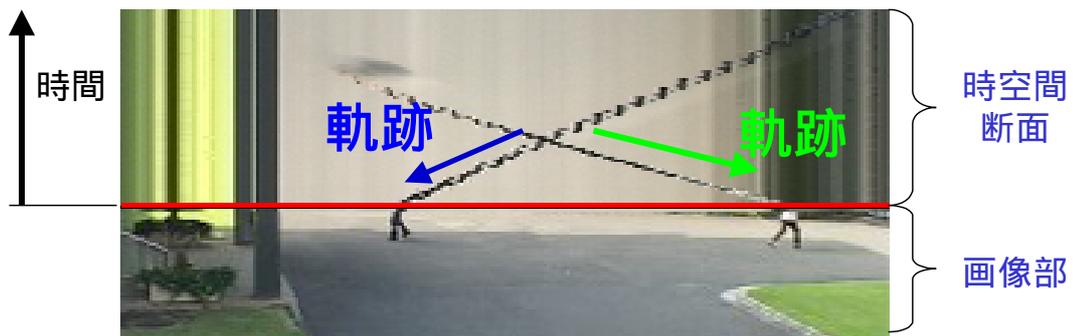
上述のブロックマッチング法や勾配法によるオプティカルフロー検出は2画像間の対応を求めたが、連続画像を並べた図2-3bのような時空間画像内では、空間内のエッジなどの軌跡を追跡することで、より正確な動きが計測できる。時空間画像からエッジなどを検出する処理は、本質的に  $x$  軸、 $y$  軸、 $t$  軸の3次元を扱うことになる。昨今、このような時空間画像やCTスキャナーの画像を処理するため、2次元の画像処理を3次元に拡張した3次元画像処理の基本的なアルゴリズム[136]が整理されだしている。しかし、3次元画像処理は計算量がかかり、画像監視システムなどの実時間処理が求められる応用では利用が困難である。



a. 監視画像

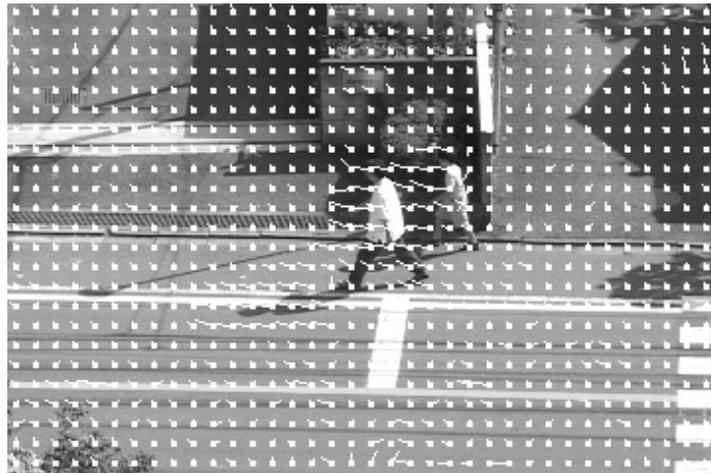


b. 時空間画像

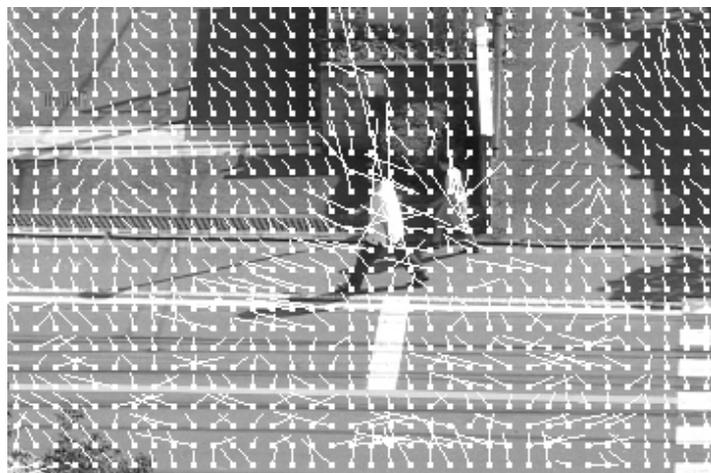


c. 時間断面画像

図 2-3 時空間画像による動き検出



a. 撮影間隔 0.06秒



b. 撮影間隔 0.2秒

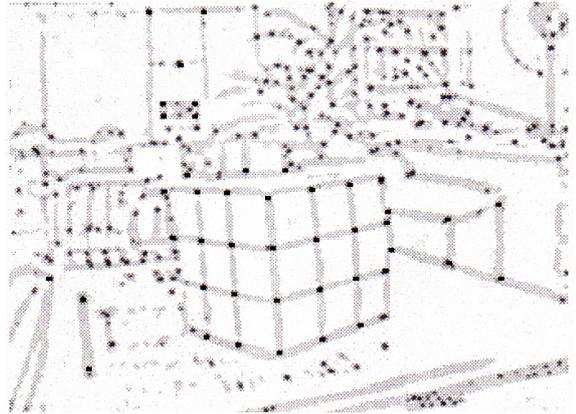
図 2-4 動きベクトルの検出結果

図2-4に勾配法の一方式[72]でオプティカルフローを計算した結果を示す。対象画像はカメラが上から下へ旋回している間に、歩行者が右から左へ移動しているシーンである。それぞれ連続画像の取り込み間隔を0.06秒と0.2秒にした時のオプティカルフローを示している。0.06秒で取り込んだときは、背景変化に比べ歩行者の動きが目立つが、0.2秒で取り込んだ画像では歩行者の動きが背景の移動の中に埋もれてしまうことが分かる。つまり、連続画像から動き変化を検出するのに時間がかかり、処理のフレームレートが落ちるようでは、背景が変化する画像や旋回中の監視カメラ画像から安定したオプティカルフローの検出が難しくなる。そのため、実用化にあたっては処理の高速化が重要になる。

例えばブロックマッチング法で、 $360 \times 240$ 画素の画像を $15 \times 15$ 画素の重ならない隣接するブロックに分割し、各ブロックの移動先を上下左右 $\pm 15$ 画素の範囲で探索すると、Pentium4, 2.26GHzの計算機で0.16秒と低速である。そのため、画像監視の実務でブロックマッチング法を使う場合、画像を縮小して処理の高速化を図るか、ハードウェア化を図る方法がとられている。



a. 現画像



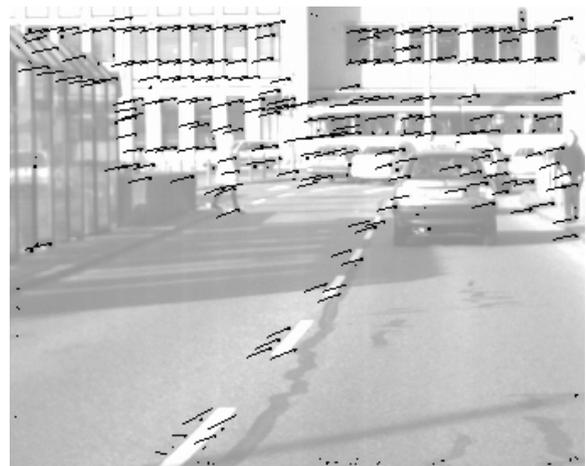
b. 特徴点の検出結果



c. 基準画像と数フレーム先の画像（走行中の車から撮影した画像）



d. 特徴点検出結果



e. 2 画像間の特徴点对応付け結果

図 2-5 特徴点の検出例と特徴点の対応付けの例

ブロックマッチングほどではないが、勾配法も繰り返し計算を行うため計算量がかかる。そのため、ブロックマッチングや勾配法を使わずに画像内の動きを検出する方式も数多く提案されている。例えば、図2-5a,bに示すように画像から物体のコーナーや変化点などの

特徴点を検出[73]し、その特徴点を画像間で対応付けることで画像内の動きを求める方式[74]も勾配法の定式化前から利用されている。図2-5c,d,eは移動中のカメラで撮影した画像から特徴点を検出し、画像間で特徴点を対応付けた結果を示している。特徴点の対応付けは弛緩法による繰り返し計算が必要になるが、画素数に比べ検出する特徴点の数が少ないため実時間処理が可能である。図2-5の実験結果では特徴点の対応付けよりも、むしろ特徴点の検出処理に計算時間がかかっている。

より高速な方式として、20年以上前の1983年に画像内のエッジを利用する方式[75]が提案されている。まず、画像 $I_t$ 内のエッジ部分をSobelフィルタ

$$S_x = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix} \quad S_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix}$$

を使い、エッジ $F_x, F_y$ を次式で求める。なお、 $m=n=1$ である。

$$F_x = \sum_{i=-m}^m \sum_{j=-n}^n I(x+i, y+j) S_x(i+m, j+n)$$

$$F_y = \sum_{i=-m}^m \sum_{j=-n}^n I(x+i, y+j) S_y(i+m, j+n)$$

その後、エッジ画像 $F_x, F_y$ と連続画像のフレーム間差分画像 $E_t$  (2.2.1参照)との積により、変化のあるエッジだけを取り出し、その輝度勾配でエッジの動きを検出する方式である。本論文の3章では、上記の文献[75]を参考にした処理の高速化を提案している。

あるいは、図2-6に示す縦横ラインで構成される小マスク上の動き変化を検出する方式[76]が提案されている。図2-6に示す小マスクを画像全体に配置し、連続した2画像間の局所的な明るさの位置変化を、小マスクの各縦横ラインだけで計算する方式である。動きの検出は各ライン上の1次元処理となるため、高速な処理が可能である。

上述のように、画像内の動きを検出する方式が多く提案されているが、目視監視を代行する画像監視システムの作成では、監視対象に応じて処理速度を優先するか、動きの検出精度を優先するかを明確にし、採用する検出方式を決める必要がある。

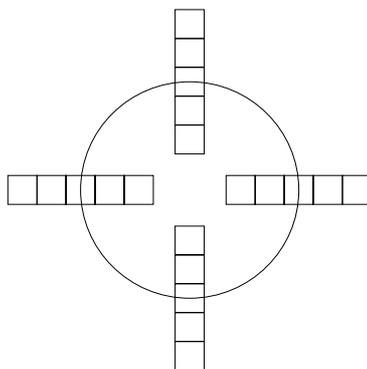


図 2-6 動き検出用に画像内に配置する小マスク[76]

## (2) 動きによる変化領域検出

画像内の動き特徴により移動体を検出する多数の方式が提案されている。例えば、連続した画像間で得られる画素 $(x,y)$ での動き変化を $F=(v_{xy},u_{xy})$ とし、 $O_{xy}=\tan^{-1}(v_{xy}/u_{xy})$ で各画素での動き方向を求め、 $O_{xy}$ に対しラプラシアンフィルタで方向が変化する境界を検出し、その境界に囲まれる領域を移動体として検出する方式[43]も多い。あるいは、画像間の動き変化を座標の歪として以下のようなアフィン変換

$$u_{x,y} = a_1x + a_2y + a_3, \quad v_{x,y} = b_1x + b_2y + b_3$$

または

$$u_{x,y} = a_1x + a_2y + a_3x^2 + a_4xy + a_5, \quad v_{x,y} = b_1x + b_2y + b_3y^2 + b_4xy + b_5$$

などのフィッティングモデルで表現し、そのモデルの係数 $a,b$ で画像内を領域分割する方式[44][45]などがある。

移動カメラで移動体を検出する方法として、周囲の画素との相対的な移動量から画像内の3次元形状を推定し、周りとの差が際立つ部分を検出する方式[46][47]が提案されている。他の方法として、画像内の動き成分を空間に投影して背景と移動体の動きを分離する方式[48][49]などがある。しかし、いずれの方式も処理に時間がかかり、実時間が要求される画像監視に使うには、特殊なハードウェアなどの利用が不可欠である。

一例として、図2-7と図2-8を使い文献[46]の方式による動き変化領域の検出結果を示す。静止カメラだけが移動している場合、2枚の連続画像間で次のような幾何学的な拘束が発生する。カメラの位置 $O$ が $O'$ へ移動したとき、静止物体 $P$ の画像内での移動 $\mu$ は図2-7aに示すような仮想平面 $\Pi$ を仮定することで次式のように表現できる。

$$\vec{\mu} = \frac{H}{Z} \frac{T}{d} (\vec{P}_w - \vec{e})$$

ここで、 $H$ は仮想平面と物体までの距離、 $Z$ は最初のカメラ位置 $O$ から物体までの奥行きである。 $T$ は最初のカメラ位置 $O$ から次のカメラ位置 $O'$ までの奥行きであり、 $d$ はカメラ位置 $O'$ から仮想平面までの距離である。また、カメラと画像面の距離は1としている。さらに、 $e$ は画像面でのエピポール(あるいはFOE: Focus of expansion)と呼ばれる点の位置を表す。また、 $P_w$ は画像面(Image Plane)での物体 $P$ の移動位置を示す。いずれも図2-7a中の値である。2つの静止物体 $P_1, P_2$ を移動カメラで撮影した場合、画像内でのそれぞれの物体の移動ベクトル $\mu_1, \mu_2$ は、

$$\vec{\mu}_1 = \frac{H_1}{Z_1} \frac{T}{d} (\vec{P}_{w1} - \vec{e})$$

$$\vec{\mu}_2 = \frac{H_2}{Z_2} \frac{T}{d} (\vec{P}_{w2} - \vec{e})$$

となる(図2-7b参照)。この2式から $e$ を消去し、 $r_i = H_i/Z_i$ とすると、

$$\vec{\mu}_1 r_2 - \vec{\mu}_2 r_1 = r_1 r_2 \frac{T}{d} (\vec{P}_{w1} - \vec{P}_{w2}) \quad (2-1)$$

となる。式(2-1)で $r_1 r_2 T/d$ はスケールパラメータの役割をしているため、式(2-1)の両サイドのベクトルは平行であることを意味している。そのため、

$$\overrightarrow{\Delta P_w} = (\overrightarrow{P_{w1}} - \overrightarrow{P_{w2}})$$

とし、 $\overrightarrow{\Delta P_w}$ に直交するベクトルを $(\overrightarrow{\Delta P_w})_{\perp}$ とすると、

$$(\overrightarrow{\mu_1 r_2} - \overrightarrow{\mu_2 r_1})^T (\overrightarrow{\Delta P_w})_{\perp} = 0$$

となる。この式から容易に以下の式(2-2)が導き出せる。

$$\frac{\overrightarrow{\mu_2}^T (\overrightarrow{\Delta P_w})_{\perp}}{\overrightarrow{\mu_1}^T (\overrightarrow{\Delta P_w})_{\perp}} = \frac{r_2}{r_1} = \frac{Z_1 H_2}{Z_2 H_1} \quad (2-2)$$

さらに仮想平面が遠くにあり $H_2/H_1 \approx 1$ と仮定すると、式(2-2)は2点の相対的な奥行き関係を示すことになる。もしカメラが画像面に対し並行に移動した場合、 $T=0$ となり式(2-1)が

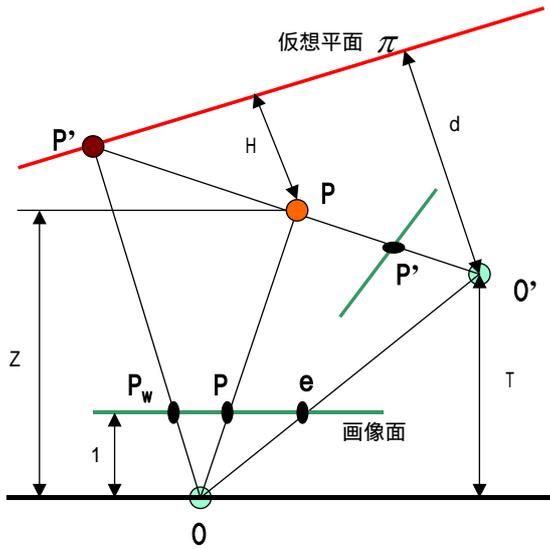
$$\overrightarrow{\mu_1 r_2} - \overrightarrow{\mu_2 r_1} = 0$$

となるので

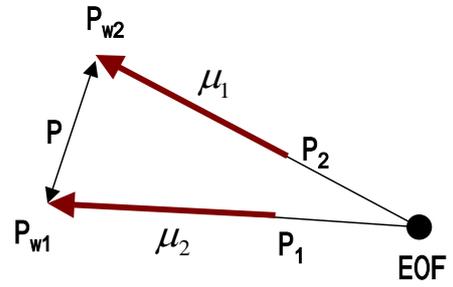
$$\frac{\overrightarrow{\mu_2}}{\mu_1} = \frac{r_2}{r_1}$$

となり、ステレオビジョンの原理と同様にベクトルの移動量が奥行きの関係を示すことになる。以上のことより、被写体が静止し1台のカメラが任意に移動する状態で撮影した連続画像から2枚の画像を取り出し、画像間の対応付けを行った後に式(2-2)を使うことで、映像内の3次元的な相対的奥行き情報を抽出することができる。図2-7c,dに式(2-2)を使い2画像から求めた相対的な奥行き画像を示す。図2-7dは奥行きを輝度値で表示した結果であり、白い部分がカメラに近く暗いほどカメラから遠いことを示している。もし、画像内に移動体を含む場合、周囲と極端に異なる奥行き部分が現れるので、その周囲と異なる部分を移動体として検出できる。図2-8aは走行中の車から撮影した画像であり、前述までの処理により奥行きを抽出し、カメラに近い奥行き部分を抽出した結果が図2-8bである。図2-8bでは人物や車両の一部がカメラに近い部分として抽出できている。図2-8bは、カメラに近い側として計測した領域を表示したが、逆に実際の奥行きよりも遠くにあるとして計測されている移動領域もある。

計算量のかからない動き領域の検出方式として、フレーム間差分などの履歴（動き履歴画像、MHI: Motion History Images）を用いる方式[77][78]がある。MHIによる変化領域の検出処理は、ビデオレートよりも速く0.02秒である。MHIによる動き検出は、広く利用されているHorn-Schunck[79]、Lukas-Kanade[80]、ブロックマッチング法に比べ2.8~9.5倍の処理速度である[81]。そのため、既存のパソコンを使いビデオレートで変化領域を検出できる有望な方式である(MHIの詳細は3.3.2参照)。



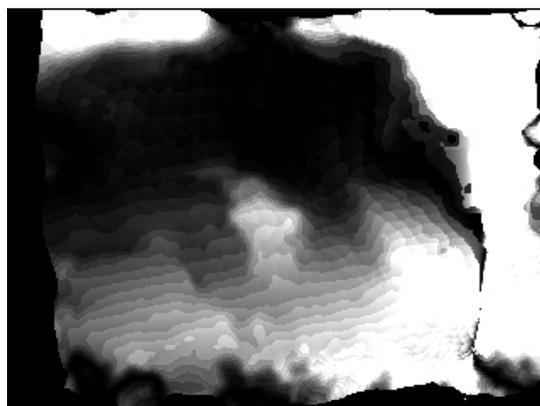
a. 物体  $P$  と 2 画像の幾何学的な関係



b. 2 物体  $P_{w1}, P_{w2}$  の画面内変化



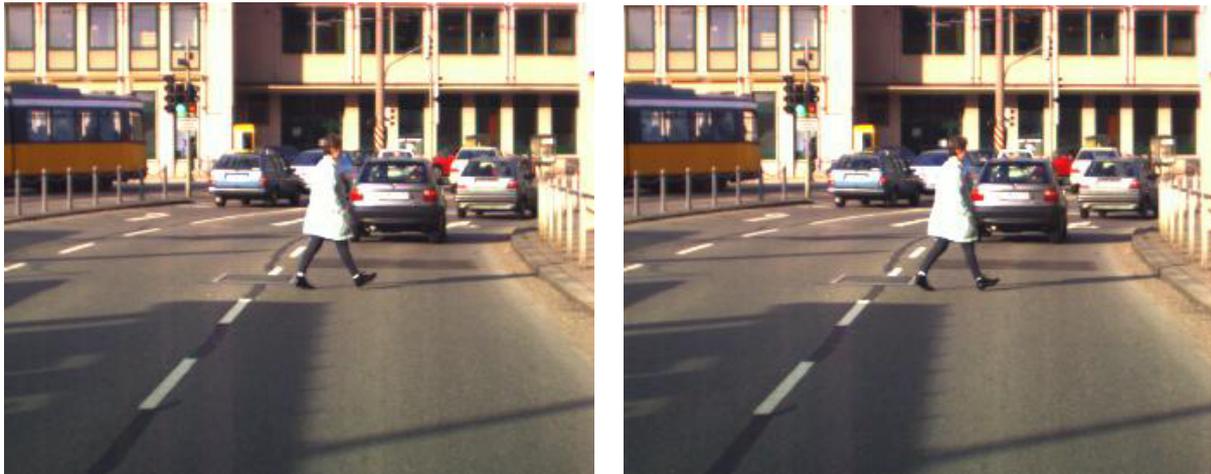
c. 基準画像と数フレーム先の画像（移動カメラで撮影した画像）



d. 相対的奥行き算出結果

移動カメラに近いほど明るく、遠いほど暗くなっている

図 2-7 3次元形状の推定（方式[46][47]）



a. 走行中に撮影した連続画像



b. カメラに近い部分として計測された領域の検出結果

図 2-8 奥行きによる移動体の検出結果例



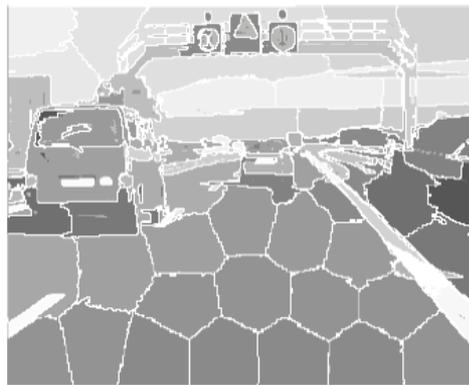
図 2-9 MHI を用いた移動体領域の検出結果例

一例として図2-9に、固定カメラで右から左へ通過する歩行者を撮影した画像から、歩行者を検出した結果を示す。図2-9内の円が検出した変化領域を示し、円内から伸びている直線は移動方向を示している。MHIは画像監視システムの画像処理として、処理速度と精度の面から有効な方式である。そのため、本論文の3.3節ではMHIの詳細を述べ、3.3節で提案する検出方式とMHIの検出方式の処理結果を詳細に比較する。

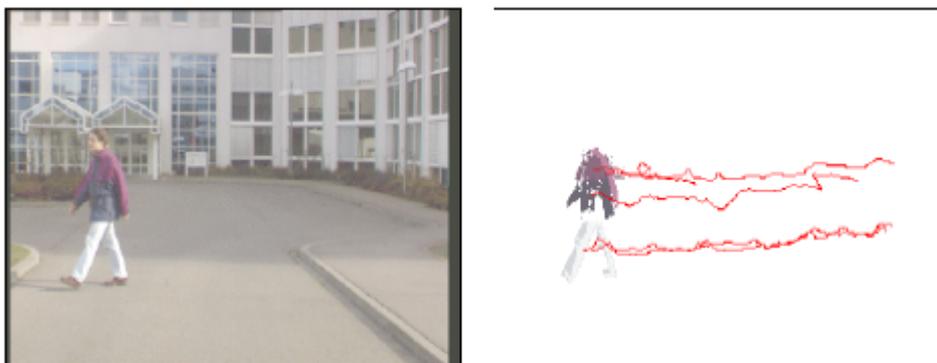
### (3) 動きと色・テクスチャ等の連携による変化領域検出

画像内の動きと輝度値を使う移動体検出方式[50]が1980年に提案されて以来、動きと他の特徴を組み合わせた多くの移動体検出方式が報告されている[45]。例えば、動きと色による領域分割を繰り返す方式[51][52]、動きとテクスチャによる領域分割を利用する方式[53]、時間方向のテクスチャ相関と時間及び空間方向の領域相関に加え、隣接領域との動きの相関を使い移動領域を検出する方式[54]などが提案されている。一例として、文献[82]の色による領域分割と、領域分割後の動き領域の追跡結果を図2-10に示す。

一般に、扱う特徴を増やすことで精度の高い移動体検出が可能になるが、扱う情報量が増えるに従い変化領域を検出するまでの繰り返し計算が増え処理時間がかかる問題がある。そのため、既方式を実時間監視に利用するには計算量を減らすアルゴリズムを考案するか、特殊なハードウェアを利用する必要がある。なお、画像処理による顔検出などでは肌色を選択的に検出することが頻繁に行われているが、肌色の検出処理は照明変化の少ない室内



a. 色と座標による領域分割例



b. 領域分割後の動き検出結果例

図 2-10 動きと色による領域の追跡例 (文献[82]より抜粋引用)

の監視のみで有効であり、照明変化の激しい屋外では使用できない等の課題がある。

### 2.2.3 パターン認識による人物検出

歩行者の画像例を機械学習し、画像内から歩行者を検出する研究[58][59]が進められている。機械学習を利用した歩行者の検出は、図2-11示すような人物の全身画像と人物を含まない背景画像から、図2-12に示すHaarウェブレットで画像特徴を算出し、機械学習（サポートベクターマシンなど:2.5節参照）により図2-13に示すような人物画像と背景画像の2クラスに分離した学習結果を事前に作成する。その後、Haarウェブレットで新規画像の特徴を計算後、機械学習結果を参照し人物領域を選択的に検出する方式である。



図2-11 学習用人物画像例 (M.I.T.-CBCLの人物画像DBより引用)

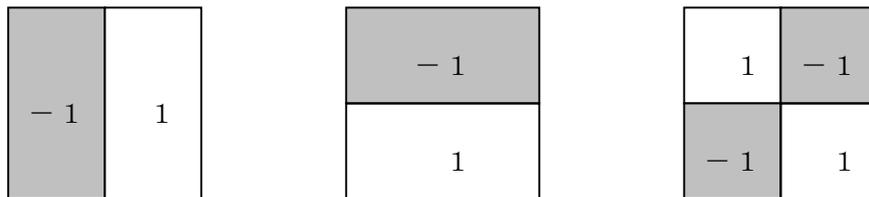


図2-12 画像から特徴を抽出するためのHaarウェブレット

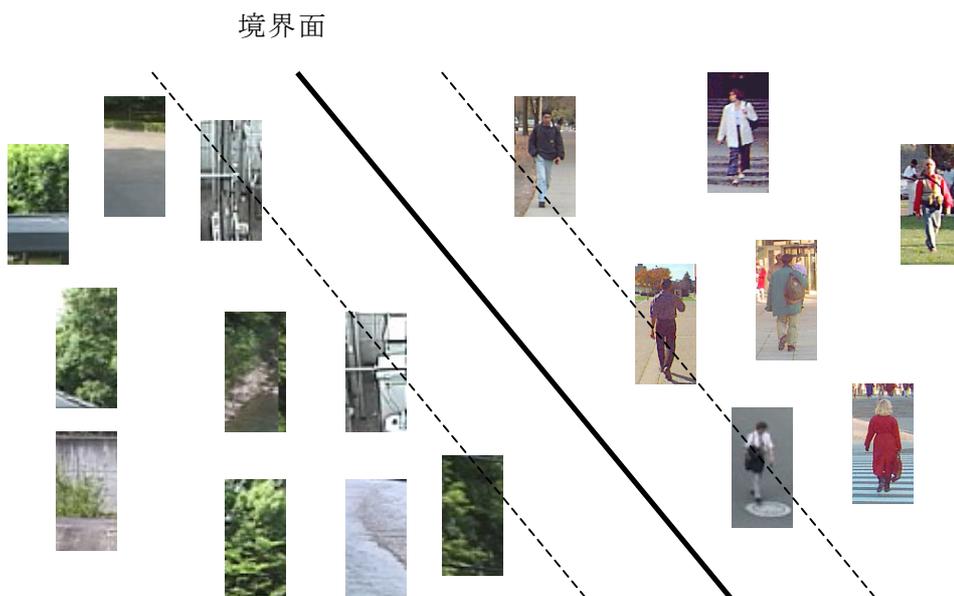
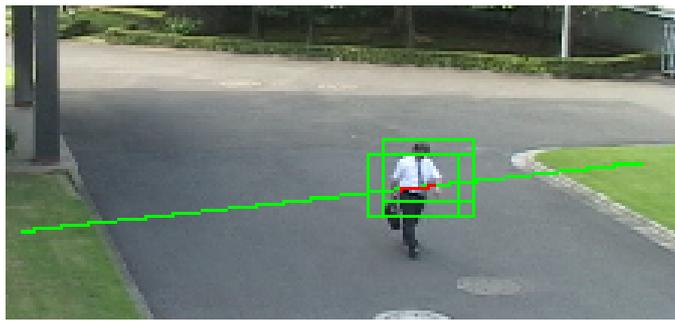
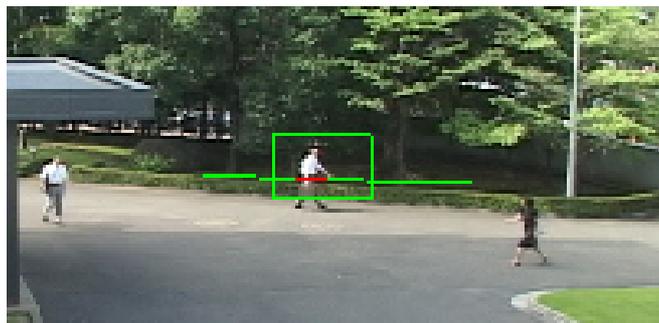


図2-13 人物画像と背景画像の学習イメージ

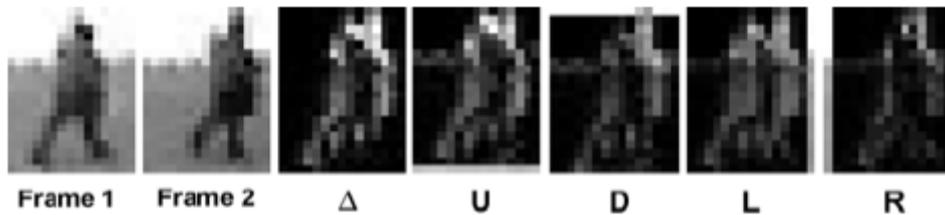


a. アスファルト上での歩行者検出例



b. 樹木の前での歩行者検出例

図2-14 歩行者の検出例



Frame1, Frame2は連続画像

図2-15 検出に利用する歩行者の画像特徴例 (文献[61]より引用)

Haarウェーブレットによる特徴抽出とSVMによる機械学習結果を使い、画像内から歩行者を検出した例を図2-14に示す。ただし、上記で示した人物検出処理を画像全体に行うと実時間処理ができないため、図2-14では画像内のラインが引いてある領域近傍だけに歩行者検出処理をかけた結果である（詳細は3.2節参照）。

監視カメラの画像内から歩行者を検出する研究では、画像内の動き変化がある部分に検出処理を行うと、歩行者の検出率が向上する実験結果が得られている[60]。そこで連続画像 $I_t$ と次画像 $I_{t+1}$ との間の変化を、

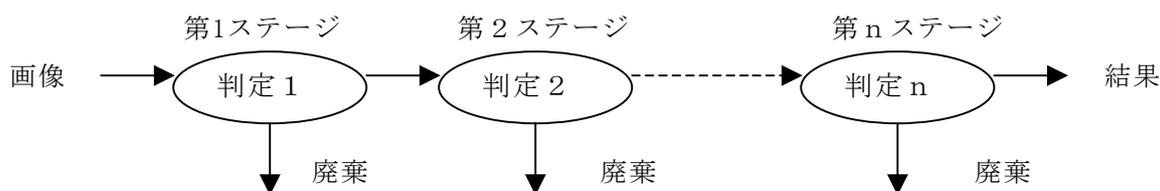


図2-16 カスケードによる処理の高速化

$$\Delta = \text{abs}(I_t - I_{t+1})$$

$$U = \text{abs}(I_t - I_{t+1} \uparrow)$$

$$L = \text{abs}(I_t - I_{t+1} \leftarrow)$$

$$R = \text{abs}(I_t - I_{t+1} \rightarrow)$$

$$D = \text{abs}(I_t - I_{t+1} \downarrow)$$

の5つの特徴 ( $\Delta, U, L, R, D$ ) で機械学習 (AdaBoost) を行い、歩行者を検出する方式[61]が近年提案されている。ここで、 $\{\uparrow, \downarrow, \rightarrow, \leftarrow\}$ は画像 $I_{t+1}$ を矢印の示す方向に1画素移動させることを意味している。図2-15に連続画像(図2-15,Frame1,Frame2)から検出した5特徴の画像例を示す。このような特殊な画像特徴は、従来のブロックマッチングや勾配法による動き計算では処理時間がかかり、実時間監視ができないために考案されたものである。画像内のオプティカルフローを計算するかわりに上記5特徴を使い、4fpsの処理速度で歩行者を検出することに成功している。

なお文献[61]は、上記5特徴の画像からHaarウェブレット形状のマスクを使い画像特徴を計算する必要があるが、次のような処理で計算の高速化を図っている[83]。はじめに、SAT (Summed Area Table) と呼んでいる配列を対象画像 $I$ から作成する。

$$SAT(x, y) = \sum_{x'=0}^x \sum_{y'=0}^y I(x', y')$$

この $SAT(x, y)$ には画像 $I$ の左上原点(0,0)から着目点 $(x, y)$ に囲まれる長方形内の値が合計されている。この $SAT$ をそのまま計算したのでは計算量がかかる。しかし、

$$SAT(x, y) = SAT(x, y-1) + SAT(x-1, y) + I(x, y) - SAT(x-1, y-1)$$

とすると、画像を1回処理するだけで $SAT$ を計算できる。この $SAT$ を使うと、領域 $r=(x, y, w, h)$ 内の合計 $Sum(r)$ は、

$$Sum(r) = SAT(x-1, y-1) + SAT(x+w-1, y+h-1) - SAT(x-1, y+h-1) - SAT(x+w-1, y-1)$$

の4回の配列参照で計算でき、Haarウェブレット形式の特徴を高速に算出できる。

文献[61]の歩行者検出では、さらなる処理の高速化のために図2-16に示すような段階的に人物領域を絞り込むカスケード方式[62]が利用されている。図2-16に示すように、はじめの第1ステージで人物の存在する部分と存在しない部分を分離する。次に、第2ステージで、第1ステージで人物領域として選択した部分に、より精度の高い判定を行う。同様な処理を $n$ 回繰返し、段階的に候補を絞り込むことで高速な人物検出を可能とする方式である。

このような機械学習を用いる人物検出方式は、学習画像を変更することで人物以外にも様々な物体を検出対象にできる利点がある。しかし、監視カメラの設置高さや撮影角度ならびに被写体までの距離ごとに事例画像を準備する必要があり、手間がかかる問題もある。また、既存の成果はいずれも固定監視カメラが前提であり、カメラが移動する場合、画像全体に変化が現れ既存の高速化処理が利用できない問題がある。

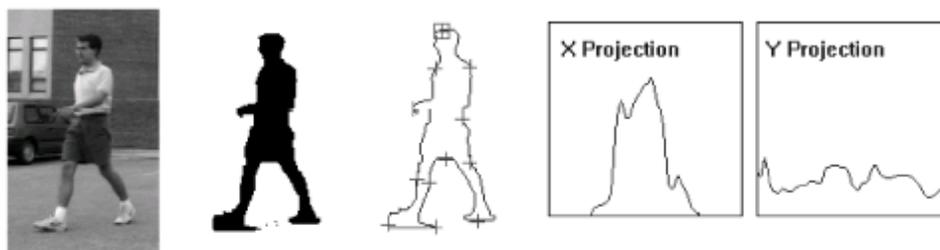
## 2.3 人物の認識と姿勢の認識

### 2.3.1 人物の認識

近年、画像内の人物と車両を区別したり、顔画像などから個々人を認識したりする研究成果が多く報告されだしている。

多数の固定監視カメラをネットワークで連携させ、監視画像内の人物や車両を検出し、その種別を分類する研究としてカーネギーメロン大学の VSAM (System for Video Surveillance and Monitoring)プロジェクト[63]や、混合分布モデルを使い背景画像から差分により人物や車両を検出するマサチューセッツ工科大学 AI ラボの Forest of Sensors プロジェクト[64]がある。人の監視に特化した画像監視システムとして IBM で販売している  $W^4$ [65]などがある。いずれも固定監視カメラを前提とした画像差分処理を利用している。一例として図 2-17 に  $W^4$  で利用している人物判定用の画像特徴の一例を示す。図 2-17 は画像から差分処理で人物領域を検出した後、人物領域のシルエットの凹凸を記録し、縦横方向の画素のヒストグラムを認識に利用することを示している。図 2-17 のヒストグラム以外にも、 $W^4$  では体の検出や追跡にテンプレートを使うなど様々な体の特徴を利用した方式となっている。しかし、いずれのプロジェクトや方式も人物や車の検出とその分類に特化しており、検出した人物を認識する研究は行なわれていない。

人物の認識に関しては、顔、虹彩、指紋、静脈など個人ごとに明確に異なる特徴を用いた認証技術が、銀行の ATM や入国審査など多くの分野で利用されだしている。さらに、昨今、連続画像から得られる歩き方の特徴を使い個人認証を行う研究[56]が始まっている。しかし、監視画像から検出した全身画像から個人を特定する方式は、個人を特定するための安定した特徴が検出困難なこともあり、研究成果が報告されていない。

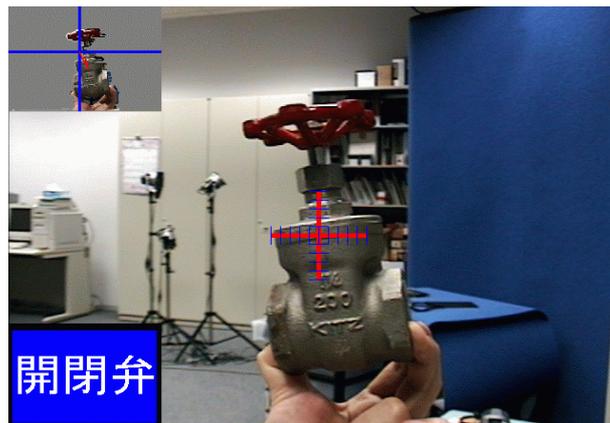


a.画像      b.人物領域      c.シルエット      e.縦横方向へ投影したヒストグラム

図 2-17  $W^4$ での人物特徴の検出例 (文献[65]より抜粋引用)



a. 人物を認識した例



b. 物体を認識した例

図 2-18 人物と物体の認識例

そこで、本論文の 5 章では全身画像を用いた人物認識方式を提案する。図 2-18 は本論文で提案する認識方式を用いた結果である。図 2-18a は人物に適用した例であり、画面の左上に認識した人物の名前、左下に認識した人物の向きを表示している。また、図 2-18b は同方式を物体認識に利用した例であり、左上に差分処理の結果を示し、左下に認識結果を示している。図 2-18 に示すように、人物以外の物体でも同じ枠組みで認識できる汎用的な方式である（詳細は 5 章参照）。

### 2.3.2 姿勢の認識

人物の姿勢や動作の検出用に、赤外光に反射するマーカを体に取り付け、赤外光を投光して人物の動作を取り込む装置が 1 千万円程度で市販されており、TV キャラクタの振り付けなどに広く利用されている。このような特殊な装置を用いることなく、画像内から人物の動作を簡易に検出するシステムの需要は非常に高く、現在も研究[129]が続いている。しかし、画像監視システムのように監視カメラが設置される様々な環境下で、簡易に利用できる姿勢認識方式はまだ実現できていない。

特に、目視機能を代行する画像監視システムには、個人の認識以外にも人物の動作や姿勢から何をしているか判断する機能が必要となる。上記で示した画像監視システムW<sup>4</sup>も人物のシルエットから特定姿勢を判断する機能を備えている。しかし、現場レベルで姿勢認識機能を使うには、監視現場の人物動作に合ったチューニングが容易にできなければならない。そのような、現場レベルで使える姿勢認識機能を備えた画像監視システムは存在しないが、今後発展が期待できる研究分野である。そのような研究成果として以下のような事例がある。

画像内の人物の姿勢を CG 化する研究成果として、腕、足、腰、胴体等が関節でつながる 2 次元あるいは 3 次元の身体モデルを画像内の人物に当てはめ、動作を正確に CG 化する研究が進められている[55][66][67][68]。図 2-19 は身体モデルを画像に当てはめた例である。マサチューセッツ工科大学での Pfinder[84]などもその一つの研究成果である。Pfinder

は、室内に一人の人物が現れる前提で、身体モデルを人物画像に実時間でフィッティングする方式である。このような従来方式は、カメラの位置や高さなどが大きく変わると、人物の見え方が変わるため、撮影環境に合わせて身体モデルのフィッティング処理を作成する必要がある。

現在、機械学習による汎用的な枠組みで姿勢を認識する研究成果[55][116]は少ないが、事例画像を学習させるだけで姿勢を認識できれば、人物の動作を把握する画像監視システムの開発に貢献する成果となる。そこで、本論文の4章では身体モデルの当てはめ方式ではなく、事例画像を機械学習する汎用的な方式を用いた姿勢認識を検討する。図 2-20 は発電所の運転員を撮影した監視カメラの画像から、4章で示す提案方式で運転員の姿勢を認識し、認識結果を CG 表現した例である。図 2-20 b の人体 CG の足元の線は、運転員の移動軌跡を示している（詳細は4章参照）。



図 2-19 人体モデルの当てはめ例（文献[66]より抜粋引用）



a. 原画像例



b. 姿勢の認識結果と人物の移動軌跡

図 2-20 人物の姿勢認識例

## 2.4 画像監視の処理時間

固定監視カメラの映像を使う人物検出に特化した画像監視システム $W^4$ は、 $320 \times 240$ 画素の画像をDual-Pentium 400MHz を使い25fpsの速度で監視ができる。しかし、 $W^4$ は人物検出に特化した画像監視システムとなっているため、車両などの他の移動体検出への適用は困難である。また、固定カメラを前提としているため、旋回中の監視カメラに対応できない問題がある。

機械学習を利用した汎用性のある人物検出方式[58][59]は、学習事例を変更することで車両や動物など多様な物体を検出対象にできる。しかし、機械学習を利用した初期の人物検出方式は、画像内を探索する必要から1画像の処理に数分程度かかっていた。近年、歩行者の移動で発生するフレーム間差分変化を画像特徴とし、機械学習結果をカスケード化して歩行者を検出する方式[61]が考案(2.2.3参照)されたことで、 $360 \times 240$ 画素の画像に対しPentium4 (2.8GHz)の計算機で4fps程の処理速度が可能となった。また、直線的に移動する人物と背景の違い(「画像内の局所変化を空間平均した動きの強さ」、「時間平均した動きの強さ」、「時間的な動きの一様性」の3特徴)を学習し、歩行者と背景を分離する方式[42]では、Pentium3 (933 MHz)の計算機(Matrox Meteorキャプチャーカード)を使い $320 \times 240$ 画素の画像を7.5fpsで処理している。

目視に代わる画像監視システムは、ビデオレート(30fps)で侵入者監視ができることが望ましい。しかし、今のところ高速化を図った機械学習による人物検出方式は4~7.5fps程度の処理速度であり、実時間監視に向け更なる高速化の研究が必要である。また、いずれも固定した監視カメラを前提としており、旋回中のカメラや背景が動く場合、背景全体が変化するため利用できない課題がある。この課題に対し、本論文では3章で固定監視カメラと旋回中の監視カメラの画像から実時間で侵入者を検知する方式を提案する。図2-21に携帯ビデオカメラで移動しながら撮影した画像から、人物領域を実時間で検知した実験結果例を示す(詳細は3章参照)。

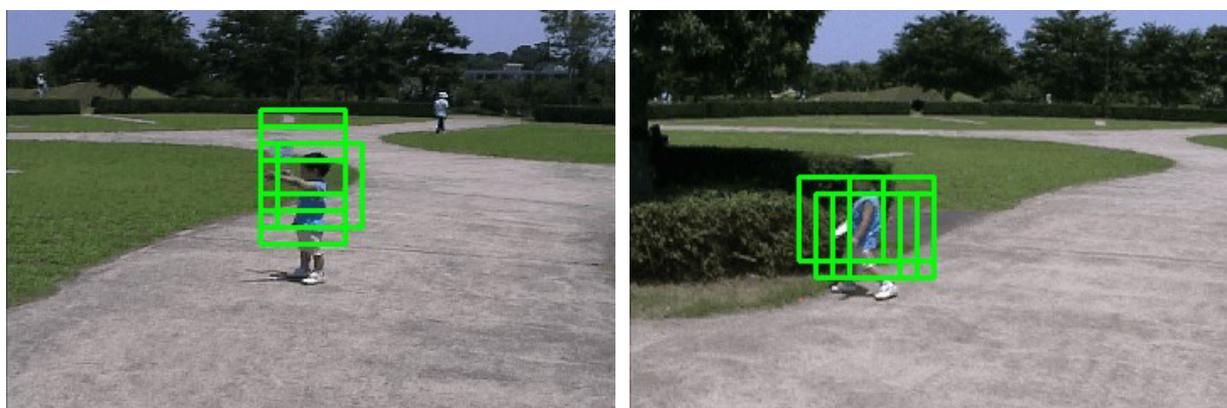


図 2-21 携帯ビデオカメラで撮影した画像からの実時間検知例

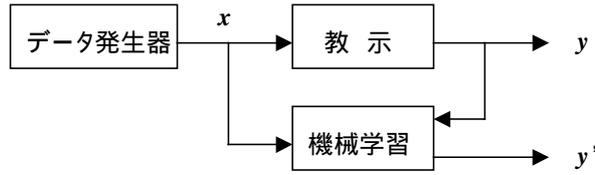


図 2-22 教師付き学習のモデル[57]

## 2.5 検出と認識に利用する機械学習

前節までに述べた検出や認識の基礎となる機械学習を説明する。機械学習は、教師無し学習と教師付き学習と呼ばれている2通りに大分類できる。教師無し学習は、クラスタリングあるいはベクトル量子化などにより、パターンの分布を幾つかのクラスに分ける方式である。一方、教師付き機械学習は、サポートベクターマシンやニューラルネットワークなどのように学習パターンとその正解クラスを与え、パターンを正しく分類する関数を作成する方式である。

本論文では、監視対象の人物画像が収集できる前提で、教師付き学習方式を説明する。この教師付き学習は図2-22に示すモデルで表現できる。データ発生器から得られたパターン（データ） $x_i$ に教示段階でラベル付け $y_i$ を行い、機械学習器に $\{x_i, y_i\}, i=1, \dots, n$ ,を与え分類関数を作成する。その関数にパターン $x_i$ を与えたときに得られるラベル付け $y_i'$ と $y_i$ との不一致を少なくする方式と見なせる。

本節では学習で最も基本的なベイズ識別を説明した後、本論文で利用する最近傍決定則とSVMならびにAdaBoostについて解説する。

### 2.5.1 ベイズ識別

認識の古典的な手法にベイズ識別がある。ベイズ識別は、事前にパターンの分布状況あるいは確率密度が分かっている場合に有効な方式である。

あるC個のクラス $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ に対し、1つの入力パターン $\mathbf{x} = \{x_1, x_2, \dots, x_d\}$ がd次元ベクトルで与えられるとする。クラスを表す確率変数を $\omega$ とし、 $\omega = \omega_i$ となる確率を $P(\omega_i)$ としクラス $\omega_i$ の生起確率と呼ぶ。クラス $\omega_i$ 内のパターン $\mathbf{x}$ の確率分布関数を大文字の $P(\mathbf{x}|\omega_i)$ 、確率密度関数を小文字の $p(\mathbf{x}|\omega_i)$ と書くことにする。ここで、パターン $\mathbf{x}$ が与えられたときのクラス $\omega_i$ の条件付確率を $P(\omega_i|\mathbf{x})$ 、パターン $\mathbf{x}$ の確率分布関数を $P(\mathbf{x})$ 、確率密度関数を $p(\mathbf{x})$ 、パターン $\mathbf{x}$ とクラス $\omega_i$ の同時確率密度を $p(\omega_i, \mathbf{x})$ とすると次の関係が成り立つ[87]。

$$0 \leq P(\omega_i | \mathbf{x}) \leq 1, \quad \sum_{i=1}^c P(\omega_i | \mathbf{x}) = 1$$

$$p(\mathbf{x}, \omega_i) = p(\mathbf{x} | \omega_i)P(\omega_i) = p(\mathbf{x})P(\omega_i | \mathbf{x})$$

$$P(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)P(\omega_i)}{p(\mathbf{x})}$$

$$p(x) = \sum_{i=1}^c p(x, \omega_i) = \sum_{i=1}^c P(\omega_i) p(x | \omega_i) = \sum_{i=1}^c p(x) P(\omega_i | x)$$

$$P(\omega_i) = \int p(x, \omega_i) dx = \int P(\omega_i | x) p(x) dx$$

最も単純な2クラス  $\omega_1, \omega_2$  に分類する問題を例にすると、

$$P(\omega_1) + P(\omega_2) = 1$$

で、パターン  $\mathbf{x}$  が与えられる前に2クラス  $\omega_1, \omega_2$  のどちらが出るかを予想する場合、事前確率を利用し  $P(\omega_1)$  と  $P(\omega_2)$  で大きいクラスを選ぶのが普通である。パターン  $\mathbf{x}$  が与えられたときは、入手した  $\mathbf{x}$  を使いクラス  $\omega_i$  の生起する確率を用いることが一般的である。つまり、

$$\lceil \max\{P(\omega_1 | x), P(\omega_2 | x)\} \rceil \text{となるクラス } \omega_1 \text{ を選択}$$

する方式が考えられる。事前知識として、 $P(\omega_i | \mathbf{x})$  よりも  $p(\mathbf{x} | \omega_i)$  のほうが得やすいことが多いため、ベイズの定理を使い

$$P(\omega_j | x) = \frac{p(x | \omega_j) P(\omega_j)}{\sum_{i=1}^2 p(x | \omega_i) P(\omega_i)} = a \cdot P(\omega_j)$$

と変形できる。この決定規則でパターン  $\mathbf{x}$  が誤って分類される確率は、

$$P(\text{誤分類} | x) = \begin{cases} P(\omega_2 | x), & \mathbf{x} \text{ が } \in \omega_1 \text{ と決定したとき} \\ P(\omega_1 | x), & \mathbf{x} \text{ が } \in \omega_2 \text{ と決定したとき} \end{cases}$$

であり全ての  $\mathbf{x}$  での平均的な誤り確率は

$$P_e = \sum_x P(\text{誤分類} | x) p(x)$$

となる。この  $P_e$  は  $P(\mathbf{x} | \omega_1)$  と  $P(\mathbf{x} | \omega_2)$  の小さいほうを選ぶことで最小になるので、

$$P(\text{誤分類} | x) = \min\{P(\omega_1 | x), P(\omega_2 | x)\}$$

と表せる。以上の2クラス問題を発展させ、パターン  $\mathbf{x}$  を  $C$  個のクラス  $\{\omega_1, \omega_2, \dots, \omega_c\}$  のいずれかに入れる場合、

$$p(x | \omega_i) P(\omega_i) = p(x, \omega_i) = p(x) P(\omega_i | x)$$

を最大にするクラス  $\omega_i$  を選択すればよい。つまり、次のような選択規則となる。

$$\lceil P(\omega_i | x) > P(\omega_j | x), \forall i \neq j \rightarrow x \in \omega_i \rceil$$

これは最大事後確率法 (Maximum a posteriori probability method) と呼ばれている。

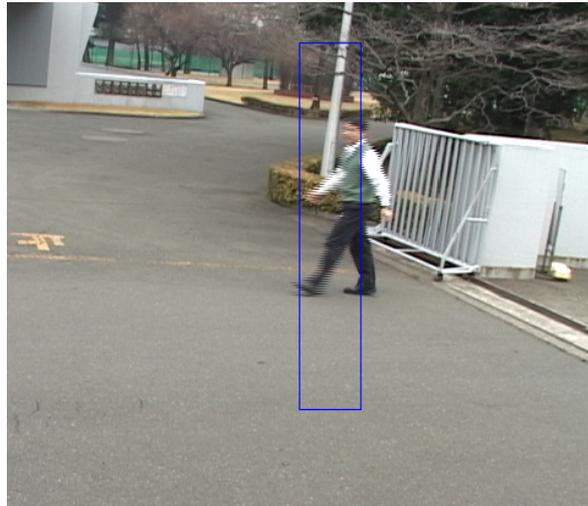
特に、パターン認識の正解に0、誤り時に1を与える損失関数の下での誤り確率  $R_b$  は

$$R_b = 1 - \sum_x p(x) \max_k P(\omega_k | x) = 1 - \sum_x \max_k p(x | \omega_i) P(\omega_i)$$

となり理論的な性能限界となる [86]。事前に確率分布  $P(\omega_i)$  や  $p(\mathbf{x} | \omega_i)$  が分かっているならば、パターン認識の問題は最適化問題に帰着し、どのような識別方式もこれ以下の誤り率を達成することができない。

以下、具体的な事例を用いて説明する。例えば、図2-23に示すような門の入退所者数を計測する問題を取上げる。図2-23内での計測領域は画像中央の青枠である。青枠の横幅はおおむね人物の横幅程度である。この計測領域の動き方向で入所と退所を判断できる。もし、歩行者が重ならずに通過する場合、計測領域内の画素全体の移動方向から入退所を判

定できるだけでなく、計測領域内の変化回数で入退所者数を正しく計測できる。しかし、歩行者が重なりながら通過する場合は、簡単に人数を計測できない。



画像内の青枠内を物体が通過する時間を計測

図 2-23 門の入退出者数を計測するための領域

表2-1 計測領域内の動き変化時間の長さとお過人数の頻度

| 長さ  | 1人   | 2人  | 3人  | 4人 | 5人以上 | 合計   |
|-----|------|-----|-----|----|------|------|
| 0   | 0    | 0   | 0   | 0  | 0    | 0    |
| 4   | 0    | 0   | 0   | 0  | 0    | 0    |
| 8   | 5    | 0   | 0   | 0  | 0    | 5    |
| 12  | 39   | 0   | 0   | 0  | 0    | 39   |
| 16  | 49   | 1   | 0   | 0  | 0    | 50   |
| 20  | 158  | 2   | 0   | 0  | 0    | 160  |
| 24  | 462  | 10  | 0   | 0  | 0    | 472  |
| 28  | 518  | 23  | 0   | 0  | 0    | 541  |
| 32  | 377  | 56  | 1   | 0  | 0    | 434  |
| 36  | 204  | 74  | 7   | 0  | 0    | 285  |
| 40  | 125  | 73  | 6   | 1  | 0    | 205  |
| 44  | 72   | 95  | 6   | 0  | 0    | 173  |
| 48  | 46   | 76  | 10  | 3  | 0    | 135  |
| 52  | 34   | 64  | 8   | 3  | 0    | 109  |
| 56  | 24   | 51  | 17  | 3  | 0    | 95   |
| 60  | 9    | 41  | 15  | 5  | 0    | 70   |
| 64  | 11   | 35  | 10  | 3  | 1    | 60   |
| 68  | 6    | 21  | 15  | 4  | 0    | 46   |
| 72  | 6    | 9   | 10  | 5  | 0    | 30   |
| 76  | 2    | 11  | 17  | 9  | 1    | 40   |
| 80  | 1    | 9   | 5   | 7  | 0    | 22   |
| 84  | 4    | 5   | 5   | 6  | 5    | 25   |
| 88  | 1    | 5   | 7   | 2  | 2    | 17   |
| 92  | 1    | 2   | 2   | 7  | 1    | 13   |
| 96  | 2    | 1   | 4   | 2  | 2    | 11   |
| 100 | 1    | 3   | 6   | 4  | 1    | 15   |
| 104 | 1    | 1   | 1   | 6  | 3    | 12   |
| 108 | 0    | 1   | 2   | 1  | 5    | 9    |
| 112 | 0    | 1   | 1   | 3  | 7    | 12   |
| 116 | 0    | 0   | 1   | 2  | 2    | 5    |
| 120 | 1    | 0   | 0   | 0  | 1    | 2    |
| 124 | 0    | 2   | 0   | 2  | 1    | 5    |
| 128 | 0    | 0   | 1   | 0  | 1    | 2    |
| 132 | 0    | 1   | 0   | 0  | 2    | 3    |
| 136 | 0    | 0   | 0   | 1  | 2    | 3    |
| 140 | 0    | 0   | 0   | 0  | 1    | 1    |
| 144 | 0    | 0   | 0   | 0  | 1    | 1    |
| 合計  | 2159 | 673 | 157 | 79 | 39   | 3107 |

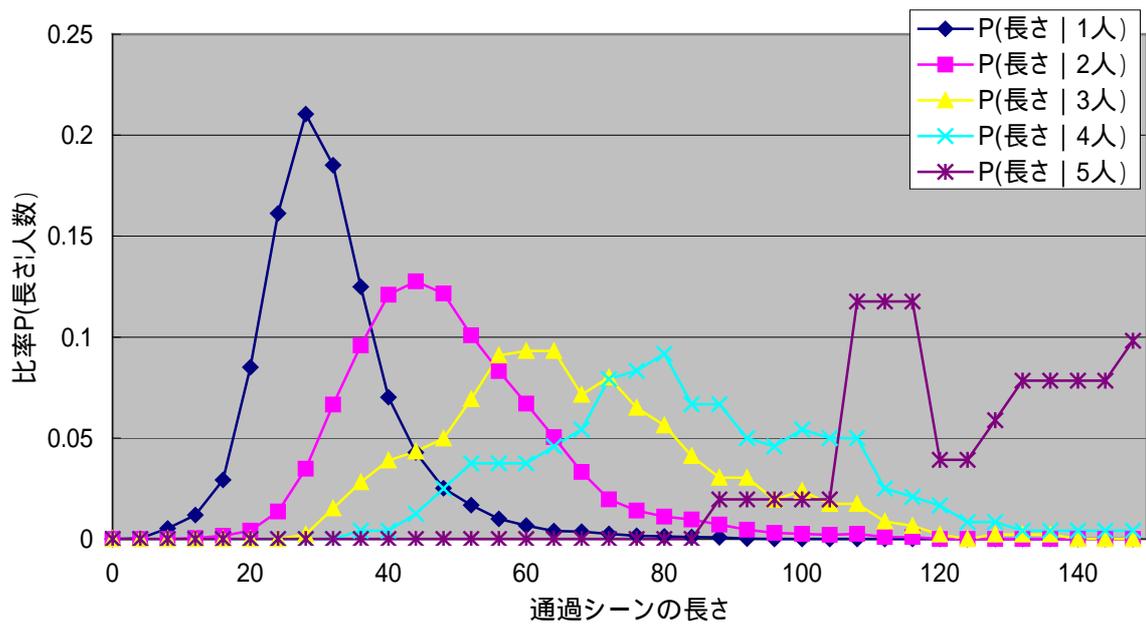


図2-24 通過時間の長さ と人数の関係 P(長さ|人数)

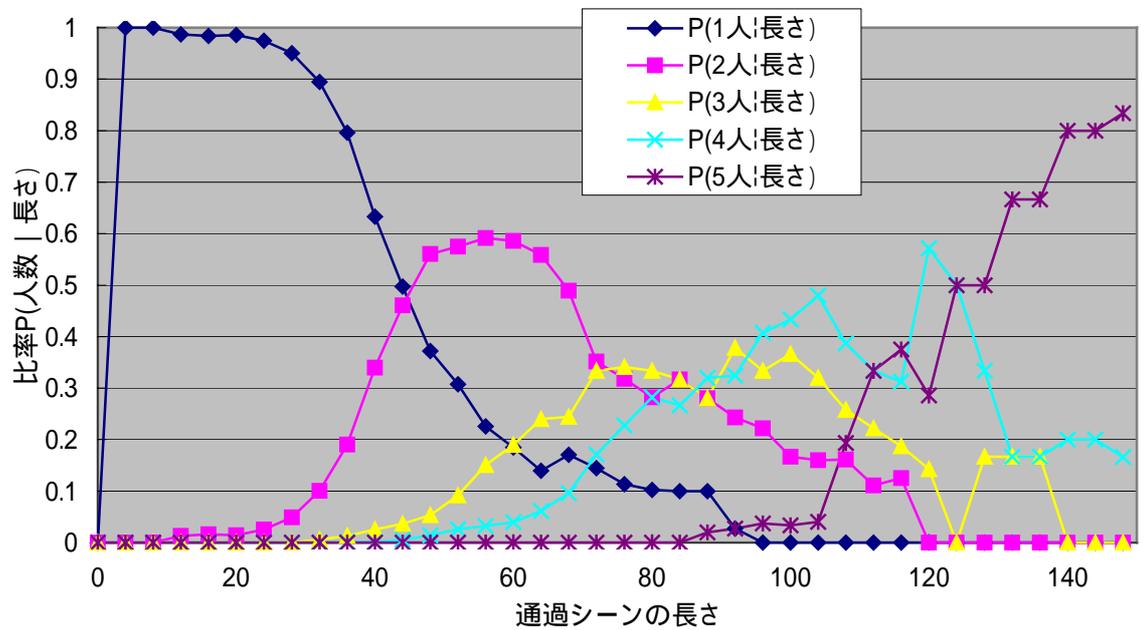


図2-25 通過時間の長さ と人数の関係 P(人数|長さ)

そこで、4,716人を含む14時間の録画画像を用い、歩行者が通過する時の計測領域内の動き変化時間と通過人数の関係を3,107シーンで調べ、計測領域内の動き変化時間と通過人数の関係を表2-1にまとめた。14時間の画像は、晴れた2日間に監視カメラで撮影したものである。表2-1には計測領域内で人が交差するシーンや、車が背景に入るシーン、あるいは車だけのシーンなどは含んでいない。つまり、入所あるいは退所のいずれかの方向だけに移動している入退所者シーンを集計した結果である。図2-24と図2-25に、表2-1から作成した通過人数ごとのシーンの長さ（図2-23の計測領域内の変化時間）と、シーンの長さごとの通過人数分布を示す。この問題は、図2-23の計測領域内での変化時間（長さ）が入力パターン $x$ であり、通過人数がクラス $C$ になるため、ベイズ誤り確率 $R_b$ は

$$R_b = 1 - \sum_{\text{長さ}} p(\text{長さ}) \max_{\text{人数}} P(\text{人数} | \text{長さ})$$

となる。

上記の例では、5人以上の通過が少ないので5人以上を1クラスとし、各クラスの分布を調べている。特に、晴れた2日間に撮影した画像を利用しているため、雨の日など傘をさして通過する場合、計測領域を通過する時間が長くなることが予想される。また、平日と休日では重なって通過する人物の割合が変わる可能性がある。

このように、実際の画像監視システムの設計時には、認識対象となるクラスの数やその分布は未知であることが多い。また限られた時間に撮影した事例画像からクラスの数や分布を決定しなければならない問題に頻繁に直面する。このような場合、訓練サンプルが無限に存在し確率分布が既知であるなどの仮定が成り立たない。

そこで、本論文では画像監視システムの開発に、パターンの確率分布の知識を用いない識別方式を用いることにした。次章以降では、パターンの確率分布を用いない識別方式として、古典的なパターン識別の中で最もよく知られている $k$ -最近傍決定則と、近年VC理論[57]でその理論的背景を得たSVM[57]、Intel社の公開しているOpenCVライブラリに付属するAdaBoostについて示す。いずれも、本論文の次章以降で利用する学習方式である。ただし、これらの方式もベイズ識別と理論的に関連しており（詳細は[119]を参照）、その限界はパターン分布が事前に分かった場合のベイズ識別精度となる。

## 2.5.2 $k$ -最近傍決定則

標本パターンの集合を $\{x_1, x_2, \dots, x_n\}$ とし、各パターンの属すクラス名も与えられているとする。このとき未知パターン $x$ と既存の標本パターン $x_i$ との距離を $d(x, x_i)$ として、距離が最小になる

$$d(x, x_i^*) = \min d(x, x_i)$$

$x_i^*$ を選び、 $x$ は $x_i^*$ と同じクラスに属すと見なす方式を最近傍決定則 (Nearest Neighbor rule) と呼ぶ[87]。特に、標本パターン中で1番近いものだけを利用するときの分類を1-NNと呼ぶ。また、与えられた入力パターン $x$ に最も近いものから $k$ 個選んだとき、その中に含まれるクラスの数によりパターンを決定する方式を $k$ -NNと呼ぶ。 $k$ -NNは直感的に分かりやすく、特

別な識別関数を構成する必要が無く、パターンの確率分布に関する知識もいらぬノンパラメトリックな方法である。

$k$ -NNの理論的限界が、以下の定理として知られている[85]。

**定理** ベイズ識別方式による最適な誤り確率と1-NNによる誤り確率をそれぞれ $R_b, R_n$ とすると、 $n$ 個の標本パターンに基づく最近傍決定側の誤り確率 $R_n$ の学習データ数が大きな極限値を $R = \lim_{n \rightarrow \infty} R_n$ とすると

$$R_b \leq R \leq \left(2 - \frac{K}{K-1} R_b\right) R_b$$

が成り立つ。ここで $K$ はクラスの個数である。

### 2.5.3 サポートベクターマシンの概要

SVMは線形の識別器であるがMercerカーネルを使うことで非線形に拡張され、近年文字認識や画像認識で高い認識率を上げているパターン認識方式である[57][86][88]。

SVMは2クラス分類器であり、学習データ $x_i \in R^d$ とそのクラス $y_i \in \{-1, 1\}$ で、事前にラベル付けされた学習データ $\{x_i, y_i\}, i=1, \dots, n$ を使い

$$F(x) = \text{sgn}(w \cdot x + b)$$

の形式で表記される識別関数を求める方式である。ここで、 $\text{sgn}()$ は-1、0、1を返す符号関数であり、「 $\cdot$ 」は内積を表す。 $b$ はバイアス項と呼ばれるパラメータである。 $w \cdot x + b = 0$ となる面がSVMでの識別面となる。識別面から各クラスまでの距離をマージンと呼び、SVMではマージンを最大化するように識別面を決めることで誤判定を少なくしている。識別面は一般に多数の学習サンプルに「サポート」されるように決定されるため、サポートベクターマシンと呼ばれている。

全ての学習データが線形分離可能ならば、

$$\begin{aligned} y_i = 1 \quad \text{の時} \quad w \cdot x_i + b &\geq 1 \\ y_i = -1 \quad \text{の時} \quad w \cdot x_i + b &\leq -1 \end{aligned}$$

となり、

$$y_i(w \cdot x_i + b) \geq 1 \quad \forall i \tag{2-3}$$

と表現できる。このときのマージンの大きさは $1/\|w\|$ と表され、このマージンを最大化する問題は

$$\min \frac{1}{2} \|w\|^2$$

$$\text{制約条件} \quad y_i(w \cdot x_i + b) \geq 1 \quad \forall i$$

と等価となる。この制約条件は、識別面が訓練サンプルを100%認識することを示している。

この制約付問題をラグランジュ乗数 $\alpha_i (\geq 0)$ を利用して表現すると、

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i ((w \cdot x_i + b) - 1) \tag{2-4}$$

となる。この最適化問題を解くには、 $w$ と $b$ に関して $L$ を最小化し、 $\alpha$ に関して最大化すればよい。この最適解では $L$ の勾配が0になるので、

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \quad (2-5)$$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (2-6)$$

となる。式(2-5)と式(2-6)を式(2-4)に代入すると、上記で定式化した問題は $\alpha$ による次の最大化問題になる。

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \text{制約条件} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad \alpha_i \geq 0 \end{aligned}$$

ここで式(2-6)を満たすペアを $(\bar{\mathbf{w}}, \bar{b})$ とすると、

$$\bar{\mathbf{w}} = \sum_{i=1}^n \bar{\alpha}_i y_i x_i$$

であり $\bar{b}$ と $\bar{\alpha}$ で上記の双対問題をKuhn-Tucker条件で

$$\bar{\alpha}_i (y_i (\bar{\mathbf{w}} \cdot x_i + \bar{b}) - 1) = 0$$

として決定できる。 $\bar{\alpha}_i$ は式(2-6)の中でゼロでなく、かつ式(2-3)を満たすものであり、その時の $x_i$ がサポートベクタとなる。新たに入力されたデータ $x$ は、単に $\bar{\mathbf{w}} \cdot x + \bar{b}$ の符号で分類されることになる。

一般に、特徴空間の次元数 $d$ が学習サンプル数 $n$ よりも大きければ、学習サンプルは超平面によって線形分離可能となる。線形分離が不可能な場合以下の2通りの方法がある。一つは、多少の識別誤りを許すように制約を緩める方法でソフトマージンと呼ばれている。他の一つは、 $x$ をMercerカーネル関数による非線形変換 $\phi(x)$ により、高次元空間に写像する方法でありカーネルトリックとも呼ばれている。

## (1) ソフトマージン

誤分類を許容する度合いを表す $\xi_i \geq 0, i=1, \dots, n$ と、誤分類とマージン最大化の関係を調整するパラメータ $C$ を導入し、最小化問題を以下のように定義する。ただし、パラメータ $C$ の値は実験的に決定する必要がある。

$$\min \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

$$\text{制約条件} \quad y_i (\mathbf{w} \cdot x_i + b) \geq 1 - \xi_i$$

この制約条件付きの最適化問題は、上記と同様にラグランジェの乗数法を用いると

$$\max \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{制約条件} \quad \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \quad \forall i$$

の双対問題に帰着し、 $\alpha_i > 0$ となるポイントがサポートベクタとなる。

## (2) Mercerカーネルによる非線形変換

SVMは学習事例間の内積で記述されているので、Mercerカーネル関数に置換ることにより $\mathbf{x}$ を非線形変換 $\phi(\mathbf{x})$ により高次元の空間に写像し、識別面を求めることができる[54]。このMercerカーネルとは、カーネル関数が下記の「半正定値性」を満たしているものである[50]。

**定義** [半正定値性] 任意の $n > 1$ 、任意の $x_1, \dots, x_n \in \mathcal{X}$  に関して

$$\sum_{i,j} c_i c_j K(x_i, x_j) \geq 0, \quad \forall c_1, \dots, c_n \in \mathbb{R}$$

が成り立つとき、関数 $K$ は半正定値であるという。

この半正定値性の定義から、次のような定理が導かれる。

**定理** 任意のMercerカーネル $K$ に対し、

$$K(x, x') = \sum_{k=1}^D \phi_k(x) \phi_k(x')$$

を満たすような写像

$$\Phi(x) := \{\phi_k(x)\}_{k=1}^D$$

が存在する。

$\Phi_k(x)$ を基底とする $D$ 次元のベクトル空間は、特徴空間と呼ばれる。それに対し、もとの空間 $\mathcal{X}$ を入力空間と呼ぶ。上の定理は、入力空間における2点 $\mathbf{x}, \mathbf{x}'$ のカーネル関数が、特徴空間に写像された点 $\phi(\mathbf{x}), \phi(\mathbf{x}')$ に内積として表されることを示している。Mercerカーネルの代表的な例として下記のRBFカーネルや多項式カーネルなどがある。

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

$$K(x, x') = (x \cdot x' + 1)^p$$

ここで $r$ と $p$ はパラメータであり、実験的に決定する必要がある。カーネル関数を用いると式(2-4)は、

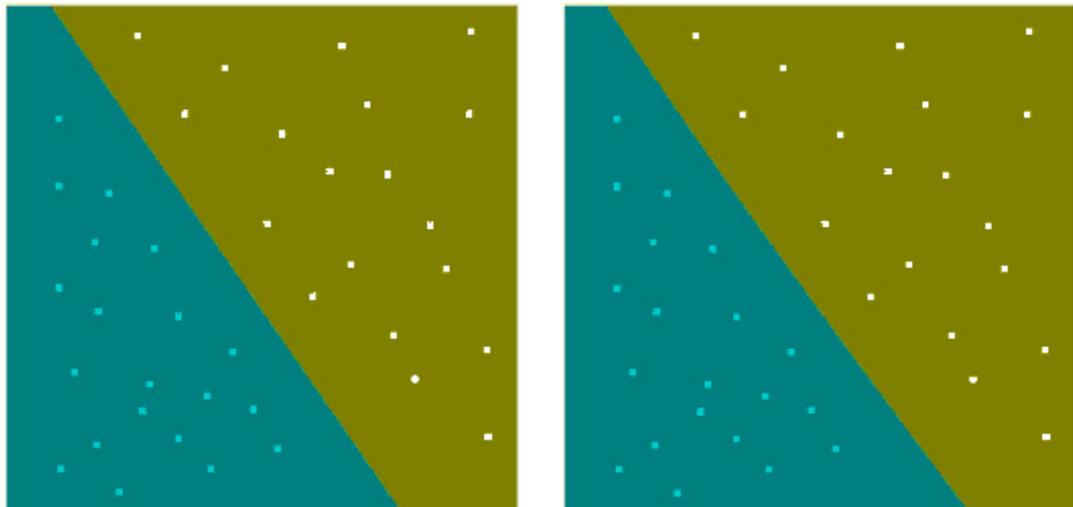
$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

と表現でき、識別関数は

$$f(x) = \text{sgn}\left(\sum_{i \in \text{sv}} \alpha_i y_i K(x_i, x) + b\right)$$

と表現できる。ここで、svはサポートベクタを表す添え字の集合である。

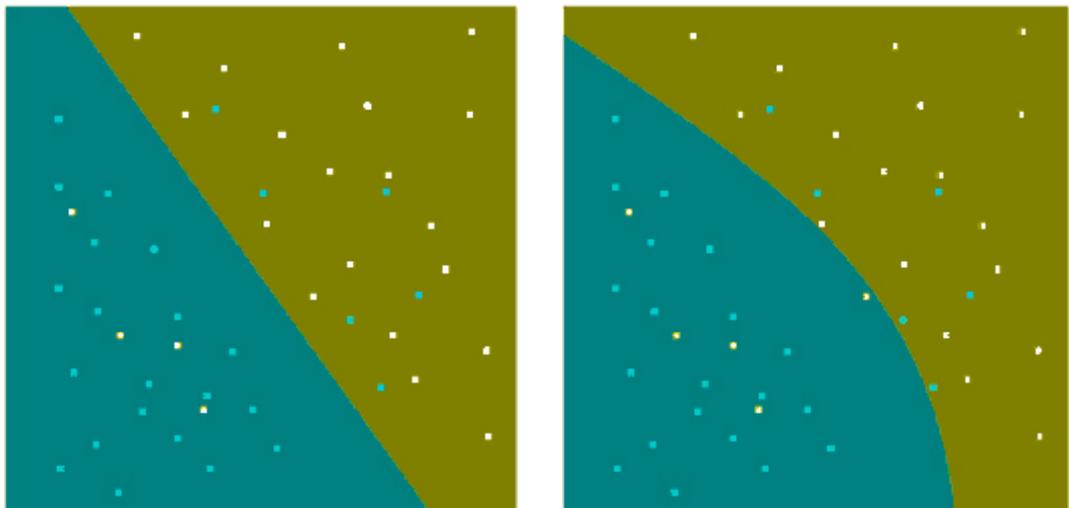
以上のように、SVMを利用した機械学習は、母集団の確率分布の知識を用いないパターン認識方式として、パターンの分布ではなく、パターン同士の境界を2クラスの分離に利用している。SVMの実行例として、LibSVM[89]による2次元領域の分類結果を図2-26に示す。図2-26は、2次元で分離可能なデータと不可能なデータに線形カーネルとRBFカーネルを利用した結果である。図2-26の2次元領域の分離に示すように、SVMは2クラス分類問題用の識別器なので、2クラス以上の判断が必要な場面では、SVMを組み合わせる必要がある。この組み合わせには、すでに多数の方式が提案されているが、未だ理論的な議論が続いている状態である[115][127]。



a. 線形カーネルの結果

b. RBFカーネルの結果

(1) 線形分離可能なデータ



a. 線形カーネルの結果

b. RBFカーネルの結果

(2) 線形分離不可能なデータ

図 2-26 SVM による分類結果 (LibSVM[89]を使用)

## 2.5.4 SVMによる多クラス分類

SVMは2クラス分類問題用の学習方式であり、標準正則化の観点から

$$f(x) = \sum_{i=1}^m d_i K(x, x_i)$$

と表記できる[130]。複数の対象を分類するにはSVMを組み合わせる必要がある。つまり、多クラス分類には上記の関数を

$$f_k(x) = \sum_{i=1}^m d_{ki} K(x, x_i)$$

のように多数作り、最大値を出力する $f_k(x)$ に対応するクラスを

$$m = \arg \max \{f_1(x), \dots, f_k(x)\}$$

として選ぶ方式などで対応できる。

そのようなSVMの組み合わせ方式は、1クラス対残りのクラスを学習する方式(図2-27a)と、各クラスを個別に比較する方式(図2-27b,cなど)に大分類できる[88][94][95][115]。図2-27はいずれも4クラス(A, B, C, D)の分類例を示している。本論文では図2-27a, b, cの各方式を図の書き方から便宜上、一対他方式、ボトムアップ方式、トップダウン方式と呼ぶことにする。また、分類したいクラスの数を $n$ として、各方式を説明する。なお、各方式の優劣は現在でも議論が続いている状況である[115][127]。

### (1) 一対他方式

図2-27aに示す一対他方式は、1つのクラスに対し残り全てのクラス( $n-1$ )を1クラスと見なしSVMによる学習を行う方式である。つまり、Aと残り全て(BCD)の間でSVMの学習を実行し超平面を決定する。残りB,C,Dに対しても同様の処理を行う。このタイプの分類法はOne vs. All othersとも呼ばれ、 $n$ クラスに対し $n$ 回の学習を行う必要がある。また、新規データの分類も、各トレーニング結果と新規データを比較するためSVMによる $n$ 回の判定が必要となる。ただし、 $n$ 回の判定の結果、複数のクラスが候補として残った場合、あるいはどのクラスにも判定されなかった場合、超平面からの距離や投票により最終的なクラスを選ぶ必要がある

### (2) ボトムアップ方式

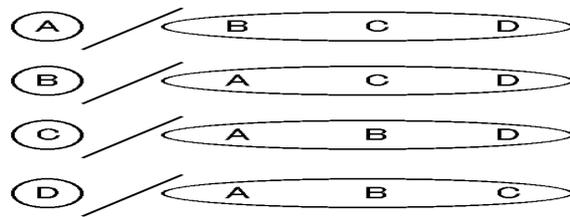
図2-27bに示すボトムアップ方式は、テニストーナメントのような勝ち抜き選抜方式であり、SVMによる学習を2クラス間で行う方式である。この方式は、あらかじめ全てのペアに対しトレーニングを行う必要があるため、 $n(n-1)/2$ 回のトレーニングが必要となる。新規データの判定も2クラス間で行い、最終結果を得るまでに $n-1$ 回の判定が必要になる。図2-27bに示すように、4クラスであれば2階層の木構造となる。この方式では、比較するクラス数を $n$ とすると、 $\log_2 n$ の階層ができる。

### (3) トップダウン方式

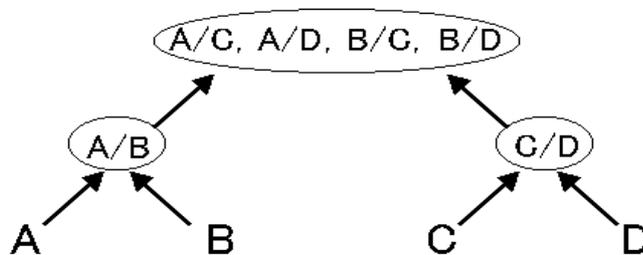
テニストーナメントのようなボトムアップ方式に比べ、図4-25cに示すトップダウン方式

が提案されている。ボトムアップ方式と同様にあらかじめ全てのペアに対し学習を行う必要がある。新規データの判定はグラフの最上位ノード(A/D：クラスAとクラスDの判定)から分類が実行され、判定されたクラスの方へ向かい処理が進められる。例えば、最上位でAが選ばれた場合、次の層ではA側のノード(A/C)が判定される。同処理を繰り返し最後に到達したクラスが認識結果となる。そのため、新規データの判定回数は $n-1$ 回となる。また、各階層の数がノード数と一致し、全体のノード数は $n(n-1)/2$ となる。

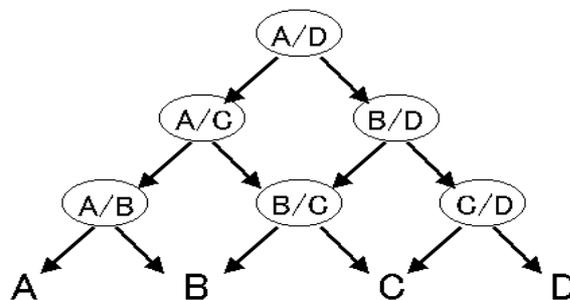
いずれのマッチング戦略もほぼ同様の学習回数を必要とするが、一対他方式よりも他の戦略のほうがSVの数が少なくなり、計算量を削減できる利点がある。これは、一対他方式は1度の学習に全てのサンプルデータを使うため、他の方式よりもSVの数が多くなる傾向があるためである。また、他の方式に比べ学習データ数が多くなるので、学習時間が最も長くなる。一方で、一対他方式は結果が常に一定だが、ボトムアップとトップダウンではマッチングの組み合わせや順番により、結果が異なる可能性がある。



a. 一対他による比較方式



b. ボトムアップによる比較方式



c. トップダウンによる比較方式

図2-27 SVMによる多クラス分類方式

## 2.5.5 AdaBoost の概要

マージンによる機械学習は上記のSVM以外にもBoostingやBaggingと呼ばれる方式がある。本論文では、3章でIntel社が公開しているOpenCVライブラリに含まれるAdaBoostによる機械学習を利用するため、以下でAdaBoostの概要を説明する。

AdaBoostは分類の難しい学習サンプルを集中的に学習し、すでに正しく分類した学習サンプルの重要性を下げる学習方法である。AdaBoostは学習仮説の重み付けに、学習の収束性が理論的に補償された方法を用いている。この方法を使えば、任意に重み付けされた学習サンプルに対して、学習誤差  $\epsilon_t$  が1/2未満の学習仮説を常に探索することができれば、有限の学習仮説生成回数で学習誤差0の仮説を生成できる。AdaBoostの学習アルゴリズムを図2-28に示す。

AdaBoostの学習結果は、SVMと同じように学習データを2分類している。AdaBoostは、前仮説  $h_t$  と重み  $\alpha_t$  の積を合計し、その値が重みの合計の半分以上なら、目的とした対象

**入力：** N 個のサンプル  $(x_1, y_1), \dots, (x_n, y_n)$ .  $y_i=0,1$  のいずれかである。

**初期化：** 重み  $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$  を設定する。  $m, l$  は  $y_i=0,1$  のそれぞれの合計数である。

**For**  $t=1, \dots, T$ :

1. 重みの正規化  $w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$
2. 各特徴  $j$  を使った分類  $h_j$  に対し、学習誤

差  $\epsilon_j = \sum_{i=1}^n w_i |h_j(x_i) - y_i|$  を求める。

3. 最も  $\epsilon_t$  の小さい分類  $h_t$  を選択する。

4. 重みを更新  $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$

ここで、 $x_i$  が正しく分類されたときは  $e_i = 0$  であり、誤って分類されたときは  $e_i = 1$  とする。ただし、

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$$

とする。

**出力：**  $\alpha_t = \log(1/\beta_t)$  として、最終的な学習結果を

$$h(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{その他の場合} \end{cases}$$

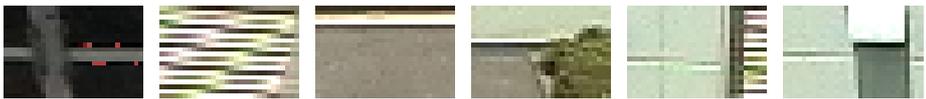
とする。

図 2-28 AdaBoost のアルゴリズム [61]

であると判定する。AdaBoost も SVM 同様、2クラス分類方式なので多クラスの分類では AdaBoost を組み合わせて利用する必要がある。

表 2-2 に示すような 120 枚の人物上半身画像 (30×20 画素) と、人物を含まない画像から切り出した 1,800 枚の背景画像 (30×20 画素) を上記の AdaBoost で学習し、新規撮影画像に学習結果を適用した結果が図 2-29 である。図 2-29 では人物以外に出口の庇部分なども誤検出している。本論文で利用する機械学習は、誤検出部分を学習事例として追加学習することで、誤りを減らすことができる方式である(詳細は 3.3 節参照)。

表 2-2 機械学習に利用した画像例

| 学習画像例 |  |
|-------|--|
| 人物画像  |  |
| 背景画像  |  |

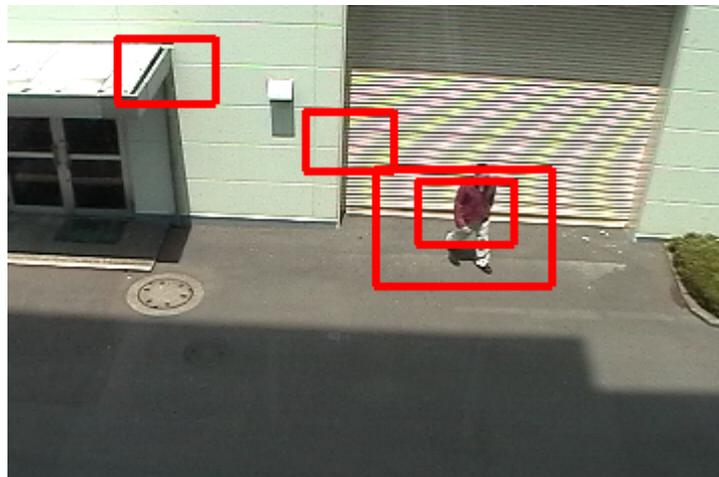


図 2-29 AdaBoost による歩行者検知例

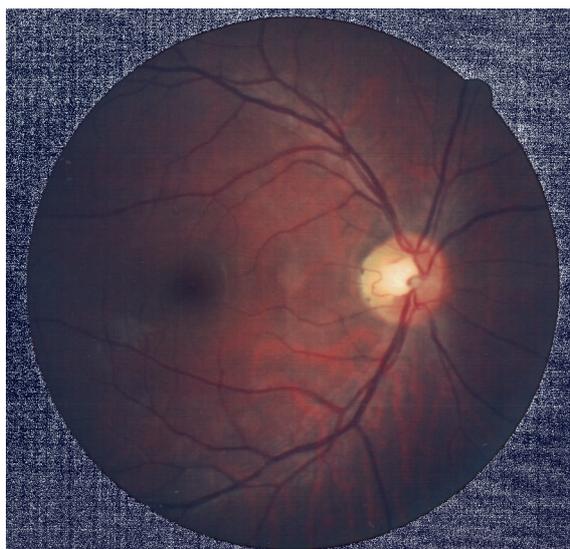


図 2-30 人の網膜 (著者の網膜写真)

## 2.6 視覚処理

1970 年代に D.Marr らが目視機能の研究と画像処理を結びつけて議論をはじめて以来、視覚の画像処理が議論されてきている。本節では、目視機能を代行する画像監視システムの作成に参考となる視覚処理の概要を示す。

### 2.6.1 視覚野モデル

生理学実験や臨床実験の知見から、視覚野の構造と機能の階層モデルが提案されている [18]-[20]。図 2-31 に提案されているモデルを簡略化した図を示す。

図 2-30 に示す網膜あるいは大脳の視覚野の構造は、DNA の設計図に従って整然と構成されている。網膜上で DNA に従い作成されるロドプシンと呼ばれるたんぱく質が、光の波長 (400nm~700nm の内、長波長、中波長、短波長) に反応し、光を信号に変えることから処理が始まり、それらの信号が大脳の視覚野へ伝達される [17]-[19]。網膜上では錐体とかん体と呼ばれる細胞中でロドプシンの反応が起きている。錐体は R G B の各波長に独立に反応する 3 種類の細胞があり、昼間の光で反応している。かん体は昼間の光では飽和状態であり、薄暗闇の中で反応する 1 種類の細胞である。そのため、我々の視覚は、薄暗闇でかん体反応による白黒画像処理となっている。監視カメラにたとえると、昼間はカラーカメラで撮影し、夜は感度の高いモノクロカメラで撮影している状態である。

網膜上の錐体あるいはかん体細胞で光が信号に置換され、その後、神経節細胞の X 型細胞と Y 型細胞と呼ばれる 2 種類の細胞により視覚野へ信号が伝達される。その途中で、各細胞の出力は外側膝状体(LGN)と呼ばれる場所で整理された後、1 次視覚野(V1)へ送られる。Y 型細胞の網膜上での受容領域が X 型細胞よりも広く、信号伝達速度も倍である。さらに Y 型細胞の信号を受ける外側膝状体の細胞も大きい。高次視覚野に向かうこの 2 種類の信号経路はそれぞれ M 経路 (大細胞系あるいは背側経路)、P 経路 (小細胞系あるいは腹側経路) と呼ばれ、その先の視覚野の機能が異なっている。M 経路は白黒画像を高速に処理

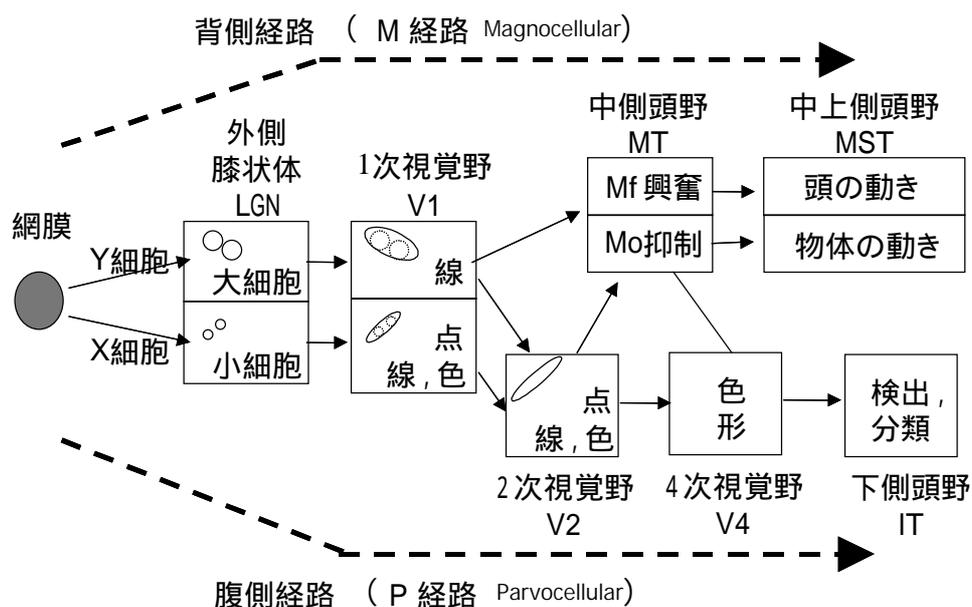


図 2-31 視覚野の構造と機能

する細胞集団なので以降本論文では動き処理経路と呼び、P 経路は色や形を扱う細胞集団なので認識処理経路と呼ぶ。動き処理経路と認識処理経路の配線が完全に分離されている状態ではなく、両経路の直接接続と階層をフィードバックした間接接続の2通りの接続関係になっている。ここでは議論を単純化するため、直接接続でかつ順方向の情報伝達として以下を説明する。なお、監視カメラにたとえると、カラー画像と白黒画像をあるメモリ（外側膝状体相当）に別々に保存し、その先の高速な動き処理経路と色や形による認識経路に個別に信号を分配していると思えることができる。

## 2.6.2 動き処理経路での MT 野の機能

動き処理経路の V1 野では画像内の局所的な動きを解析し、その結果を受ける中側頭野 (MT 野) では画像の中局所領域での動きを解析している [18]-[20]。さらに MT 野の出力を受ける中上側頭野 (MST 野) では画像全体の動きから頭部の動きを検出する細胞と、背景と独立に移動する物体の存在を判定する細胞が占めている。

MST 野ではオプティカルフロー式の回転、拡大/縮小、平行移動項に相当する機能を持つ細胞が集中している。この MST 野の網膜上の受容領域が広く、移動物体の場所を特定できる状態ではない。MT 野には受容領域内で動きがあるときに反応する Mf 型細胞と、図 2-32 のように受容領域の中心と周辺が異なる動きの時に反応する Mo 型細胞が存在する [90]。頭部の動きを検出するには Mf 型細胞の機能が適しており、動いている物体を検出するには Mo 型細胞の機能が適している。

画像処理の観点からみると、MT 野では色情報を省略した白黒画像で高速に動きを処理していることが分かる。特に、自己の動きで発生する画面内の動きと、背景内で独立に動く物体を検出する細胞が独立に存在していることになる。本章の 3.3 節で示す旋回中の監

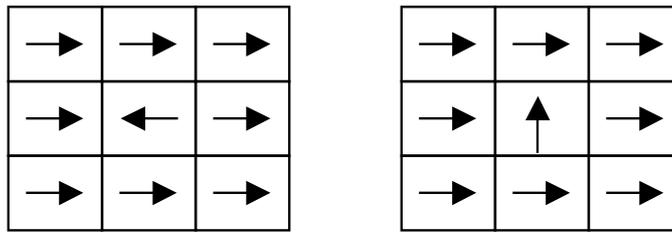


図 2-32 MT 野の Mo 型細胞が反応する動きの組合せ例

視カメラの画像から移動体を検出する方式の開発では、この MT 野 Mo 型細胞の機能を参考としている。

### 2.6.3 認識処理経路での V4・IT 野の機能

認識処理経路の V1 野と V2 野では画像内の局所的な点、線、色などを分離して処理している。V2 野から信号を受ける 4 次視覚野(V4 野)以降の詳細な機能の解明は不十分だが、V4 野は色と形を解析していること、その先の下側頭野 (IT 野) は物の識別に強く関与していることが分かっている[18]-[20]。顔に反応する細胞が集まっているのも IT 野である。

V4 野は視覚野の全体的な画像配置を維持しており、色の知覚や色の恒常性に寄与している。その先の IT 野では、検出や分類を行う細胞が群をなして集まっている。そのため、画像監視システム用の画像処理の観点から見ると、V4 野でサンプル画像の色などの正規化を行い、IT 野で正規化画像の機械学習結果が蓄積されていると見なすこともできる。特に、動き処理経路と認識経路では、動き処理経路が全般的に処理が速いため、動き処理が認識処理を行う場所を絞り込む前処理になっている可能性がある。

### 2.6.4 画像処理との関連

視覚野の低次視覚と高次視覚の主要な働きが解明されているが、その中間の V4 野の働きや、MT 野と V4 野などの相互接続の役割解明は不十分であり、システム化の観点から検討を進める必要がある。

#### (1) 低次視覚野の画像処理

網膜から V2 野までの低次視覚野を監視カメラにたとえると、網膜が CCD などの撮影素子であり、LGN が撮影素子の電気信号を蓄えるバッファである。その先の、V1 野や V2 野は画像処理フィルタによる線分、点、色などの検出処理と処理結果の保存メモリと見なせる。また、撮影素子のデータ収集段階から、白黒で動き処理を高速に行う経路と、カラー画像で時間をかけて物体認識等を行う経路の 2 つに処理が分かれ、各経路の上位に情報を送信していると思なせる。

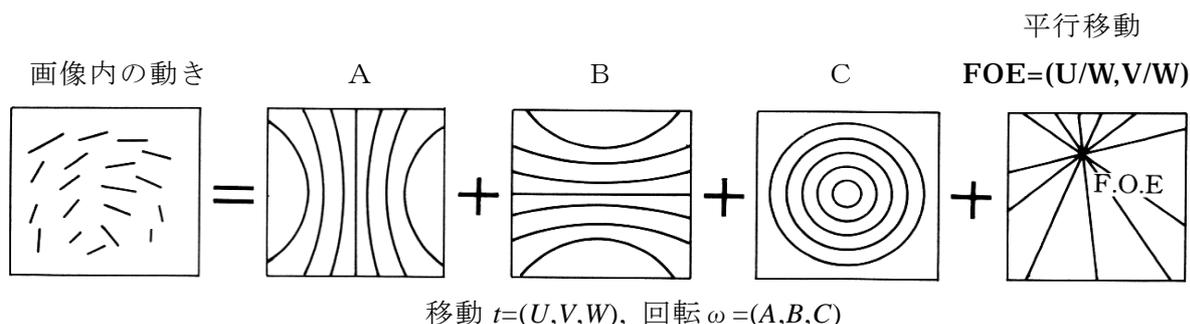


図 2-33 オプティカルフローと移動,回転の関係 (文献[76])

## (2) 動き処理経路での高次視覚野の画像処理

動き経路の到達先である MST 野は、頭の動きと物体の動きを処理する 2 領域に分かれている。画像内の座標  $(x,y)$  でのオプティカルフローを  $(u,v)$  とし、目を原点とした座標系で見ている物体  $P$  の座標を  $(X,Y,Z)$ 、目の平行移動を  $t=(U,V,W)$ 、回転を  $\omega=(A,B,C)$  とすると、画像内の座標とオプティカルフローの関係は、

$$u = Axy - B(x^2 + 1) + Cy + \frac{-U + xW}{Z}$$

$$v = A(y^2 + 1) - Bxy - Cx + \frac{-V + yW}{Z}$$

となる。MST 野には、図 2-33 のような回転  $A,B,C$ 、移動  $(U/Z,V/Z)$  に相当する動きだけに選択的に反応する細胞が集まっている領域がある[90]。そのため、MST 野では上記式を細胞レベルの働きで実現していると見なす事ができる。近年、この MST 野を参考に動きを抽出する神経回路モデル[137]なども提案されている。

## (3) 認識処理経路での高次視覚野の画像処理

見たものを認識したり分類したりする視覚野の認識処理経路は、形や色などにより事例画像を学習する機械学習の枠組みでモデル化できる[124]。特に、IT 野等での顔や特定の物体に選択的に反応する細胞の働きは、標準正則化理論で説明が可能である[119]。なお、本章で示した機械学習はいずれも標準正則化の枠組みで説明が可能であり、SVM などはその 1 つの表現形式にすぎない。下記に Tikhonov の標準正則化を示す。

視覚処理はデータ  $x$  が与えられたとき、正解  $y$  を出力する関数  $f$  を求める問題として

$$y = f(x)$$

と表現できる。しかし、この問題はデータ  $x$  から関数  $f$  を求める逆問題(inverse problem)であり、一般に不良設定問題(ill-posed problem)である。Tikhonov の標準正則化では、この逆問題を適当なパラメータ  $\lambda$  を導入し 2 次形式で

$$\|y - f(x)\|^2 + \lambda \|f(x)\|^2$$

を最小にする関数  $f$  を求める問題におきかえる。この標準正則化理論は、2.5 節で説明した図 2-22 の教師付き機械学習モデル実現方法の一つである。SVM やベイズ識別あるいはニ

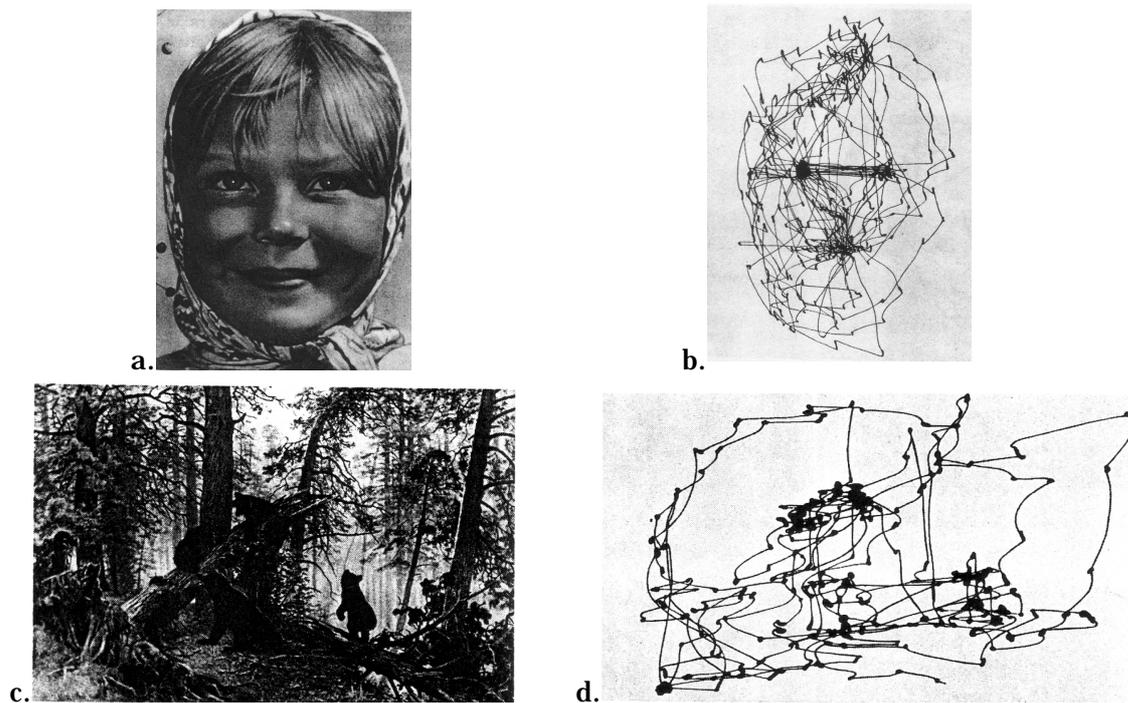


図 2-34 注視位置の変化（文献[125]より抜粋引用）  
 写真 a と c を見たときの注視点の移動軌跡 b と d

ューラルネットなどはいずれも標準正則化理論の枠組みと関連付けできる。

#### （４）処理の連携

オプティカルフローや機械学習などの研究成果により、個々の視覚領野が行っている処理の数理モデルが明確化されてきている。一方、脳では動き処理経路と認識処理経路の連携が図られているが、その連携の役割や工学応用方法は明確にされておらず、今後研究が必要な分野である。

例えば、動き処理経路の **MT** 野と、認識経路の **V4** 野では双方向の接続があるが、両経路の処理速度が違うため、**MT** 野で先に移動体領域を検出し、その結果を **V4** 野へ送り注目領域を認識している可能性がある。このような **MT** 野から **V4** 野への接続は、認識処理の高速化の観点から工学的な利点がある。別な例として、人が画像を見ると注視位置が常に変化するサッケードと呼ばれる眼球運動が生じる。図 2-34a,c のような画像を見ると、視線は無意識のうちに図 2-34b,d のような動きをする[125]。人の視覚は、無意識のうちに目や動物がいる重要と思われる部分に選択的かつ高速に注意を向けている。これまで、ボトムアップ的な特徴選択を組み合わせた注視のモデルが多数提案[126][131]されているが、学術的議論が先行しており、実時間監視に使える成果は得られていない。今後、視覚の注視行動を実時間で実現できれば、監視員による目視を代行する画像処理の実現に寄与する可能性がある。

個別視覚領野の機能解明やモデル化が進んでいるが、それらの機能連携の解明は未だ不十分であり、工学応用に関してはほとんど議論されていない状態である。今後、視覚領野

の連携の役割解明に、画像処理のシステム化の観点が発立つ可能性がある。また、逆に個々の視覚領野の接続関係から、シーン理解のブレークスルーに繋がるシステム化方式が発見される可能性もあるので、今後の進展を期待したい研究分野である。

## 2.7 本章のまとめ

本章では画像監視システムに活用できる画像処理を、画像内の変化領域の検出・認識・処理速度の観点から紹介し、画像監視システムの開発に適用する際の課題を示した。また、本研究で利用する機械学習の概要を示し、教師付学習の理論的な限界を中心に説明した。特に、次章以降で利用する SVM、 $k$ -NN、AdaBoost などの認識方式を説明した。さらに、人の脳細胞の処理速度は現在市販されている計算機よりも遅いが、膨大な数の細胞を構造化し機能分化することで、既存の画像処理による人物検出よりも高速で汎用的な処理を実現していることを示した。また、目視機能を代行する画像監視システムの設計に参考となる視覚野の構造と機能を説明した。特に、脳では動きによる移動体領域の絞り込みを先に行い、その後、移動体領域の認識を行っていると考えられる。そこで、動きによる絞り込みと、その絞り込んだ領域の認識処理を行うことで、高速な侵入者検知が実現できる可能性を示した。

本論文に直接関係する画像処理の課題として、人物検出では機械学習による汎用的な検出方式が提案されているが、現在の市販パソコンを使い実時間で動作する処理が実現できていないことを示した。検出人物の姿勢認識では、身体モデルを当てはめる研究が進んでいるのに比べ、人物の様々な姿勢を定義することが困難なため人物検出と同様の機械学習の枠組みで姿勢認識を行う研究成果が少ないことを示した。さらに、検出した人物の認識では、個人を特定する安定した特徴が得られにくいことから、全身画像を利用した人物認識が行われていないことを示した。

## 第3章

# 固定カメラと旋回カメラによる人物検出の高速化

### 3.1 はじめに

目視監視を補助あるいは代行する画像監視システムには、侵入者などを実時間で検出する機能が必要である。しかし、2章の関連研究で説明したように現在の汎用的な SVM などの機械学習を用いた人物検出は、処理速度が遅い問題がある。そこで本章では、SVM などを使う既存の人物検出を実時間で動作させる方式を提案する。

#### (1) 人物検出の高速化と課題

SVM などの機械学習を利用した人物検出の高速化方式として、以下の3通りの代表的な方式がある。

##### ① 特徴選択による高速化

人物検出に適した特徴だけを画像から選択的に抽出し、検出に利用する特徴数を減らすことで処理の高速化を図る。

##### ② 階層化による高速化

画像を縮小するなどして画像内を最初に荒く検索し、人物領域候補を絞り込む。徐々に解析精度をあげ人物領域を絞り込む処理を繰り返し、最後に残った部分を人物領域とする。顔検出などのように最初に「目」を見つけその近傍で顔を探索する方式など、はじめに特徴的な部分を見つけ探索範囲を絞り込む方式も多い。

##### ③ 監視領域を限定した高速化

画像内の監視領域以外にマスクをかけ検出処理を省略する。あるいは、マスク領域だけに検出処理を行う方式である。または、画像内の動き変化のある領域だけに人物検出を行うことで高速化を図る方式である。

特徴選択による高速化や階層化による高速化がすでに多数提案されているが、実時間で高速な検出は実現できていない。一例として、特徴選択と階層化による顔検出の高速化を図った実験では、 $320 \times 240$  画像の処理に 4fps 程度かかっている[132]。また、2章で示したが、動き変化領域だけに着目した機械学習による人物検出も、その処理速度は現在のところ約 4fps である[61]。将来的に、計算機の高速化により検出時間が短縮できると期待できるが、現時点では実応用の観点から処理の高速化を検討する必要がある。

#### (2) 具体的な監視課題

電気所構内には多数の監視カメラが設置されており、それらの監視カメラの活用法として、構内での充電部接近監視や立ち入り禁止領域での侵入監視などが検討されている[108][109]。特に、充電部近辺での作業員の安全監視は、不用意な充電部接近が事故につな

がるため実時間監視が必要である（設備工事等では監視員を配置して、作業員の安全を図っている）。監視員に代わり充電部への接近を監視し事故を防ぐには、既方式よりも高速な人物検出が必要となる。

以下、本章では既設監視カメラに接続している計算機で、計算量をかけずに充電部などの特定監視領域に接近する人物を高速に検出する方式を提案する。提案方式は、固定型の監視カメラ用と、旋回型の監視カメラ用の2方式を別々に提案する。固定カメラ用では監視領域を特定の線上に限定するライン型前処理を示し、旋回カメラ用では移動中の人物が周囲と異なる動きをすることに着目する側抑制型前処理を示す。いずれの方式も、はじめに人物の可能性が高い部分を絞り込み、次に絞り込んだ部分を機械学習による認識を行い実時間で人物を検出する方式となっている。

3.2 節で固定カメラ用の人物検出の高速化を提案し、3.3 節で旋回カメラ用の人物検出方式を提案する。3.4 節でまとめを示す。なお、本章では機械学習によるパターン認識で画像内から人物を検出する処理を侵入者検知と呼び、画像処理による前処理で人物候補領域を検出する処理と区別する。

## 3.2 固定カメラによる人物検出の高速化

本節では、固定監視カメラ用の実時間人物検出方式を提案する。3.2.1 で監視環境と監視対象を示し、3.2.2 で提案方式について述べる。3.3.3 で提案方式の実験結果と考察を示し、3.2.4 で本節の結論を述べる。

### 3.2.1 監視環境と監視対象

画像監視形態として電気所構内を俯瞰撮影した監視カメラ画像により、充電部への侵入者接近を実時間で監視することを想定する。ビデオレート(30fps)で取込む監視画像（720×480画素）の中で、侵入者サイズを縦横30×25画素から120×60画素程度と仮定する。また、歩行による移動速度を、画像間で数画素から15画素以下と仮定する。仮に、画像内のあるポイントを通過する侵入者の横幅を60画素、移動速度を15画素/画像とすると、侵入者がそのポイントを通過する4画像以内に検出するには、概ね7.5fps以上の処理速度が必要となる。なお、監視カメラは固定とし、ズームやフォーカスなども固定とする。

### 3.2.2 ライン型前処理による歩行者の高速な検出処理

#### (1) 処理概要

機械学習による既存の人物検出方式は画像内の変化を面として処理するため、解析処理に時間がかかる。そこで、提案方式では充電部領域が既知であることに着目し、監視領域を面から線（以降、この線のことを監視ラインと呼ぶ。）へ絞り込むことで処理の大幅な高速化を図る。提案方式の処理概要を図3-1に示す。なお、他と区別するために本提案方式をライン型前処理と呼ぶことにする。

図3-1に示す「ライン監視」処理では、監視画像内に設定した監視ライン上の各画素値

の変化に着目し、時系列変化が大きい点を抽出する。ライン監視処理で抽出した変化点全てに SVM による侵入者検知を行うと、実時間処理ができない可能性がある。そこで、監視画像を特定の区画に分割し、人物の可能性が高い部分を「優先区画決定」処理で絞り込み、可能性の高い部分のみに「侵入者検知」処理を行い処理の高速化を図る。なお、侵入者検知処理で優先順位の高い区画に人物が写っていないと判断した場合、ある重みを伝播（「重み更新」）し、次回の優先区画決定処理で区画の順位を下げ無用な侵入者検知処理をさける方式としている。

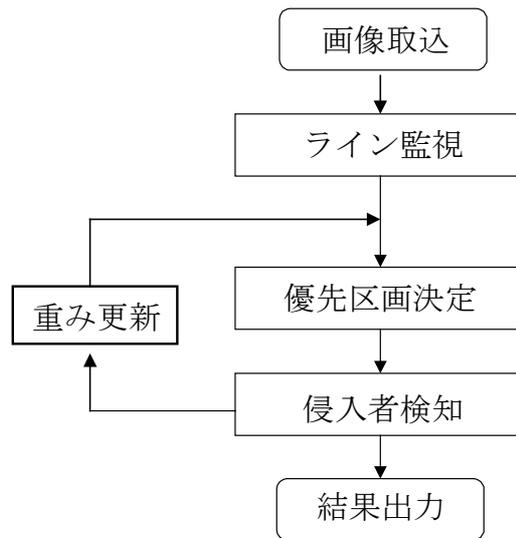


図 3-1 ライン型前処理による監視処理の概要



図 3-2 監視ラインの設定例

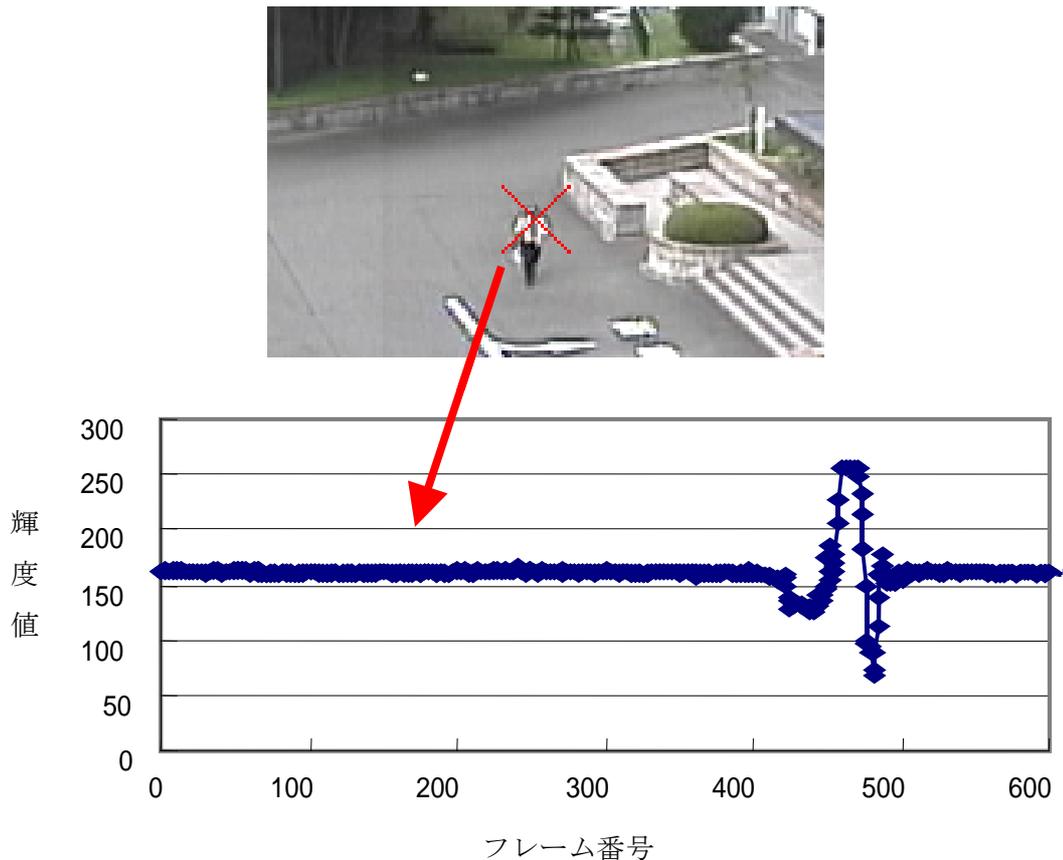


図 3-3 人物の通過に伴う輝度値の時間変化例

以下、「ライン監視」処理、「優先区画決定」処理と「侵入者検知」処理の詳細を示す。

## (2) ライン監視処理

ライン監視処理では利用者が画像内に設定した監視ライン上の変化点を抽出する。

### a. 監視ライン設定

利用者が監視画像内で充電部の周囲に監視ラインを設定する。監視ライン数とその間隔は監視対象や監視環境に依存するため、状況に応じて利用者が任意に設定する。図 3-2 に作業現場を囲むように設定した 3 本の監視ライン例を示す。

### b. 監視ライン上の変化点検出

ライン監視処理では、監視ライン上の各画素を独立に処理し、輝度値が過去の時系列変化に比べ大きく変化した画素を抽出する。例として、図 3-3 画像内の「×点」部分の輝度値（約 600 フレーム分）の時系列変化を図 3-3 のグラフに示す。図 3-3 のグラフは約 400～500 フレームにかけて着目領域を人物が通過している様子を示している。人物が通過するとき(400～500 フレーム)は輝度値が大きく変化（100 画像分の輝度差の標準偏差 11.4）するが、それ以外では平均輝度値を前後しながら微小変化（100 画像分の輝度差の標準偏差 1.4）している様子が分かる。

そこで、監視ライン上の画素の輝度値が大きく変化した点を、各画素での輝度値の標準偏差を使う  $\sigma$  バンドで抽出する。監視ライン上の各画素  $i$  での輝度値変化の有無  $P_i^k$  を、標準偏差  $\sigma$  を使い、

$$P_i^k = \begin{cases} 1 & \text{if } \sqrt{(M_i^{k-1} - I_i^k)^2} > \alpha \sigma_i^{k-1} \\ 0 & \text{else} \end{cases}$$

とする。ここで  $k$  は  $k$  番目の新規入力画像を示し、 $M_i^{k-1}$  と  $\sigma_i^{k-1}$  は 1 画像前までの各画素  $i$  の輝度値平均と標準偏差を示す。各画素  $i$  の新規入力値  $I_i^k$  と 1 画像前までの平均  $M_i^{k-1}$  との差が標準偏差  $\sigma_i^{k-1}$  の一定倍率  $\alpha$  以内に収まるかで変化の有無  $P_i^k$  を判定する。本節で示す実験は、いずれも事前実験で良好な結果を得ている  $\alpha = 4.0$  とした時の結果である。

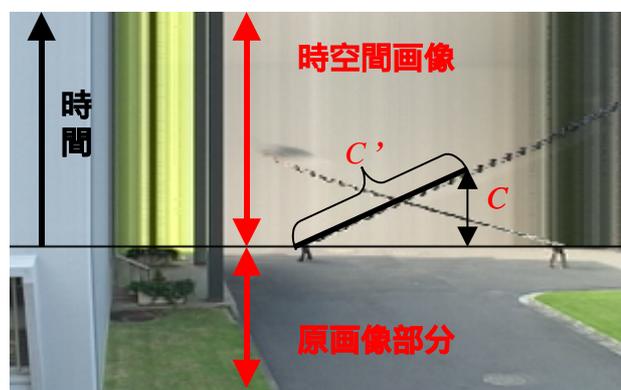
ただし、画像内に発生する雑音により単発的な輝度値変化が発生するので、監視ライン上で  $C$  画像以上連続的に変化が続いた時のみ監視ラインに進入物体が現れたものと判断する。以下、変化の判定基準となる  $C$  を「変化判定画像数」と呼ぶ。本実験では初期探索が 7.5fps 以上の処理速度で侵入者検知ができるように  $C=4$  とした。

### (3) 優先区画決定処理

監視ライン上の変化点全てにパターン認識による侵入者検知処理を行うと実時間処理ができない。そこで、監視画像を  $m$  画素単位の  $p \times q$  個の区画  $R_j$  ( $j=1, \dots, p \times q$ ) に分割し、移動中の人物がいる可能性の高い区画を見つけ、優先的に侵入者検知処理を行う。



元画像



時空間画像断面 (x-t 断面)

図 3-4 時空間画像の例

優先区画決定処理では、移動速度の速い人物を見逃さないように、各区画  $R_j$  内に速い移動物体が含まれる順に優先順位を決定する。この処理では、まず監視ライン上の各画素  $i$  に対する重み  $\omega_i^k$  を求め、次に区画  $R_j$  内で最大値を示す  $\omega_i^k$  によって区画の値を決め、その後区画値の大きい順に優先順位を決定する。ここで  $k$  は現在の入力画像番号を示す。下記に優先区画決定処理の手順を示す。

#### a. 時空間画像の作成

監視ライン上で図 3-4 に示すような時空間画像を作成する。図 3-4 は説明を容易にするため、画像内の特定の高さで時空間画像を作成した例である（2章図 2-3 参照）。時空間画像上で、移動人物は傾きを持つ線分として描かれ、動かない設備などは垂線となる。特に、監視ラインにある角度で人物が入る場合、時空間画像内のエッジの傾きを調べれば画像内の人物のおおまかな移動速度がわかる。

#### b. 変化の傾きを計測

時空間画像で移動物体が作り出す線分の傾き角を  $\theta_i$  とする。監視ラインへの入力角度が一定の時、監視ライン上での移動速度が速いと傾きが大きくなり、移動速度が遅いと傾きが小さくなる(図 3-4 参照)。そこで、監視ラインの各画素  $i$  の時空間画像内での線分の傾き  $\theta_i$  に比例する変数  $\lambda_i^k$  を

$$\lambda_i^k = \beta_i^{k-1} \frac{1}{\sin \theta} = \beta_i^{k-1} \frac{C_i'}{C}$$

とする。ここで  $C$  は変化判定画像数、 $C_i'$  は変化判定画像数  $C$  画像内での移動軌跡長( $\geq C$ )である（ $\beta$  は下記 d を参照）。ただし、 $C_i' = C$  は監視ラインに垂直に進入する場合である。

#### c. 優先区画決定

監視ライン上の各画素の重みを次式で更新する。

$$\lambda_i^k \neq 0 \Rightarrow \omega_i^k = \min(\lambda_i^k, 2.0)$$

特に、 $\lambda_i^k = 0$  の部分は人物が立ち止ったために発生する場合もあるので、時系列ゲイン用パラメータ  $\gamma$  を使い

$$\lambda_i^k = 0 \Rightarrow \omega_i^k = \max(\omega_i^{k-1} - \gamma, 0)$$

とし、短時間の立ち止りに対応している。

画像を  $m$  画素単位に区画分けし、各区画  $R_j$  内での重みの最大値  $\omega^k$  を各区画に設定する。全区画の中から最大値を持つ区画  $R_{max}$  を見つけ、その区画  $R_{max}$  に侵入者検知処理を行う。

#### d. 重み係数の更新

侵入者検知処理の結果を対応画素の重み計数  $\beta$  へ反映させる。提案方式では下記の値を使い、侵入者検知処理が人物でないと判定した区画の優先順位を下げる設定としている。

$$\beta_i^k = \begin{cases} 1.0 & : \text{人物と判定したとき} \\ 0.5 & : \text{人物でないと判定したとき} \end{cases}$$

#### (4) 背景と人物の判定処理

侵入者検知処理はパターン認識により人物画像と背景画像を区別する。本節では文献[58][59][60]を参考に、人物上半身画像と人物以外の2種類の画像を学習用データとしてSVMに与える。SVMは機械学習方式の1つであり、SVMでの侵入者検知関数 $f(x)$ は

$$f(x) = \sum_{i=1}^m d_i K(x, x_i)$$

と表記できる。ここで $x_i$ が事例画像、 $x$ が新たな入力画像、 $d_i$ はパラメータ、 $K$ は事例画像と入力画像の距離を計算するカーネル関数を示す。侵入者検知関数 $f(x)$ の符号により人物の可否を決定する。

入力画像 $x$ のわずかな変化に影響を受けないように、画像をHaarウェーブレットで変換した画像特徴をSVMに与えるパターンとする。本実験でのSVMによる侵入者検知処理は画像内の侵入者検知サイズを $60 \times 40$ 画素とし、 $60 \times 40$ 画素内を10画素ずつ飛ばしながらウェーブレット( $8 \times 8$ 画像のHaarウェーブレット: 図2-12参照)で計算した72次元( $6 \times 4 \times 3$ )の特徴をSVMでの侵入者検知処理に用いる。

### 3.2.3 実験結果

実験では3種類の撮影画像を利用し、処理速度、不特定多数の侵入者検知精度、充電部接近監視精度を示す。なお、本実験は画像( $720 \times 480$ 画素)をIEEE1394で読み込み、1GBのメモリを搭載した計算機Pentium4, 2.4GHzで実施した結果である。

表3-1に示す3種類の撮影画像を実験に利用する。各画像を使った実験目的は以下の通りである。

#### a. 処理速度計測

表3-1(A)では提案方式の処理速度を調べるため、設備近傍を歩行する1名の被験者の検知速度を計測する。

#### b. 不特定多数の侵入者検知精度計測

表3-1(B)では撮影画像15分の前半部分(7分)で収集した人物画像を使い、後半8分の不特定多数の侵入者検知結果を示す。特に、構内では設備以外にアスファルトや花壇が背景になる場合があるため、背景の違いによる侵入者検知結果を示す。

#### c. 充電部接近監視精度計測

表3-1(C)では、状況変化に対する頑健性を調べるため、表3-1(B)で学習したシステムを用いて電力設備周辺での接近監視を行った結果を示す。

#### (1) 処理速度計測

表3-1(A)に示す約1分4秒(1,900画像)の侵入者検知実験により、提案方式の処理速度を計測した。

**a. 実験画像**

表 3-1 (A)は曇りの日に撮影した照度変化の少ない画像であり、1名の被験者が電力設備に接近したり離れたりを繰り返すシーンで構成されている。

処理速度計測実験で利用した各種パラメータを表 3-2 に示す。侵入者検知処理の学習に使う人物画像は、表 3-1 (A)の約 1 分のビデオ画像から切り出した 551 画像を利用した。図 3-5 に学習に使用した人物画像例を示す。また、学習に利用する人物画像以外の背景画像は、表 3-1 (A)の無人背景から切り出した 2,838 画像を利用した。監視ラインは図 3-6 に示すように、画像内の電力設備フェンスを囲むように 3 本設定した。

表 3-1 実験に利用する画像例

| No<br>撮影時間               | ビデオ画像  |
|--------------------------|--|
| (A)<br><br>約 1 分         |   |
| (B)<br><br>約 15 分        |  |
| (C)<br><br>約 3 分<br>20 秒 |  |

表 3-2 設定パラメータ

|          | パラメータ           | 値     |
|----------|-----------------|-------|
| ライン監視処理  | 平均と標準偏差の計算用画像数  | 60 画像 |
|          | 変化判定画像数 C       | 4 画像  |
| 優先区画決定処理 | 時系列ゲイン $\gamma$ | 0.2   |
|          | 区画 R の画素数 m     | 72 画素 |



図 3-5 学習に利用した人物画像例

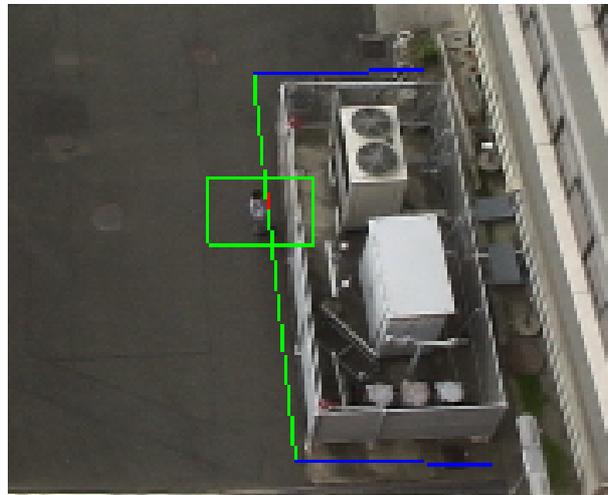


図 3-6 提案方式での人物の検出例

## b. 実験結果

図 3-6 に人物の検出例を示す。図 3-6 では、システムが自動的に検出した領域を長方形で示している。図 3-7 に撮影画像全シーンに対する侵入者検知の時系列結果を示す。図 3-7 のグラフ上側はライン監視処理が変化を検知した画像番号を示し、下側は侵入者検知処理が人物を検知した画像番号を示している。SVM による侵入者検知の処理速度は特徴算出も含め 1 箇所あたり約 0.01 秒であった。線形カーネルを使った SVM の学習率 (5-fold cross validation の検証) 等を表 3-3 に示す。SVM に非線形の RBF カーネルを用いた比較実験を行ったところ、RBF カーネルで学習率が 98.5% と 0.8% の学習率向上結果が得られたが、サポートベクタ数 (SV 数) が合計で 438 となり、線形 SVM の 4 倍となった。線形 SVM と RBF カーネルによる SVM が仮に同一の SV 数であっても、RBF カーネルは計算量がかかることから、計算量を削減するために線形 SVM の利用が妥当であることが実験で分かった。

表 3-3 人物 551 画像、背景 2,838 画像の SVM 学習結果

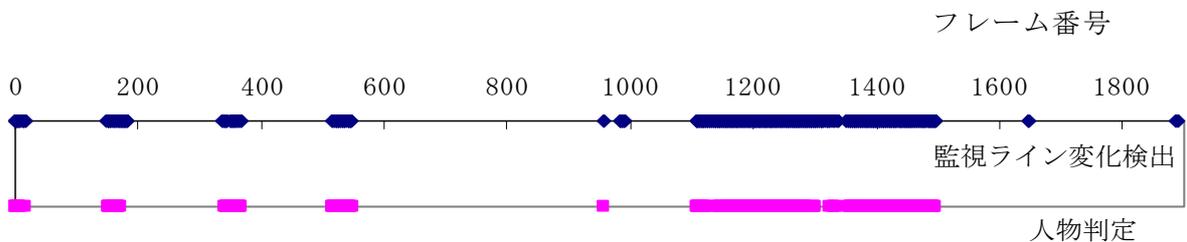
| 項目                      | 人物    | 背景   |
|-------------------------|-------|------|
| 学習画像数                   | 551   | 2838 |
| Support Vector 数        | 50    | 57   |
| 5-fold cross validation | 97.7% |      |

表 3-4 提案方式の反応結果 (1,900 画像)

|           | 正解反応                       | 誤反応   | 未反応                     |
|-----------|----------------------------|-------|-------------------------|
| ライン監視処理結果 | 458 画像                     | 12 画像 | 12 画像                   |
| 侵入者検知処理結果 | 428 画像<br>93%<br>(428/458) | 0 画像  | 30 画像<br>7%<br>(30/458) |
| 最終結果      | 97%<br>(1858/1900)         | 0%    | 3%<br>(42/1900)         |

表 3-5 1,900 画像に対する実験結果の要約

|             | 正解反応                  | 誤反応 | 未反応 |
|-------------|-----------------------|-----|-----|
| ライン監視処理結果概要 | 6 件                   | 3 件 | 0 件 |
| 侵入者検知処理結果概要 | 6 件                   | 0 件 | 0 件 |
| 最終結果概要      | 監視ラインの通過を見逃すことなく検知した。 |     |     |



横軸は連続画像のフレーム番号であり、上段は監視ライン検出が反応を示したフレーム番号、下段は上段の反応の中で侵入者検知が反応を示したフレーム番号を示している。

図 3-7 提案方式による侵入者検知結果

提案方式は約 1 分 4 秒の撮影画像 (約 1,900 画像) を実時間処理 (約 30fps) できる結果となった。1,900 画像に対する侵入者検知の正解率等を表 3-4 に示す。表 3-4 のライン監視処理欄は、侵入者による変化を正しく検知した画像数 (正解反応)、侵入者以外で誤反応を示した画像数 (誤反応) と、侵入者の通過中に反応を示さなかった画像数 (未反応) を示す。侵入者検知処理欄は、ライン監視処理に反応した 458 画像に対し侵入者検知処理を行った結果を示している。最終結果欄では、侵入者が現れない画像を含め全 1,900 画像での提案方式の反応割合を示している。

表 3-5 に人物が監視ラインに入ってから出るまでに一度でも反応があれば正解、監視ライン内に入ってから出るまでに 1 度も反応がなければ未反応、人物以外の要因で反応した場合を誤反応として集計した結果を示す。以降、この集計を「実験の要約結果」と呼ぶ。

実験の結果、提案方式により実時間で侵入者検知ができることが分かった。また、実験画像の中から抽出した学習用の人物画像と背景画像を使うことで、侵入者を高精度に検知できることが分かった。

## (2) 不特定多数の侵入者検知

上記ではビデオ画像全体から切り出した同一人物画像を学習事例として利用したが、ここでは別人物の画像を利用した不特定多数の侵入者検知実験結果を示す。特に、構内監視ではアスファルトのような人工的な部分や樹木を背景とする場合が多いため、背景の違いに対する実験結果もあわせて示す。

### a. 実験画像

構内を撮影した表 3-1(B)に示す 15 分の画像を実験に利用する。この画像には、太陽が雲で隠れるなどの大きい照度変化が多数含まれている。

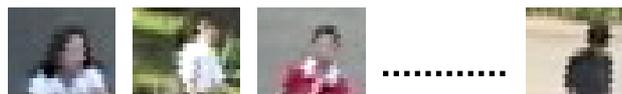


図 3-8 学習に利用した人物画像例（多人数）

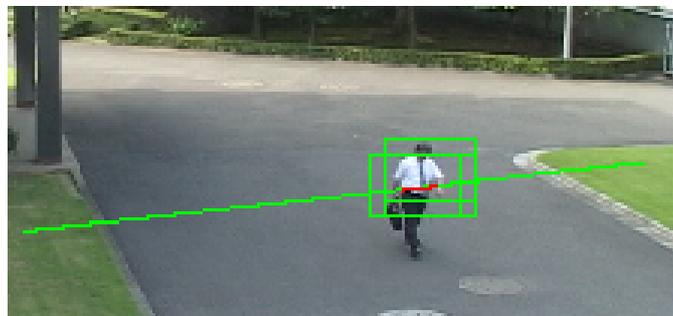


図 3-9 アスファルト上での監視結果例

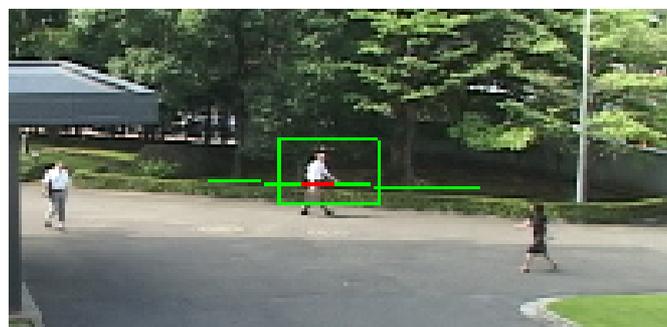


図 3-10 樹木の前での監視結果例

表 3-6 人物 2,410 画像、背景 9,281 画像の SVM 学習結果

|                         |       |      |
|-------------------------|-------|------|
|                         | 人物    | 背景   |
| 学習数                     | 2410  | 9281 |
| Support Vector 数        | 525   | 532  |
| 5-fold cross validation | 96.5% |      |

表 3-7 アスファルト背景時の監視結果 (14,400 画像)

|           |                            |        |                         |
|-----------|----------------------------|--------|-------------------------|
|           | 正解反応                       | 誤反応    | 未反応                     |
| ライン監視処理結果 | 454 画像                     | 155 画像 | 0 画像                    |
| 侵入者検知処理結果 | 417 画像<br>92%<br>(417/454) | 0 画像   | 37 画像<br>8%<br>(37/454) |
| 最終結果      | 99.7%<br>(14363/14400)     | 0%     | 0.3%<br>(37/14400)      |

表 3-8 樹木が背景時の監視結果 (14,400 画像)

|           |                            |      |                         |
|-----------|----------------------------|------|-------------------------|
|           | 正解反応                       | 誤反応  | 未反応                     |
| ライン監視処理結果 | 953 画像                     | 8 画像 | 0 画像                    |
| 侵入者検知処理結果 | 901 画像<br>94%<br>(901/953) | 0 画像 | 52 画像<br>6%<br>(52/953) |
| 最終結果      | 99.7%<br>(14348/14400)     | 0%   | 0.3%<br>(52/14400)      |

15 分の画像前半 7 分から切り出した 25 人分の画像 1,486 枚 (図 3-8 参照) と、マサチューセッツ工科大学の歩行者画像 DB(924 人分)の合計 2,410 枚を人物の学習画像とした。背景画像として人物が写っていない無人画像 1 枚から切り出した 9,281 画像を学習に利用した。その後、残り 8 分の画像に現れる不特定多数の侵入者検知実験を行った。なお、実験に使用した各種パラメータは表 3-2 の値とした。

## b. 実験結果

SVM の学習結果を表 3-6 に示し、アスファルトと樹木を背景にした実験結果を下記に示す。

### i) アスファルトの背景

アスファルト部分での監視結果例を図 3-9 に示し、ビデオ画像後半 8 分 (14,400 画像) の正解率を表 3-7 に示す。表 3-7 に示すように、ライン監視処理では外乱の影響で 155 画像の誤反応が観測された。いずれも、太陽が雲に隠れるときなどに発生した誤反応であった。提案方式は太陽光の変化によるライン監視処理の誤反応を侵入者検知処理で完全に排除できる結果となった。提案方式の時系列結果を図 3-11 に示す。

表 3-9 アスファルトを背景とした実験結果の要約

|             | 正解反応                         | 誤反応  | 未反応 |
|-------------|------------------------------|------|-----|
| ライン監視処理結果概要 | 13 件                         | 32 件 | 0 件 |
| 侵入者検知処理結果概要 | 13 件                         | 0 件  | 0 件 |
| 最終結果概要      | 監視ラインを通過する 13 人を見逃すことなく検知した。 |      |     |

表 3-10 樹木を背景とした実験結果の要約

|             | 正解反応                         | 誤反応 | 未反応 |
|-------------|------------------------------|-----|-----|
| ライン監視処理結果概要 | 17 件                         | 4 件 | 0 件 |
| 侵入者検知処理結果概要 | 17 件                         | 0 件 | 0 件 |
| 最終結果概要      | 監視ラインを通過する 17 人を見逃すことなく検知した。 |     |     |

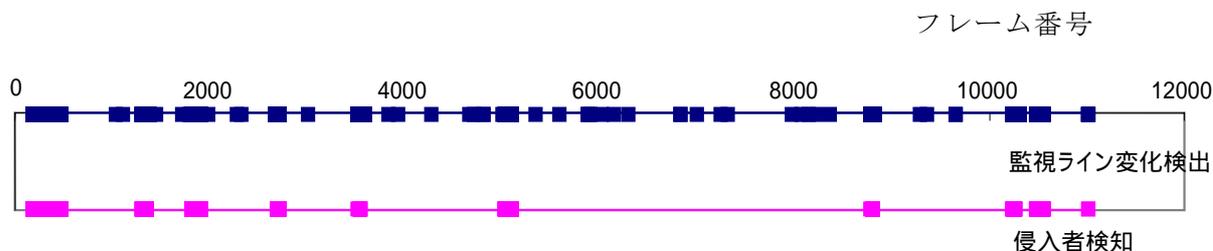


図 3-11 アスファルトを背景とした監視結果

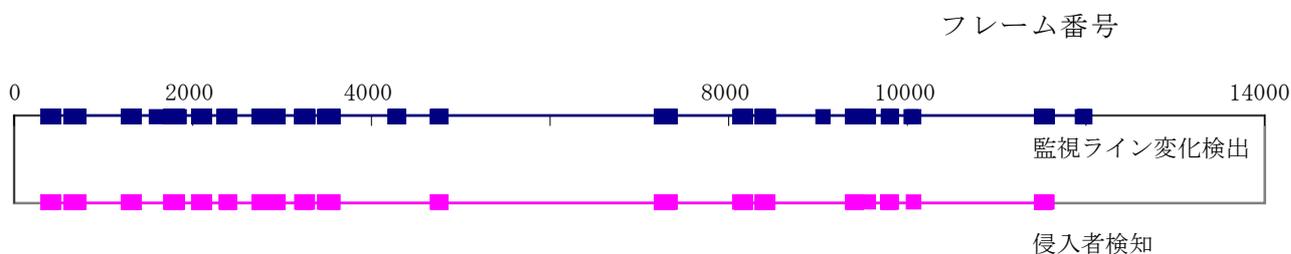


図 3-12 植物を背景とした監視結果

ii) 樹木の背景

樹木部分での監視結果例を図 3-10 に示し、正解率を表 3-8 に示す。アスファルトでは、輝度値のばらつきが小さいため太陽光のわずかな変化に反応し、ライン監視処理で多くの誤反応を起こしている。しかし樹木では、樹木部分の輝度のばらつきが大きいため、太陽光の変化にアスファルトよりも追従する結果が得られた。14,400 画像に対する提案方式の時系列結果を図 3-12 に示す。

表 3-9 と表 3-10 にアスファルトと樹木を背景にした際の「実験の要約結果」を示す。本実験結果により、学習事例と異なる不特定の人物に対して、安定に侵入者検知ができることを確認できた。また、背景が道路や樹木であっても、実時間で安定に侵入者検知ができることを確認できた。

### (3) 充電部接近監視

状況変化に対する頑健性を調べるため、上記(2)で作成した人物画像の学習結果を使い、電力設備周辺で接近監視を行った結果を示す。

#### a. 実験画像

撮影画像は3人の被験者がフェンスを触ったり登ったり、あるいはフェンスの前で立ち止るシーンを含む3分20秒(6,000画像)である。監視ラインは図3-13のように電力設備のフェンスを囲むように3本設定し、各種パラメータは表3-2の値を利用した。

#### b. 実験結果

監視実験結果例を図3-13に示し、6,000画像に対する正解率等を表3-11に示す。「実験の要約結果」を表3-12に示し、提案方式の時系列結果を図3-14に示す。

監視ライン上で人物が静止した場合、人物が背景とみなされるためライン監視処理が反応しない。そのため、監視ライン上で人物が静止している91画像でライン監視処理が未反応となった。しかし、監視ラインへの出入りでは必ず反応があり、移動しないで監視ラインに入ることは不可能なので、充電部接近監視の目的は達成できている。

フェンスの揺れにより人物以外でライン監視処理が誤反応し、その近傍を侵入者検知処理で人物と誤判定した画像(連続3画像)が発生した。監視ライン近傍で、侵入者検知処理が誤判定を起こす場所を調べるため、監視ライン近傍(幅120画素)全体に10画素飛びに侵入者検知処理を実施した。その結果が図3-15である。図3-15から、人物以外に背景の柱を人物として誤検出することが分かる。なお、図3-15の誤検出位置は、表3-12で示す1件の誤反応が発生した場所である。

本実験では、上記(2)の実験で用いた表3-1(B)の無人画像から得た背景画像のみを学習した結果を用いている。そのため、学習に利用した背景画像にない柱を人物として誤検知しているが、監視場所から検出した背景画像(柱やフェンスあるいは電力設備等)を新たに学習することで誤検知を減らすことができる。

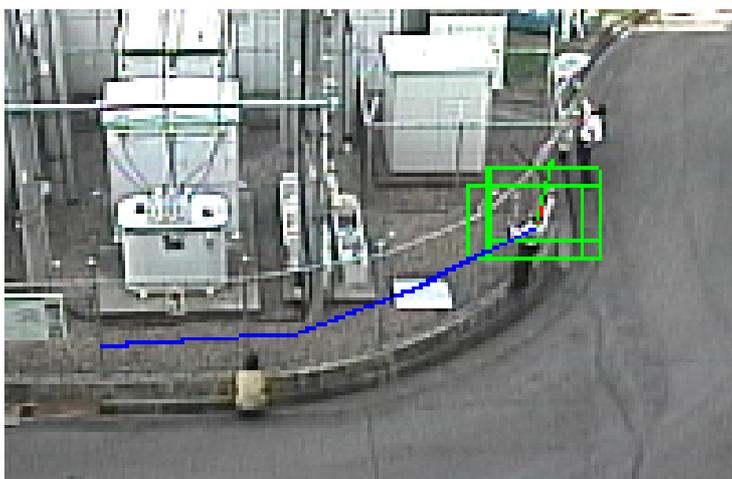


図 3-13 別シーンの学習結果を利用した監視結果例

表 3-11 電力設備への接近監視結果 (6,000 画像)

|           | 正解反応                       | 誤反応                      | 未反応                     |
|-----------|----------------------------|--------------------------|-------------------------|
| ライン監視処理結果 | 498 画像                     | 3 画像                     | 91 画像                   |
| 侵入者検知処理結果 | 478 画像<br>96%<br>(478/498) | 0 画像<br>(上記 3 画像に反<br>応) | 20 画像<br>4%<br>(20/498) |
| 最終結果      | 98.1%<br>(5886/6000)       | 0%<br>(3/6000)           | 1.9%<br>(111/6000)      |

表 3-12 電力設備への接近監視実験結果の要約

|             | 正解反応                                    | 誤反応 | 未反応 |
|-------------|---|-----|-----|
| ライン監視処理結果概要 | 17 件                                    | 1 件 | 0 件 |
| 侵入者検知処理結果概要 | 17 件                                    | 1 件 | 0 件 |
| 最終結果概要      | 1 件の誤反応が発生したが、監視ラインを通過する人物を見逃すことなく検知した。 |     |     |

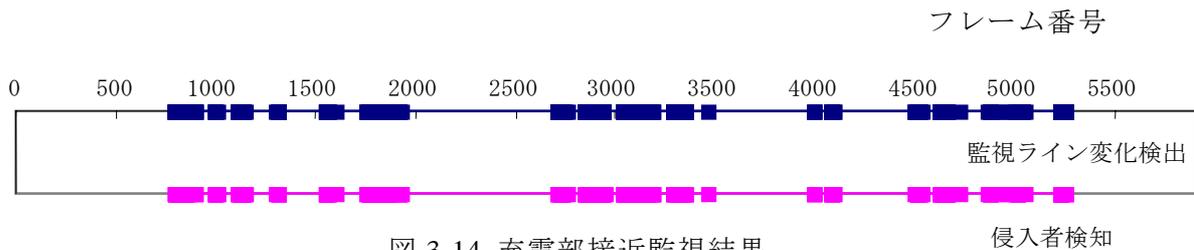


図 3-14 充電部接近監視結果



図 3-15 監視ライン近傍全体に侵入者検知を行った結果

### 3.2.4 考察

本節の実験では、侵入者が監視ライン上を通過する時に、侵入者を実時間で安定に検出する良好な結果を得た。本考察では、提案したライン型前処理を利用した侵入者検知の理論的な限界を示し、その対策を示す。

## (1) 理論的な限界

提案方式は監視ライン上の変化点を検出し、変化点に優先順位を付けて SVM による侵入者検知を行っている。また、一度人物ではないと判定した部分は、一定時間の間、侵入者検知を行わないようにしている。そのため、提案方式は以下に示す理論的な限界がある。

- ・ 侵入者候補領域の未検出

監視ライン上を速く通過した人物は、数フレーム分しか時空間画像上に記録が残らないため見落とす可能性がある。また、監視ライン上の多くの場所で変化する場合、あるいは監視ライン全体で変化が発生する場合、検出する変化領域の上位に人物領域が含まれない可能性がある。

- ・ 侵入者検知の未検知と誤検知

機械学習で侵入者検知を行っているため、事前に学習できていない想定外の迷彩服等を着た侵入者を検知できない限界がある。学習データに対する誤認識率を表す経験損失 (Empirical Error) と、学習パターンが無限にあったときの誤認識率である期待損失 (Expected Error) のバランスを考慮し、未知データに対し誤認識を少なくする発想で機械学習が開発されている。そのため、理論的に侵入者検知の未検知と誤検知は避けられない。

## (2) 対策

将来的に計算機が速くなれば、監視ライン上の変化領域全てに人物検知が行えるようになる。しかし、当分の間は既存の計算機を使う必要があるため、何らかの対策を考案する必要がある。そこで、上記の侵入者候補領域の未検出問題に対し、例えば監視ライン近傍の変化領域を画像間で追跡し、変化領域の軌跡長が特定以上の長さで監視ラインを通過した場所に侵入者検知を行う対策が考えられる。一方、侵入者検知の未検知と誤検知に対して、監視ラインを平行に多重設置し侵入者の見逃しを少なくする対策や、歩行独特の足の動きを認識する処理を加えるなどの対策が考えられる。

機械学習では誤検出と未検出の問題は避けられないため、歩き方の特徴を利用した認識や移動軌跡の解析、あるいは物理的なセンサなどを組み合わせたシステム化の観点から信頼度を上げる検討が必要である。

### 3.2.5 本節のまとめ

本節では画像監視システムに必要な機能の一つである固定監視カメラによる実時間監視を議論した。ここでは屋外監視を対象とし、画像内の変化が人物によるものか人物以外によるものかを、機械学習を使い実時間で判定する問題を取上げた。特に、電気所構内の監視カメラを対象として、計算量のかからないライン型前処理方式を提案し、実験により実時間(約 30fps)で侵入者検知ができることを示した。

従来の機械学習を利用した画像処理では実時間で人物を検出できなかった。しかし、提案方式により従来の機械学習を使いながら実時間で人物を検出できるようになった。また、提案方式は計算量のかからないソフトウェアとして実現しているため、監視現場の既存監視カメラに計算機をつなぐだけで使うことができる。従来のような高価な専用ハードウェアを導入する必要がないため、業務応用にあたり費用対効果の面から大きな利点がある。

### 3.3 旋回カメラによる人物検出の高速化

屋外の広域を監視する場合、監視員は監視カメラを旋回させながら周囲を常時監視している。しかし、前節で提案したライン型前処理や2章で示した既存の人物検出方式は、監視カメラを旋回させながら同時に人物等の移動体を高速に検出し認識することができない。そこで、本節ではカメラの旋回中に移動体候補を検出し、実時間で侵入者を検知する方式を議論する。

既存の画像監視システムで広域を監視する方法として、監視カメラを任意方向に一定時間静止させながら移動体検知を繰り返すこともできる。しかし、監視カメラを一定時間静止させるため、画像範囲外を通過する移動体を見逃してしまう問題が発生する。そこで、監視カメラのパンあるいはチルト中に、背景と異なる移動体を検出し、その移動体が人物ならば監視員に警報を出す方式を提案する。

提案方式では、監視領域に入る人物が周囲と異なる動きをすると仮定し、周囲と異なる動きを持つ領域を移動体候補領域として高速に検出する処理を示す。なお、本節では提案方式と機械学習の侵入者検知を組み合わせる事で、旋回カメラから実時間で人物を検出できることを示す。以降、提案方式を前節の方式と区別するために側抑制型前処理方式と呼ぶことにする。

本節では、3.3.1で監視環境と監視対象を示し、3.3.2で参考とした既存の移動体検出方式と、提案方式について述べる。3.3.4で既存方式と提案方式の実験結果ならびに機械学習による侵入者検知を連携した結果を示す。3.3.5で考察を示し、3.3.6で本節の結論を述べる。

#### 3.3.1 監視環境と監視対象

旋回監視カメラは旋回速度や旋回角度あるいは歩行者の移動速度との関係等、固定監視カメラに比べ不確定要因が増える。そこで、監視状況や環境を明確にするため旋回監視カメラの例として、図 3-16 に示す建屋入口とシャッタ付近に接近する歩行者を等速旋回中の監視カメラで実時間検知する問題をあつかう。図 3-16 に示すような監視環境を想定し、処理の開発にあたり撮影条件と監視対象に以下の3つの仮定を設ける。

##### 1) 等速旋回する監視カメラ

監視カメラの旋回は左右に等速とする。旋回範囲は  $360 \times 240$  画素の画像を横に2枚並べた  $720 \times 240$  画素の範囲とする。8~12秒で監視範囲をカバーできるように、画像内の背景が 1~1.5 画素/画像の速度で動く旋回速度とし、画像の取り込みはビデオレート(30fps)とする。

##### 2) 背景と侵入者の状態

静止背景内を侵入者が歩行しているものとし、歩行速度は監視カメラを固定してビデオレートで画像を取込んだ時に最大3画素/画像程度とする。

##### 3) 画像サイズと侵入者のサイズ

監視画像サイズを  $360 \times 240$  画素とし、歩行姿勢を想定し侵入者の画像サイズを横20~50画素、縦40~60画素程度とする。



図 3-16 対象とする監視場所の例

さらに使い易さと経済性の観点から、以下の3つの開発方針を設けた。

A. 背景の事前知識が不要なこと

監視対象としている電力設備が増強されたり、一時的に物が置かれたりするため、背景の事前知識が不要な方式とする。

B. 既設監視カメラの活用を前提とすること

電気所構内で日常業務の監視に活用している既設監視カメラを利用するため、既設監視カメラに魚眼レンズや全方位撮影装置などを新たに取り付ける、あるいは精度の高い旋回装置を新たに導入するなどの設備増強を前提としない方式とする。

C. ソフトウェアにより高速に動作すること

費用の面から特殊な装置を用いることなく、ソフトウェアで高速かつ高精度に動作する方式とする。

### 3.3.2 既存検出方式の枠組みと既存の前処理

#### (1) 侵入者検出方式の枠組み

前節までの固定監視カメラ用の侵入者検出方式の一般的な枠組みを図 3-17 に示す。図 3-17 に示すように、背景差分などにより画像内の変化部分を侵入者候補領域として検出する前処理と、検出した侵入者候補領域を侵入者と背景に区別するパターン認識処理の組み合わせが広く利用されている。一般に、機械学習を利用したパターン認識処理による侵入者検知[61]を画像全体に行うと、4fps 程度の処理速度しか得られない。そのため、ビデオレートで動作する侵入者検知を実現するには、前処理による侵入者候補領域の検出が重要となる。

しかし、旋回中の監視カメラの画像は画像全体が変化するため、前節で提案したライン型前処理や既存の背景差分などによる前処理が機能しなくなり、実時間での侵入者検知が不可能となっている。そこで以下、本項では旋回中の監視カメラの画像に前処理として利用できる従来方式を示し、その特徴と課題について述べる。

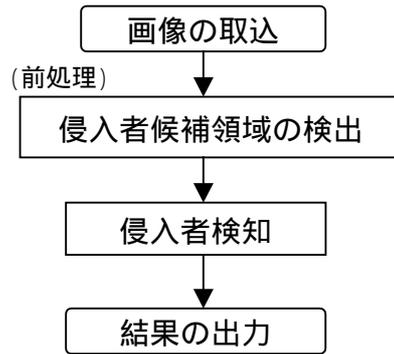


図 3-17 侵入者検出方式の一般的な枠組み

## ( 2 ) 旋回角度を利用した侵入者候補領域の検出方式

監視カメラの旋回度を旋回装置などから直接取得し、侵入者候補領域の検出に利用する方式を示す。

### a. 旋回方向の背景画像の活用

旋回度を利用した従来方式として、監視カメラの全旋回方向の背景画像を作成しておき、最新画像と旋回方向の背景画像の差分で侵入者候補領域を検出する方式[112]が提案されている。しかし、この方式では背景に物理的な変化や日照の変化が発生するたびに全旋回方向の背景画像を更新する必要があるため、屋外の監視用として実用的でなく開発方針 A を満足できない。

### b. 旋回度による画像補正

監視カメラの旋回度分だけ直前の画像を移動させ、最新画像とのフレーム間差分で侵入者候補領域を検出できる。図 3-18 に等速旋回中のカメラで撮影した連続画像のフレーム間差分(図 3-18 a)と、旋回度分だけ直前の画像を移動させて作成したフレーム間差分(図 3-18 b)を示す。図 3-18b から、旋回度を利用することで侵入者領域が強調されることが分かる。

しかし、撮影時の量子化誤差やレンズの歪み、旋回装置の振動や旋回精度などが原因となり、図 3-18b に示すように侵入者以外の背景も検出してしまう問題がある。この問題は、高精細画像や歪みの少ないレンズ、高精度な旋回装置を使うことで解決できるが、新たな設備を導入する必要があるため、開発方針 B を満たさない。

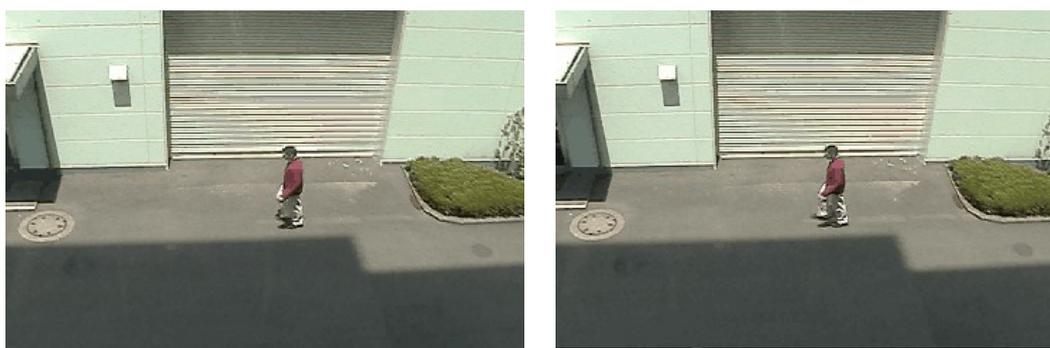
## ( 3 ) 画像処理による侵入者候補領域の検出方式

物理的な装置を使わずに、画像処理により侵入者候補領域を検出する方式を示す。

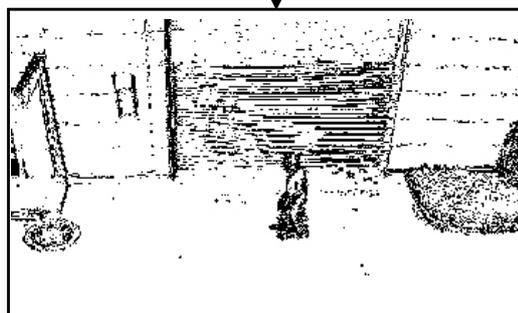
### a. ブロックマッチングの活用

連続画像間の画素の移動量をブロックマッチングなどで計算し、その移動量を利用して侵入者候補領域を検出する方式が提案されている。例えば、直前の画像を移動させて最新画像とのフレーム間差分で変化領域を検出する方式[113]や、画素の移動方向の違いから背景と移動体の動きを分離する方式[48]などがある。

このブロックマッチングは一般に処理時間がかかる問題がある。一例として、 $360 \times 240$  画素の画像を  $10 \times 10$  画素の重ならない隣接するブロックに分割し、各ブロックの移動先を上下左右  $\pm 10$  画素の  $30 \times 30$  画素の範囲で探索すると、Pentium4, 2.26GHz の計算機で



a. フレーム  
間差分画像



b. 旋回度分移動  
した画像による  
フレーム間差分  
画像

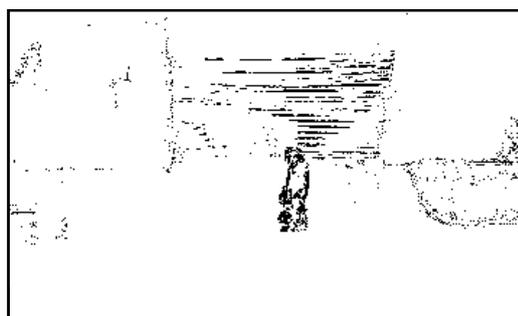


図 3-18 動き補正後のフレーム間差分画像

12fps と低速である。そのため、前処理にブロックマッチング処理専用の LSI を利用する方式[114]なども多いが、特別な装置を導入するのでは開発方針 C を満足できない。

しかし、ブロックマッチングは監視環境に特化した設計を行えば、ソフトウェアでもビデオレートで動く従来方式の 1 つである。

#### b. 動き履歴画像の活用

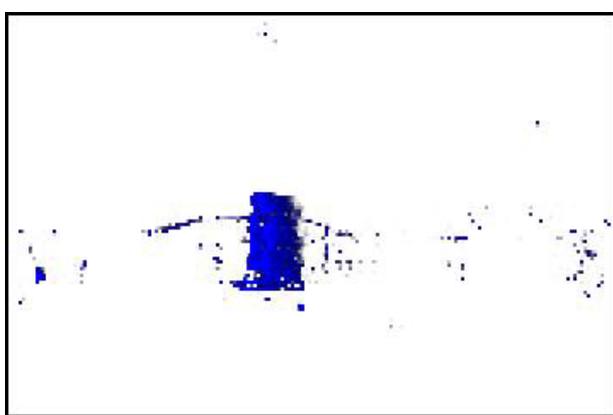
ソフトウェアによる高速で高精度な侵入者候補領域の検出方式として、フレーム間差分などの履歴画像（動き履歴画像、MHI: Motion History Images）を用いる方式[78][79]がある。MHI による移動体検出は、背景の事前記録やカメラの歪み調整などの必要がなく、既設監視カメラの画像からソフトウェアで移動体を検出できるという開発方針を満たす方式である。



a. 固定カメラで撮影した画像[56]



b. 旋回中のカメラで撮影



c. 固定カメラで撮影した画像の動き履歴画像



d. 旋回カメラで撮影した画像の動き履歴画像



e. 固定カメラで撮影した画像での検出結果



f. 旋回カメラで撮影した画像での検出結果

図 3-19 固定/旋回カメラからの MHI による移動体検候補の検出例

MHI による移動体検出例を図 3-19 に示す。図 3-19a,b は静止した背景の前を右から左に歩く人物を固定カメラと旋回中のカメラで撮影した画像である。図 3-19c,d は 0.5 秒間の連続画像で作成した MHI であり、移動方向と反対に影のように伸びている部分が動き履歴である。MHI の影は最新の画像変化部分が濃く、時間がたつに従い薄くなっている。図 3-19e,f

は、MHI 内の最も濃い画素から近傍の薄い画素を統合する処理を繰返して領域分割を行い、縦横サイズの合計が 40 画素以上の領域を円で示している。図 3-19e,f の画像内の円内から伸びている直線は移動方向を示している。

MHI の具体的な処理を下記に示す。MHI 自体を下記の式で更新する。

$$MHI(x, y) = \begin{cases} \tau & \text{if } \Psi(x, y) = 0 \\ 0 & \text{その他、あるいは } MHI(x, y) < \tau - \delta \end{cases}$$

ここで  $\Psi(x, y)$  は利用者が設定する背景差分あるいはフレーム間差分などであり、座標  $(x, y)$  での差分値が閾値  $T$  を超えた時は 1、それ以外は 0 とする。  $\tau$  は画像取得時のタイムスタンプを示す。遅延時間  $\delta$  よりも過去は 0 とする。

次に、MHI を 255 階調のグレー画像に変換する。その際、現在の時刻に近い部分を明るく、遅延部分を暗く表現する。その後、グレー画像に対しソーベルフィルタ

$$S_x = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix} \quad S_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix}$$

を使い、グレー画像の輝度勾配

$$F_x = \sum_{i=-m}^m \sum_{j=-n}^n I(x+i, y+j) S_x(i+m, j+n)$$

$$F_y = \sum_{i=-m}^m \sum_{j=-n}^n I(x+i, y+j) S_y(i+m, j+n)$$

を求め、その縦横の勾配から動き方向を

$$\theta = \arctan \frac{F_y}{F_x}$$

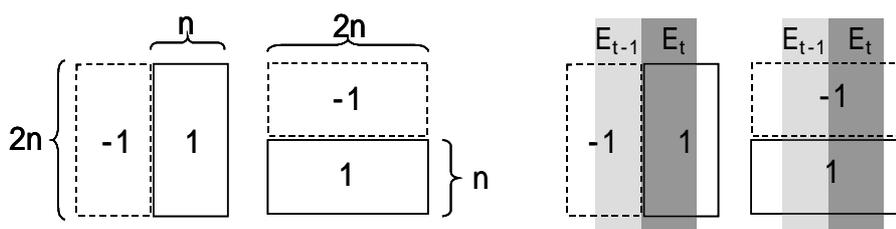
で求める。なお、 $m=n=1$  である。MHI での領域分割は、差分処理で得られるシルエットの輪郭から、floodfill と呼ばれる着目画素値よりも低い近傍画素を統合する処理を繰返す。その後、分割した領域のサイズが  $S$  以上の部分を検出する。

処理速度はビデオレートよりも速く、固定カメラの画像で約 55fps、旋回カメラの画像で約 45fps である。MHI による動き検出は、広く利用されている Horn-Schunck[80]、Lukas-Kanade[81]、ブロックマッチングに比べ 2.8~9.5 倍の処理速度である[82]。ただし、MHI は固定カメラの画像を前提に開発が進められているため、図 3-19f に示すように旋回中の監視カメラの画像では背景変化を多数検出するという問題がある。

### 3.3.3 側抑制型前処理による歩行者の高速な検出処理

対象としている旋回監視カメラの画像は、周囲と異なる動きを示す部分を検出すれば歩行中の侵入者である可能性が高い。そこで、周囲と異なる動きを示す部分を高速に検出する方式を提案し、ブロックマッチングとMHIを使う方式と比較する。

以下、提案方式による侵入者候補領域の検出方式を提案検出方式と略す。また、ブロックマッチングによる検出方式をBM検出方式、MHIによる検出方式をMHI検出方式と略す。



a. 縦横の Haar ウェブレットマスク      b. 2 値画像  $E_t$  でのマスク位置

図 3-20 縦横方向の動き検出マスク

### (1) 提案検出方式

まず連続画像から動きを検出し、次に周囲と異なる動き部分を検出する。その後、周囲と異なる動きが大きい部分を優先的に検出する手順となっている。

#### a. 動き方向の検出

画像内の動きを検出するために、まず連続画像  $I_t, I_{t-1}$  からフレーム間差分を計算し 2 値化した画像  $E_t$  を作成する。

$$E_t(x, y) = \begin{cases} 1 & \text{if } |I_t(x, y) - I_{t-1}(x, y)| \geq T \\ 0 & \text{閾値 } T \text{ 未満} \end{cases}$$

画像  $E_t$  は図 3-18a に示すように背景や侵入者の縁部分が動いた方向に動いた量だけ線幅として現れる。そのため、連続した 2 値化画像  $E_{t-1}$  と  $E_t$  を局所的に比較すると、画像  $E_t$  で新たに線が現れた方向にその線幅だけ移動したことが分かる。そこで、線の現れた方向と線幅を使い、次のように  $x, y$  方向の動きを計測する。まず、フレーム間差分の 2 値化画像  $E_{t-1}$  と  $E_t$  内の線の境界を中心に図 3-20a に示す  $2n \times 2n$  画素の Haar ウェブレットのマスクをかける（例：図 3-20b 参照）。その後、図 3-20a の縦長と横長のマスクを画像  $E_t$  に畳み込んだ結果をそれぞれ  $x$  と  $y$  方向の動きとする。

実験では、背景画像の動きが 1.2 画素/画像で侵入者の動きが 3 画素/画像以下と想定しているため、マスクのサイズに余裕をもたせ  $n=10$  画素とした。なお、提案検出方式はフレーム間差分で得られた縁部分を局所的に比較するため、BM 検出方式や MHI 検出方式に比べ動きの検出精度は落ちるが、高速な処理が可能である。

#### b. 周囲と動きが異なる部分を検出

画像内から周囲と異なる動きをしている部分を検出する。まず、画像全体を隣接する小領域  $\Phi_i$  (小領域サイズ  $l_i \times l_i$  画素:  $i=1, 2, \dots, M \times N/l_i^2$ ) に分割する(図 3-21a 参照)。次に、各小領域  $\Phi_i$  内の動きを持つ画素の動き平均を  $A_i$  とする(図 3-21b 参照)。その後、小領域  $\Phi_i$  を囲む小領域（周囲の 8 近傍）の動きの合計を  $V_i$  とし、

$$\frac{V_i \cdot A_i}{|V_i| |A_i|} \leq \cos \theta$$

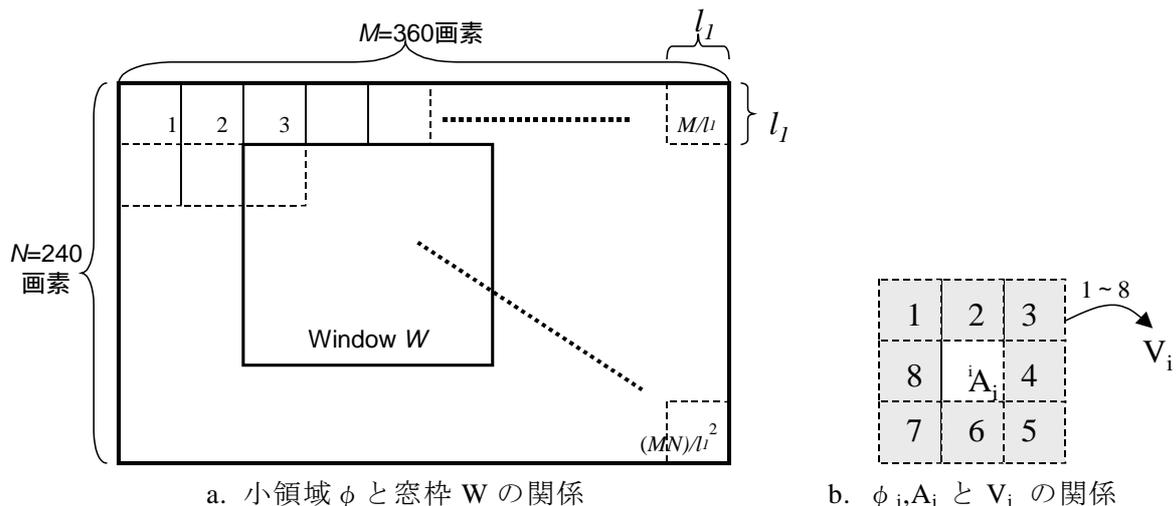


図 3-21 提案検出方式のパラメータの関係  
(小領域  $\phi$  と窓枠  $W$ , 小領域  $\phi_i$  と  $A_i, V_i$  の関係)

を満たす小領域、あるいは周囲の動きよりも大きい ( $|A_i - V_i| > \text{閾値 } T$ ) 小領域を周囲と異なる動きを示す小領域  $\phi_i'$  とする。ここで  $\theta$  は利用者が与える角度閾値であり、 $V$  は動きを持つ 8 近傍の小領域の動き平均である。 $M$  と  $N$  は画像サイズ ( $M=360$  画素,  $N=240$  画素) である。本実験では縁の動きを図 3-20 に示す短辺  $n=10$  画素のマスクで検出するため、 $l_1$  を  $n$  にあわせ  $l_1=10$  画素とした。閾値  $T$  は周囲の動き平均  $V$  の 2 倍 ( $T=2|V|$ ) とした。

### c. 検出優先順位の設定

パターン認識を行う部分を少なくするため、検出領域に優先順位をつける。図 3-21a に示ようなサイズ  $l_2 \times l_2$  画素 ( $l_2 = m \times l_1$ ) の窓枠  $W$  を準備する。窓枠  $W$  を横と縦方向に  $l_1$  画素単位でずらしながら (画像全体で  $(M/l_1 - m + 1) \times (N/l_1 - m + 1)$  箇所)、周囲と異なる動きを持つ小領域  $\phi_i'$  ( $\phi_i' \subset W$ ) の動きを窓枠  $W$  内で合計する。その後、動きの大きい上位の箇所だけを機械学習による侵入者検知を行う。

実験では、窓枠  $W$  のサイズを BM 検出方式のブロックサイズと同じ  $l_2=40$  画素とした。また、順位をつけた上位  $L$  箇所に侵入者検知を適用した。

## (2) BM 検出方式

ブロックマッチングの高速化には、監視カメラの旋回速度や侵入者の歩行速度、侵入者のサイズに合わせた設計が必要である。そこで、まず、ブロックマッチングの処理速度の基礎データを示す。次に、本実験に利用する監視カメラに特化した BM 検出方式を示す。

### a. BM 検出方式設計の基礎データ

ブロックマッチングは  $p \times q$  画素のブロックのテンプレートをそれよりも大きい  $P \times Q$  画素上の探索範囲内 ( $(P-p+1) \times (Q-q+1)$ ) で動かし、類似度が最大になる部分を探す方式である。そのためブロックサイズと探索範囲を変えることで処理速度を制御できる。図 3-22 にブロックサイズと探索範囲を変更した時の処理速度を示す。

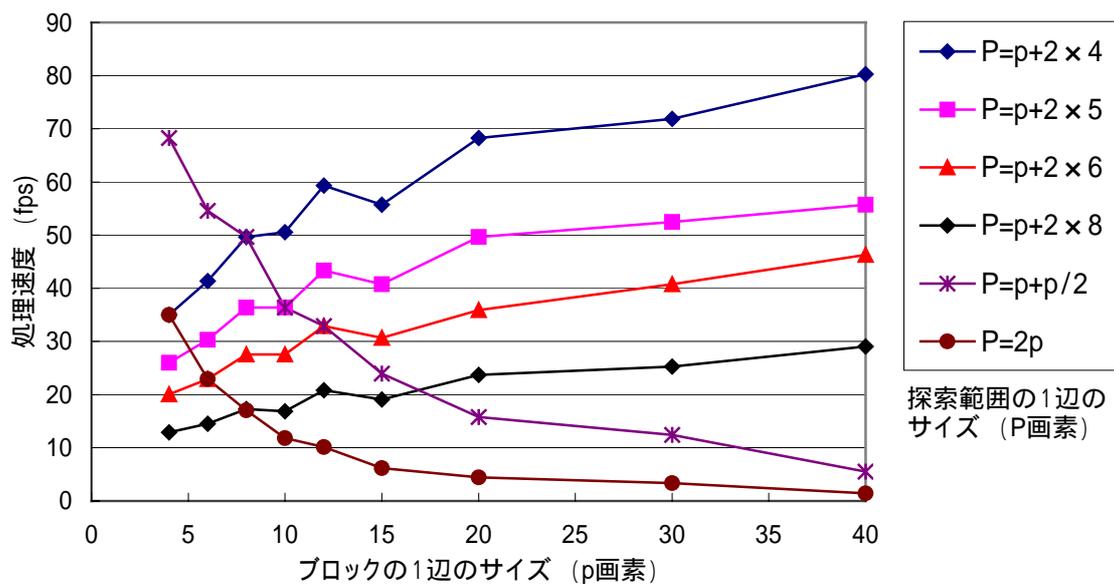


図 3-22 ブロックサイズと探索範囲に依存する処理速度

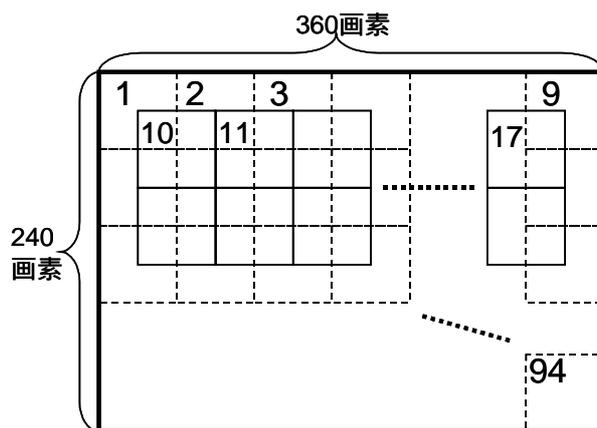


図 3-23 ブロックの配置

図3-22はブロックのサイズを $p \times p$ 画素とし、探索範囲を $P \times P$ 画素として $360 \times 240$ 画素の画像を使い計測した結果である。ブロック同士を隣接して配置し、ブロックの1辺 $p$ を4~40画素、探索範囲の1辺 $P$ を6~ $3p$ 画素まで変化させている。利用した計算機はPentium4, 2.26GHzである。

### b. BM検出方式の設計

図3-22よりブロックサイズが $40 \times 40$ 画素の時、探索範囲が $48 \times 48$ 画素であれば80fpsと高速であり、画像に2回ブロックマッチングを行ってもビデオレートよりも速い処理ができる。

そこで、本実験ではブロックの1辺を侵入者の縦サイズの40画素とし、各ブロックが4分の1ずつ重なる図3-23に示すブロック配置を利用する。図3-23に示す配置は、画像内のブロック数が94個であり、ビデオレートよりも速い処理が可能である。また、着目ブロッ

クを中心に探索範囲が48×48画素であれば、背景と侵入者の移動を検出するのに十分である。

ブロックマッチングの後、背景の動きが全体の動き方向と違う部分を

$$\frac{V \cdot A_i}{|V \parallel A_i|} \leq \cos \theta$$

で検出する。あるいは、背景の動きと異なる $(|A_i - V| > \text{一定値} C)$ 部分を検出する。ここで、 $V$ は画像全体の動きであり、 $A_i, i=1, \dots, 94$ は各ブロックの動きである。 $\theta$ は角度閾値である。

### (3) MHI 検出方式

MHI 検出方式で検出する領域の大きさを、本実験で対象とする侵入者のサイズにあわせ、横幅  $W$  が 20 画素以上 ( $W \geq 20$ ) かつ縦  $H$  が 40 画素以上 ( $H \geq 40$ ) の領域とする。また、履歴記録時間は侵入者候補領域の検出精度に大きな影響を与えるため、0.1～1 秒の間で検出精度が高くなる 0.5 秒とした。例として図 3-24 に 0.1 秒、0.5 秒、1 秒の MHI と検出結果を示す。

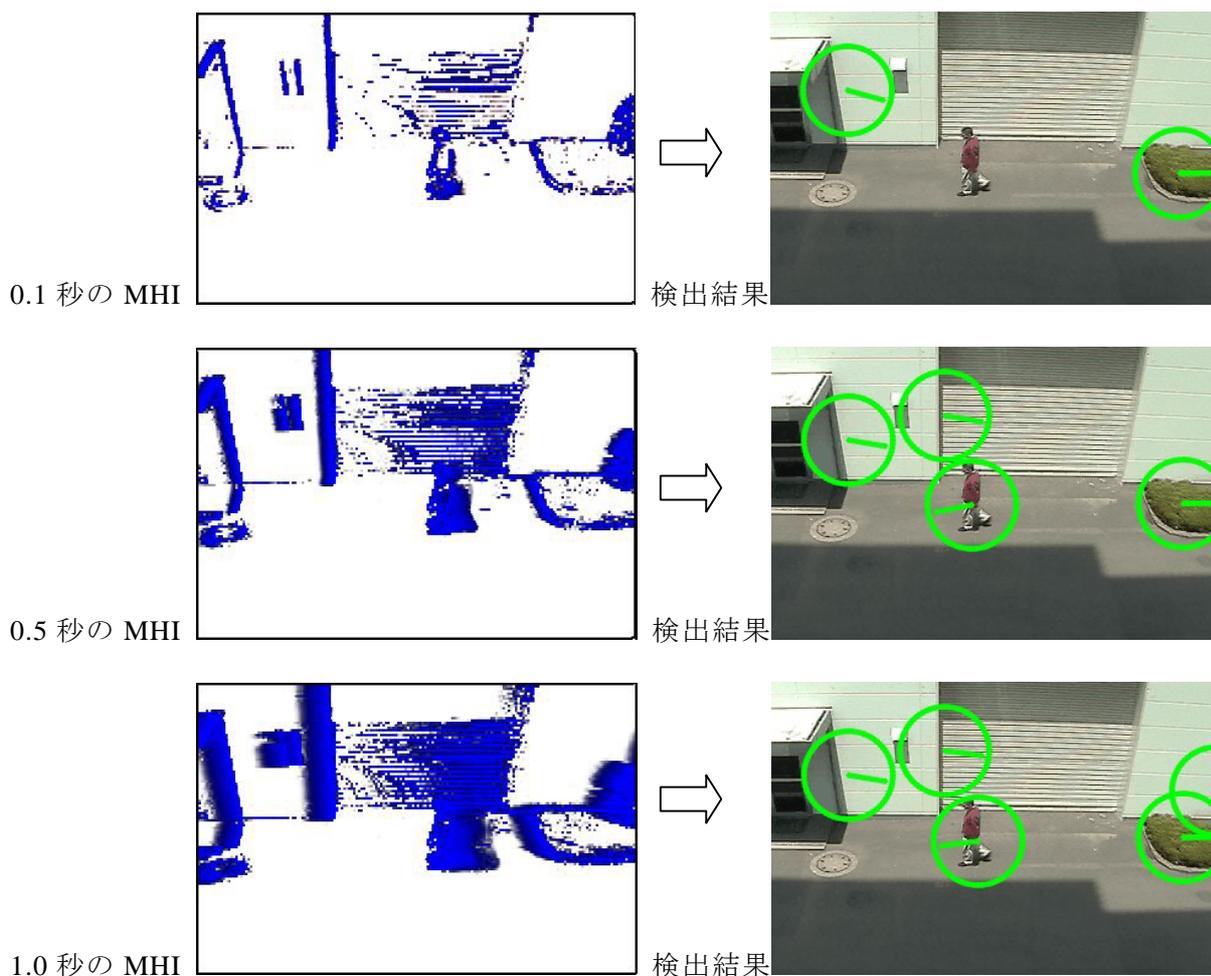


図 3-24 記録時間を変えたときの MHI と検出結果

### 3.3.4 実験結果

機械学習を利用したパターン認識による侵入者検知[58]-[62]では、画像からHaarウェットレットで特徴を抽出した後、サポートベクターマシンあるいはBoostingなどの機械学習結果を参照して人と背景画像を分離している。前節の侵入者検知ではSVMを利用したが、本実験ではSVMの代わりにAdaBoostを利用する。その理由は、機械学習の高速化を図った従来検知方式と比較するためである。本実験では、従来方式としてIntel社が公開している画像処理ライブラリOpenCV (Intel Open Source Computer Vision Library) に含まれるViola-Jones検知方式[61]を利用する。このViola-Jones検知方式は、機械学習にAdaBoostを使い、カスケードと呼ばれる段階的な認識処理で処理の高速化を図っている。なお、実験では機械学習結果もOpenCVに含まれるデータを侵入者検知に利用する。

実験では始めに、監視カメラの向きを固定して撮影した侵入者の画像を使い、3種類の検出方式が機能することを確認する。次に、監視カメラを等速回転させながら侵入者を撮影した画像を使い、従来方式と提案検出方式による検出実験結果を比較する。さらに、提案検出方式による前処理と機械学習による侵入者検知処理を連携した実験結果を示す。

実験に用いる監視カメラの設置状況を下記に示す。

- ・監視カメラと建屋入口までの距離：約 12m
- ・監視カメラの設置位置：高さ 6.6m
- ・回転による背景の移動速度：1.2 画素/画像

実験は1.5GBのメモリを搭載したPentium4, 2.26GHzの計算機を利用した。また、BM検出方式と提案検出方式では、侵入者候補領域の検出に用いる動き角度 $\theta$ を10度とした。MHI検出方式と提案検出方式では、フレーム間差分から2値化画像を作成する時の閾値(0-255の間の値)を50とした。

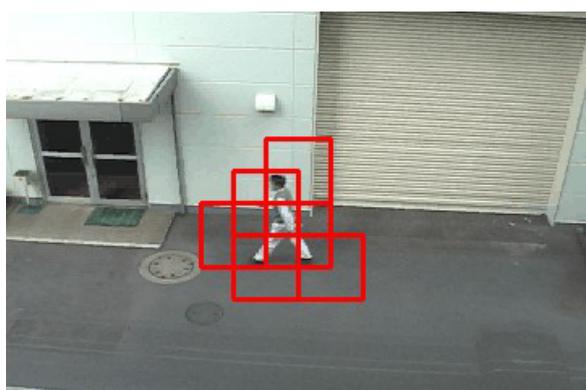
#### (1) 固定監視カメラ

実験では固定監視カメラの前を侵入者が右から左に通過する約 10.1 秒(305 画像)のシーンを利用した。このシーンでは 3.3~9 秒(100~270 番目の画像)までの間に侵入者が画像内を通過している。図 3-25 に侵入者候補領域の検出例を示し、表 3-13 と図 3-26 に実験の集計結果を示す。図 3-25a が BM 検出方式の検出結果例であり、図 3-25b が MHI 検出方式の検出結果例である。図 3-25c が提案検出方式の検出結果例である。

図3-26は各画像で、各検出方式が検出した侵入者候補領域の数を示している。図3-26aに示すBM検出方式では、侵入者がいない0~100画像で不規則に1箇所を誤り検出している様子が分かる。また、図3-26bのMHI検出方式は、動き履歴画像を作成するのに15画像必要なため、はじめの15画像までの間に誤検出を起こしている。また、同様の理由からMHI検出方式による侵入者の検出も提案検出方式に比べ7画像遅れて検出を始めている。さらに、侵入者が画面から消えたあとも、10画像程度余分な誤検出が発生している。図3-26cの提案検出方式では、シーンの立ち上がり1箇所、侵入者が現れる前に1箇所の合計2箇所で誤検出を起こし、侵入者の体が画面の外に出だした時点で3箇所の未検出を起こしているが、それ以外は安定に検出が行われた。

表3-13 固定カメラでの侵入者候補領域検出結果（130画像から260画像の集計結果）

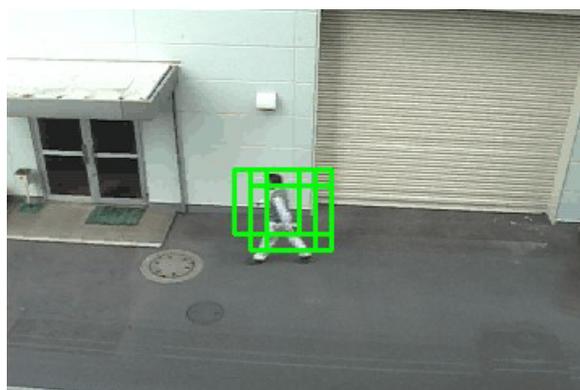
| 前処理     | 検出率  | 未検出率 | 過剰検出率 |
|---------|------|------|-------|
| BM検出方式  | 100% | 0%   | 23%   |
| MHI検出方式 | 100% | 0%   | 0%    |
| 提案検出方式  | 100% | 0%   | 0%    |



a. BM 検出方式の検出例



b. MHI 検出方式の検出例



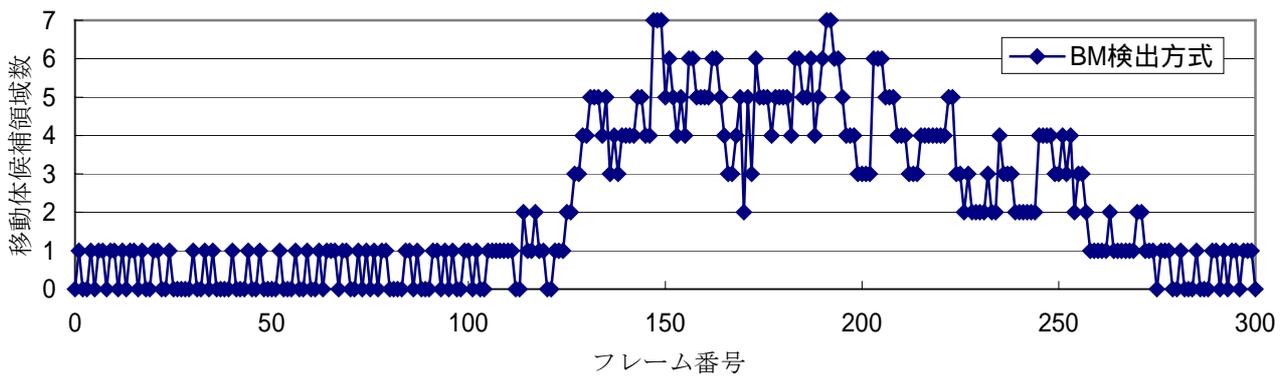
c. 提案検出方式の検出例

図 3-25 固定カメラでの侵入者候補領域の検出結果例

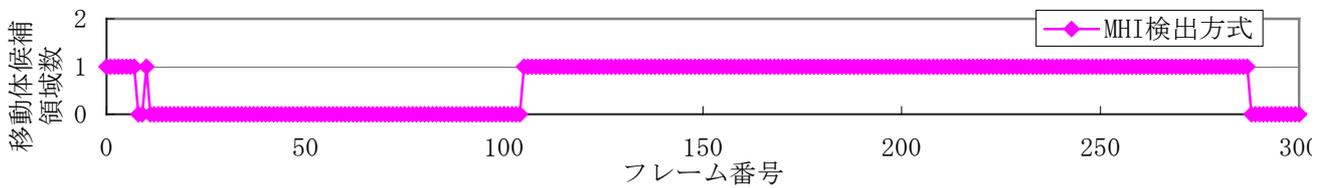
実験シーンの中で侵入者が画面の縁から20画素以上離れている130～260画像までの検出結果を表3-13に集計した。表3-13は、130～260画像の中で、侵入者を検出した画像数の割合を「検出率」欄に記載し、未検出であった画像数の割合を「未検出率」欄に記載している。さらに、侵入者以外の領域を検出した画像数の割合を「過剰検出率」欄に示している。

いずれの検出方式も、侵入者を100%検出できた。ただし、BM検出方式は雨よけの特定部分で不規則に誤検出が発生したため、過剰検出率が23%となった。

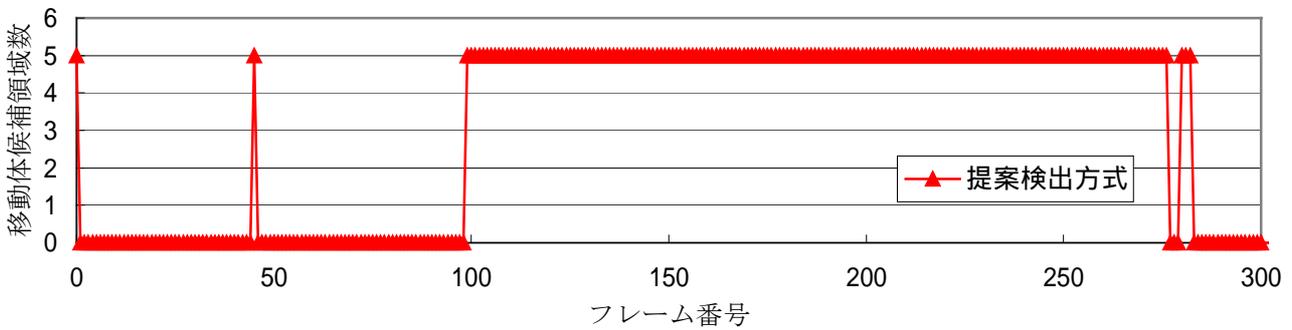
本実験により固定監視カメラで撮影した画像から、3種類の検出方式で侵入者が検出できることを確認できた。



a. BM 検出方式による侵入者候補領域の検出数



b. MHI 検出方式による侵入者候補領域の検出数



c. 提案検出方式による侵入者候補領域の検出数

図 3-26 固定カメラによる侵入者候補領域の検出数の推移

表3-14 各実験シーンの録画時間

| カメラに対する侵入者の移動方向 | 各撮影シーンの長さ    | 合計  |
|-----------------|--------------|-----|
| 逆方向             | 6秒、6秒、5秒、4秒  | 21秒 |
| 同方向             | 8秒、8秒、7秒、7秒  | 30秒 |
| 垂直方向            | 6秒、5秒、5秒、5秒  | 21秒 |
| 斜め方向            | 9秒、7秒、5s秒、4秒 | 25秒 |

## (2) 等速監視カメラ

等速旋回中に撮影した画像を使い侵入者候補領域を検出する実験を行った。実験では監視カメラの旋回方向に対し逆方向に移動する侵入者、同方向に移動する侵入者、垂直に移動する侵入者、斜めに移動する侵入者の4ケース（合計16シーン：表3-14参照）の実験結果を詳細に解析した。その後、1時間の画像で検出実験を行った。

一例として、監視カメラの旋回方向と逆方向に移動する侵入者を撮影した5秒のシーンから検出した侵入者候補領域の例を図3-27に示し、検出数の時系列推移を図3-28に示す。表3-15に侵入者領域を正しく検出した画像数の割合を示し、表3-16に過剰に検出した割合を示す。なお、BM検出方式の1画像内の最大検出数が20箇所であったため、表3-15と表3-16には提案検出方式の検出数を上位5箇所以外に上位20箇所の結果をあわせて示す。

図3-28に1シーン内での侵入者候補領域の検出数の推移を示したが、16シーンではBM検出方式が1画像当たり平均7箇所、最大20箇所を検出し、MHI検出方式が1画像当たり平均3箇所、最大7箇所を検出している。一方、提案検出方式は設定通り1画像あたり5箇所あるいは20箇所の一定数を検出している。

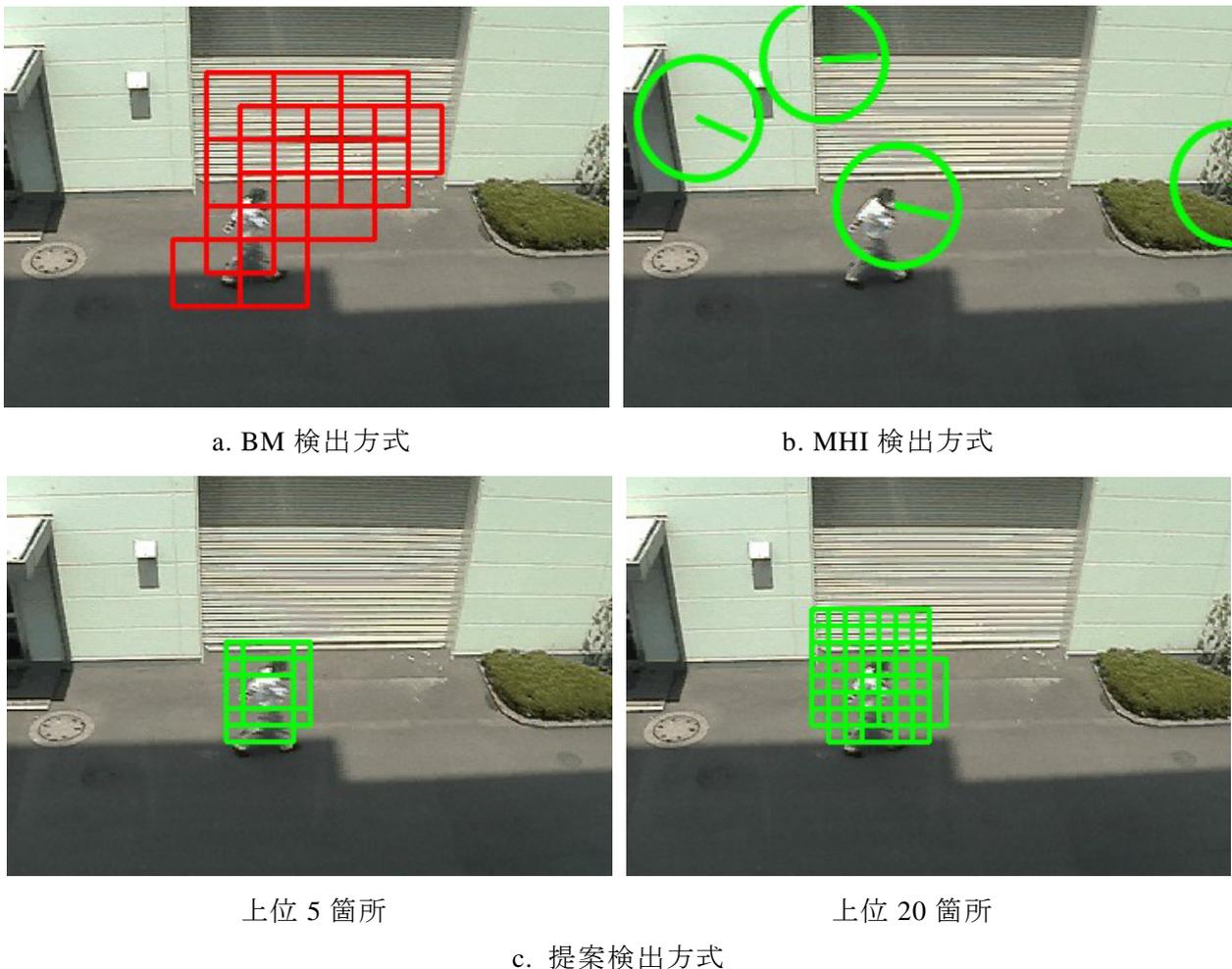
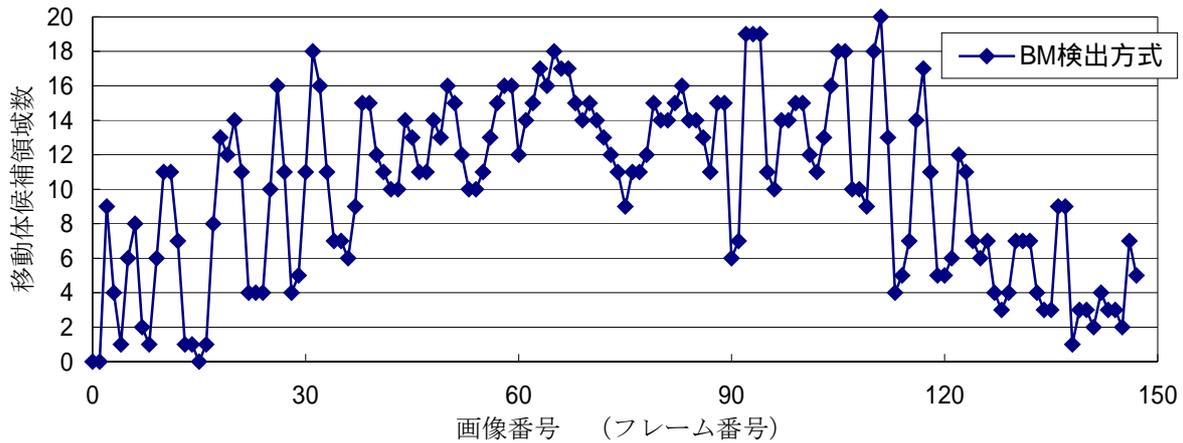
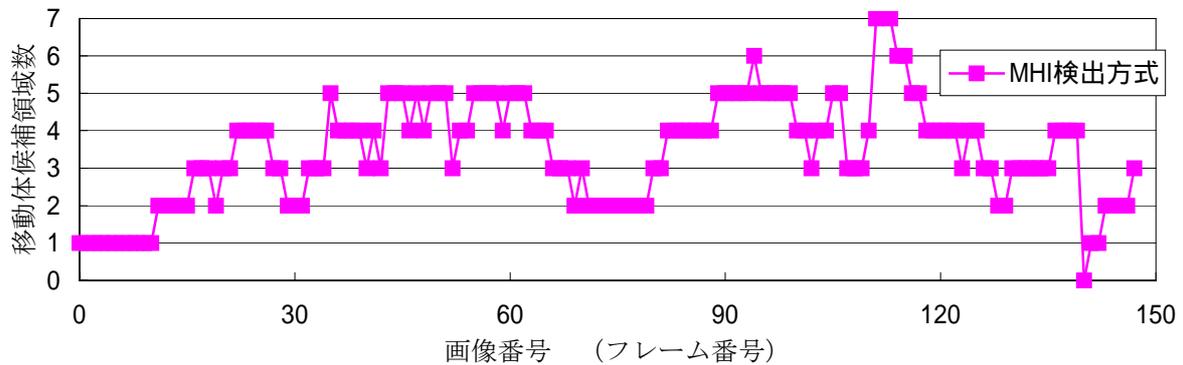


図 3-27 旋回カメラからの侵入者候補領域の検出結果例

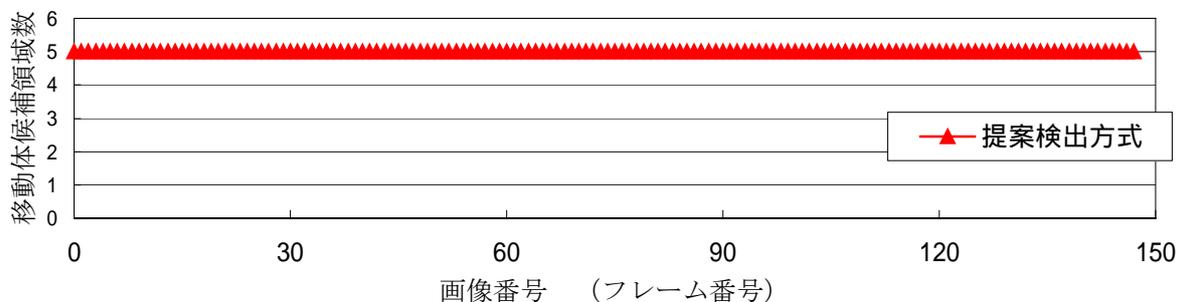
侵入者の検出率の平均は、**BM**検出方式と上位20個所を検出した提案検出方式が95%前後であり、上位5個所を検出した提案検出方式が90%、**MHI**検出方式が約80%であった。表3-16の過剰検出率は**MHI**検出方式と**BM**検出方式がいずれも95%以上であるのに対し、提案検出方式は上位5個所で約25%、上位20個所で約55%と、他方式よりも安定した検出結果となった。過剰検出が多いと次処理のパターン認識を行う部分が増え全体の処理速度が低下するため、提案検出方式のように過剰検出が少ないことが望ましい。



a. BM 検出方式による侵入者候補領域の検出数



b. MHI 検出方式による侵入者候補領域の検出数



c. 提案検出方式による侵入者候補領域の検出数

図 3-28 旋回方向と逆方向に移動するシーンでの侵入者候補領域の検出数推移

巡回中の監視カメラで55人の歩行者の通過を撮影した1時間のシーンに対し、3種類の検出方式を適用した。その結果、歩行者が画像内を通過する際に1度以上歩行者を検出した割合は3種類の方式とも100%であった。いずれの検出方式も、対象とした巡回監視カメラを使い、侵入者候補領域を検出できることが確認できた。

表 3-15 前処理での侵入者候補領域の検出率

| カメラに対する侵入者の移動方向 | BM 検出方式 (%) | MHI 検出方式 (%) | 提案検出方式 (%) |       |
|-----------------|-------------|--------------|------------|-------|
|                 |             |              | 上位 5       | 上位 20 |
| 逆方向             | 100         | 77           | 95         | 98    |
| 同方向             | 93          | 73           | 92         | 93    |
| 垂直方向            | 91          | 90           | 84         | 92    |
| 斜め方向            | 98          | 85           | 89         | 93    |
| Total           | 96          | 81           | 90         | 94    |

表 3-16 前処理での侵入者候補領域の過剰検出率

| カメラに対する侵入者の移動方向 | BM 検出方式 (%) | MHI 検出方式 (%) | 提案検出方式 (%) |       |
|-----------------|-------------|--------------|------------|-------|
|                 |             |              | 上位 5       | 上位 20 |
| 逆方向             | 98          | 100          | 12         | 43    |
| 同方向             | 90          | 100          | 36         | 62    |
| 垂直方向            | 97          | 100          | 33         | 63    |
| 斜め方向            | 94          | 100          | 23         | 47    |
| Total           | 95          | 100          | 26         | 54    |

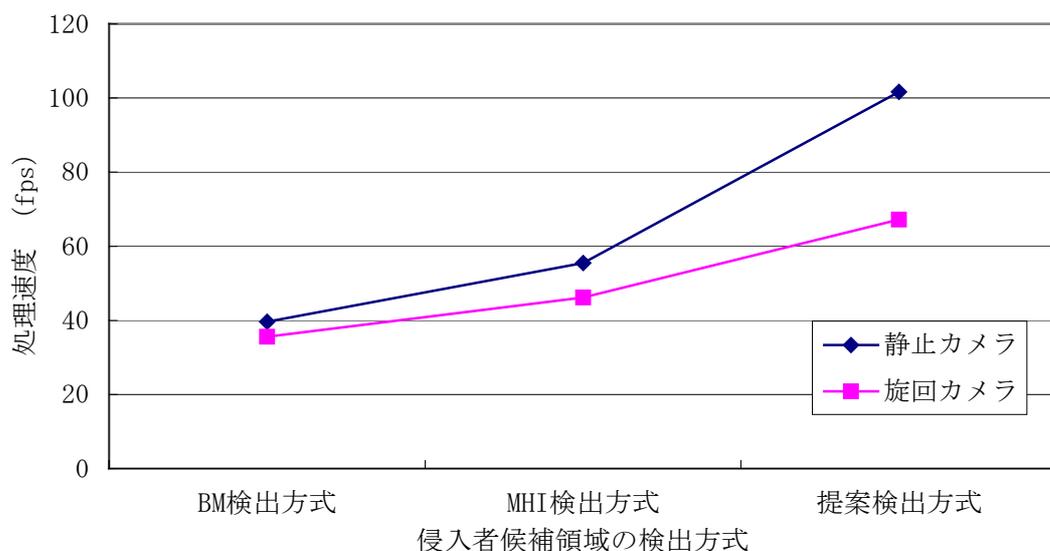


図 3-29 侵入者候補領域の検出処理速度

### (3) 処理速度の比較

監視カメラを固定して撮影したシーンに対する処理速度と、等速旋回中の監視カメラで撮影した16シーンの処理速度の平均を図3-29に示す。いずれの検出方式も監視カメラを固定した画像の方が、旋回中の監視カメラの画像よりも処理速度が速い。また、旋回中の監視カメラでも3種類の検出方式はビデオレート（約30fps）以上の速度で動作することが図3-29より分かる。

BM検出方式は、監視カメラの静止時に39.6fpsと設計通りの40fpsに近い速度が得られている。MHI検出方式は、BM検出方式に比べ処理が速く、監視カメラの静止時が55.5fpsで旋回時が46.2fpsであった。提案検出方式は3種類の検出方式の中で最も処理速度が速く、監視カメラの静止時に101.7fps、旋回時に67.2fpsであった。

### (4) 侵入者検知との連携

3種類の検出処理の中で最も高速な提案検出方式とパターン認識による侵入者検知を組み合わせ、等速旋回中の監視カメラの画像から侵入者をビデオレートで検知できるか調べた実験結果を示す。

パターン認識による侵入者検知を行う領域は、提案検出方式が侵入者候補領域を40×40画素で検出するため、その4倍の面積（80×80画素）とした。また、侵入者検知の重複処理を防ぐため、画像内で一度侵入者を検知した領域の中心から近傍40画素以内は検知処理を行わない設定とした。実験には旋回中の監視カメラの前を斜めに侵入者が移動する9秒のシーンを用いた。表3-17に実験結果を示す。

前処理をせずに画像全体にパターン認識による侵入者検知を行うと処理速度は5.9fpsだが、前処理として提案検出方式を使うとビデオレートよりも速い33.9fpsとなった。また、画像全体に侵入者検知を行うと検知精度は88%であったが、提案検出方式を前処理に使うと検知精度は74%であった。ただし、侵入者検知にはOpenCVに含まれる既存の機械学習結果を用いているので、監視カメラの画像から切り出した侵入者画像と背景画像を新たに学習することで、84%まで検知率を引き上げることができる。（実験に用いた9秒のシーンでは、提案検出方式による侵入者候補領域の検出率が84%であったため。）

この実験結果により、対象とした旋回監視カメラの画像から、提案検出方式を用いることとビデオレートの2倍以上の処理速度で侵入者候補領域を検出できることが分かった。また、パターン認識による侵入者検知と連携しても、侵入者を実時間で検出できることが分かった。

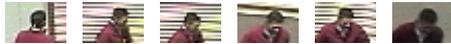
表3-17 侵入者検知処理との連携結果

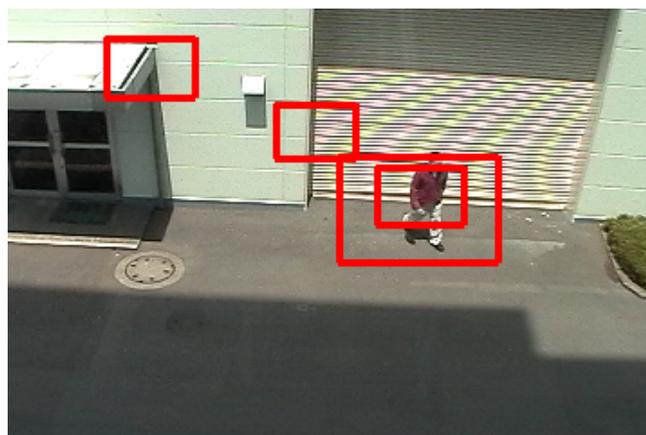
|   | 前処理    | 検知箇所 | 処理速度    | 検知率 |
|---|--------|------|---------|-----|
| 1 | なし     | 画像全体 | 5.9fps  | 88% |
| 2 | 提案検出方式 | 5領域  | 33.9fps | 74% |

参考として、実験対象としたシーンを機械学習で新たに学習した侵入者検知処理と提案検出方式を連携した実験結果を示す。

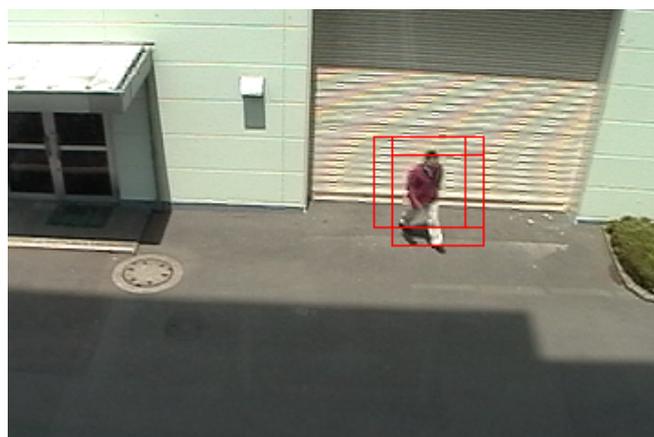
実験に用いた画像から切り出した 120 枚の人物上半身画像 (30×20 画素) と、人物を含まない画像から切り出した 1,800 枚の背景画像 (30×20 画素) を AdaBoost で学習させた。表 3-18 に学習画像例を示す。その学習結果を用いた侵入者検知処理を等速回転カメラで撮影した 9 秒のシーンに単独で適用したところ、人物検出率は 98.6%、過剰検出率は 85.7% となった。図 3-30a に侵入者検知処理を単独で実施した例を示す。一方、提案検出方式を前処理に使うと図 3-30b のように歩行者のみを検知でき、過剰検知が発生しなかった。

表 3-18 機械学習に利用した画像例

|                 | 学習画像例  |
|-----------------|--|
| 人物画像 (120 画像)   |  |
| 背景画像 (1,800 画像) |  |



a. 侵入者検知単独での検出結果



b. 提案検出方式と侵入者検知処理を連携した検出結果

図3-30 学習結果後の人物検出結果例

表 3-19 マスクサイズに対する処理速度と検出率

| $n$ -画素    | 6  | 8  | 9  | 10 | 11 | 12 | 14 | 16 | 18 |
|------------|----|----|----|----|----|----|----|----|----|
| 処理速度 (fps) | 80 | 75 | 70 | 68 | 66 | 63 | 57 | 53 | 50 |
| 検出率 (%)    | 66 | 74 | 83 | 84 | 85 | 78 | 69 | 71 | 63 |

### 3.3.5 考察

#### (1) 前処理に適した検出方式

今回設計したBM検出方式は、旋回中の監視カメラに対し35.6fpsとビデオレート(30fps)よりも速い処理速度が実現できた。しかし、BM検出方式の侵入者候補領域の検出数は1画像当たり最大20個所と多く、20個所全ての侵入者検知を5.6fps (=35.6fps-30fps)で処理できない問題がある。

MHI検出方式は、46.2fpsの高速でかつ高い精度で侵入者候補領域を検出できるため、前処理として有効な一方式である。また、侵入者候補領域の検出数は、1画像当たり最大7個所と、BM検出方式に比べ少ない利点がある。しかし、提案検出方式はMHI検出処理よりも速い67.2fpsで動き、MHI検出方式より高い検出精度が得られる利点がある。さらに、提案検出方式が検出する領域は、真の侵入者領域に集中する傾向がある。

以上より、本実験対象の監視カメラに対し、3種類の方式の中で提案検出方式が最も適切な前処理方式であるといえる。

#### (2) パラメータ値と処理速度

提案検出方式では、画像内の動きを図3-20のマスク( $2n \times 2n$ 画素)を使い計算している。このマスクサイズ $n$ の大きさが処理速度と検出精度に大きな影響を与える。そこで、 $n$ のサイズを変えた時の処理速度と検出精度を表3-19に示す。表3-19より、 $n$ のサイズを小さくすると処理速度が向上するが、検出率が低下することが分かる。また、 $n$ を大きくすると処理速度と共に検出精度が下がることが分かる。そのため、提案検出方式を利用するには、対象とする監視カメラの環境に合わせたパラメータ調整が必要である。

#### (3) 仮想監視カメラによる考察

電力設備を監視する監視カメラの画像からパノラマ画像を作成し、そのパノラマ画像上に歩行中の人物画像を合成することで、旋回中の監視カメラの画像を仮想的に作成した。作成した仮想画像には、3.3.1で設定した撮影条件を超える例も含まれている。

歩行者の速度は2画素/画像とし、背景の移動速度を1画素/画像と2画素/画像の2通りとした。また、背景画像の移動方向と歩行者の移動方向を、同一方向と逆方向の2通りの画像を作成した。特に、背景の移動速度が2画素/画像で背景と歩行者が同一方向に動く場合、歩行者は手足を動かし歩行しているものの、常に画像の中心から移動しない映像となる。なお、侵入者サイズは3.3.1で示した想定よりも大きい縦80画素であり、背景移動の2画素/画像も想定よりも速い移動速度である。



パノラマ背景画像と歩行者画像を合成して作成した仮想カメラ画像

図 3-31 仮想の巡回監視カメラ画像による検出結果例

表 3-20 仮想画像による実験結果

| 背景と歩行者の動き方向 | 背景の移動速度 (pixel/frame) | 検出率 (%) | 過剰検出率 (%) |
|-------------|-----------------------|---------|-----------|
| 逆方向         | 1 p/f                 | 92      | 16        |
|             | 2 p/f                 | 89      | 37        |
| 同一方向        | 1 p/f                 | 83      | 35        |
|             | 2 p/f                 | 74      | 76        |

図 3-31 に合成画像に対する提案検出方式の実験結果例を示し、表 3-20 に検出率と過剰検出率の集計結果を示す。この実験では、窓枠  $W$  のサイズを侵入者サイズにあわせ 80 画素とした以外、他のパラメータは前節の実験時と同じ値を用いている。

表 3-20 から、背景と歩行者の移動方向が異なる場合、背景の移動速度に影響を受けずに、前節の実験結果とほぼ同じ 89% 以上の検出率が得られた。一方、背景と歩行者の移動が同一方向になると、背景の移動速度が遅い場合 83%、両者の速度が同じ場合 74% の検出率となった。特に、背景と歩行者が同一速度で同一方向に移動する場合、逆方向と比べ過剰検出率が約 2 倍の 76% となった。この理由は、背景と歩行者が同一速度になると、歩行者の体の動きが背景の動きに埋もれ、手足の動きだけで歩行者を検出する状態となるためである。そのため、手足が体と重なる時などに他の背景部分を過剰検出する傾向がある(図 3-31 右画像参照)。

今後、背景と歩行者の移動速度と移動方向が同期した場合の検出率向上方法を検討する必要がある。

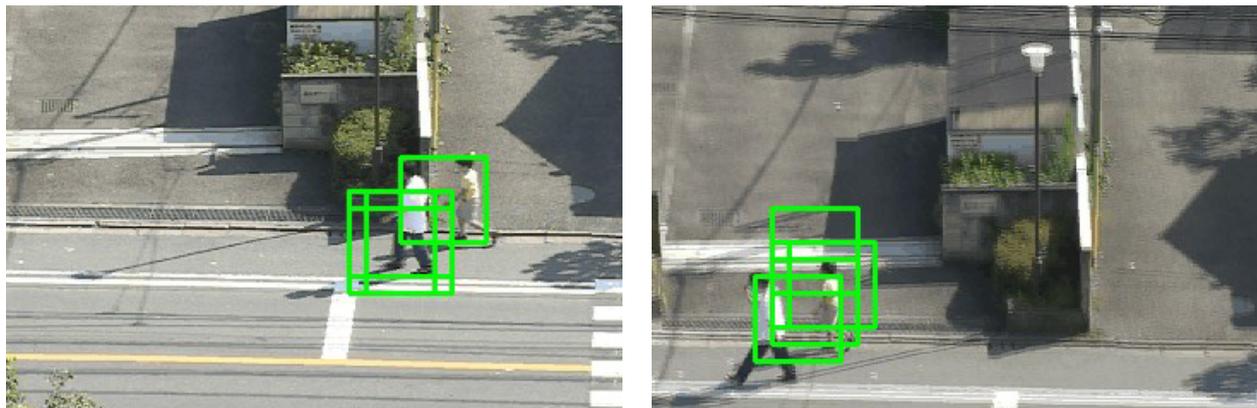


図 3-32 上下にランダムに旋回したカメラで撮影した画像からの移動体候補の検出結果例



図 3-33 携帯ビデオカメラで撮影した画像からの移動体候補の検出結果例

#### (4) 等速旋回以外の画像への適用結果

##### a. ランダムに旋回するカメラの画像への適用結果

監視員が監視カメラを手動で旋回させた状態を想定し、手動でカメラを上下にランダムに旋回させながら撮影した画像から歩行者を検出する実験を行った。実験画像は、歩行者が画像内を右から左へ通過する約7秒間に、カメラを下から上さらに途中から再度下へ向かうチルト中に撮影したシーンである。実験結果例を図3-32に示す。提案検出方式による歩行者の検出率は約92%で、過剰検出率は22%であった。

##### b. 携帯ビデオカメラの画像への適用結果

これまでの実験シーンは、画像内を通過する人物を追跡していないため、いずれも10秒以下の短いシーンとなっている。ここでは、画像内の人物をカメラで追跡することを想定し、公園で遊ぶ子供を監視対象に見立て、携帯ビデオカメラで子供を追跡撮影したシーンを実験に利用した。実験シーンは36.2秒で、子供は常に動いている状態である。また、子供が画像の中心付近に来るように撮影している。ただし、子供が画像内の中心にいる間は、ビデオカメラの動きは手ブレ程度である。

携帯ビデオカメラで撮影した36.2秒のシーンに提案検出方式に適用したところ、人物検

出率は98%、過剰検出率は約5%となり安定した検出結果が得られた。実験結果を図3-33に示す。実験シーンの中で、ビデオカメラの移動量が小さい部分は、固定監視カメラと同程度の安定した検出結果が得られた。

本実験により、ランダムに動く監視カメラや携帯ビデオカメラで撮影した実験シーンでも、提案検出方式により安定した侵入者候補領域の検出ができる可能性が得られた。

## (5) 提案検出方式の理論的境界と対策

### a. 理論的な限界

提案検出方式と MHI 検出方式は、いずれも高速に侵入者候補領域を検出できるが、画像内に侵入者が存在しない状態でも常に背景の一部を侵入者候補領域として検出する問題がある。特に、提案検出方式は局所的な動き変化を検出しているため、画像内の正確な動きが検出できない。そのため、画像全体が同一方向に動く場合、物体のコーナーなどを誤検出する傾向がある。

また、提案検出方式による侵入者候補領域の検出後に、機械学習で侵入者検知を実施する必要があるが、3.3.4 で述べたように学習画像に合わない場合、侵入者の誤検知や未検知が避けられない問題を抱えている。

### b. 対策

複雑な背景でなければ BM 検出方式は侵入者候補領域を正しく検出できる利点がある。そこで、提案検出方式の間に、間隔をあけて BM 検出方式を挿入し、検出結果を補完する方式などが考えられる。例えば、数秒に1回程度 BM 検出方式を挿入し、提案検出方式の結果を比較することで、精度を上げる対策が考えられる。将来的に計算機が速くなることで、BM 検出方式と提案検出方式を同時に実時間処理できるようになると予想できるが、当面は上述のような組み合わせにより、上記の課題に対応する必要がある。

機械学習では誤検知や未検知が避けられない。そこで、提案検出方式で検出した侵入者候補領域を数秒に渡り追跡し、監視カメラの旋回と異なる一定以上の軌跡を観察した段階で、たとえ機械学習の結果が「背景」であっても監視員に警報を出す方法が考えられる。ただし、誤報が増える可能性があるため、画像監視システムの他のセンサなどとの連携を図り、システム化の観点から誤報を減らす対策を検討する必要がある。

## 3.3.6 本節のまとめ

等速旋回中の監視カメラで撮影した画像から、侵入者候補領域を高速に検出する方式を提案した。提案検出方式を従来の検出方式と比較し、従来の検出方式よりも速くかつ安定に侵入者候補領域が検出できることを実験で示した。また、提案検出方式とパターン認識による侵入者検知を組み合わせることで、旋回中の監視カメラの画像から侵入者をビデオレートで検出できることを示した。さらに、パノラマ画像と歩行者画像の合成で、提案検出方式の前提条件を超えた画像を擬似的に作成し、提案検出方式の汎用性とその限界を実験で示した。

これまで、機械学習を利用した人物検出は、固定監視カメラでも実時間処理が不可能であった。そのような中、本提案方式は従来の機械学習方式を使いながら、旋回中の監視カ

メラの画像に対し、実時間で人物検出を可能にしている。また、市販の計算機で機能するため、既存の画像監視システムに導入が容易な成果である。

### 3.4 本章のまとめ

従来の SVM などの機械学習を利用した人物検出方式は、画像内から人物を選択的に検出できるが、実時間(30fps)で処理できない問題があった。そこで本章では、機械学習を使い実時間で人物を検出するため、固定カメラ用にライン型前処理を考案し、旋回カメラ用に側抑制型前処理を考案した。

ライン型前処理は、監視領域内に設定した監視ライン上で動きの速い部分に優先順位をつけ、優先順位の上位を人物検出の探査範囲とする方式である。一方、側抑制型前処理は、移動中の人物が周囲の背景と異なる動きをすることに着目し、周囲と異なる動きを示す部分を人物検出の探査範囲とする方式である。監視カメラ画像を用いた評価実験により、従来の機械学習による人物検出方式を使いながら実時間検出を可能にする研究成果であることを示した。

本章で示した人物検出方式は、市販の計算機で実時間動作するソフトウェアであり、既設画像監視システムの高度化を容易に図ることができる。

## 第4章

### 作業中の人物の姿勢認識

#### 4.1 はじめに

3章の提案方式により、室内や電気所構内のいずれでも実時間で人物検出が可能なため、本章では検出した人物の全身画像を使い、人物の動作に応じて警報を出すための姿勢認識方式を提案する。以下、発電所の運転を監視する監視カメラの画像を使い、運転動作を把握するための姿勢認識を議論する。

##### (1) 姿勢認識方式と課題

人物の姿勢を画像で認識する方式として以下の代表的な方式がある。

###### ① マーカの活用

人物の頭部や間接部分などに赤外に反応するマーカなどを取り付け、赤外光を当てた時のマーカの反射で各部位の位置を計測し、人体の動きを取り込む方式。

###### ② 身体モデルの当てはめ

2次元あるいは3次元の身体モデルを画像内の人物にあてはめ、体の動きに身体モデルを追従させる方式。

###### ③ 姿勢形状とのマッチング

代表的な姿勢を登録しておき、人物画像と登録姿勢のマッチングを行う方式。

画像監視システムでは、事前に社員や侵入者にマーカをつける事ができない。また、身体モデルの当てはめ方式は、初めに画像内の人物と身体モデルをフィッティングする作業が必要となる。このフィッティングが自動化できれば有効な方式だが、現状はこのフィッティング作業を監視員に依頼する必要がある。一方、姿勢形状のマッチングは、作業内容や状況に応じて人体が様々な形状をするため、一般的に典型的な姿勢を定義することが難しい。しかし、次に示すような訓練姿勢が定義できる監視課題では有効な方式となる。

##### (2) 具体的な監視課題

人物の動作内容の把握を主眼に設置している画像監視システムに、原子力発電所の運転訓練室の監視システムがある。発電所の運転訓練では実機と同じフルサイズ・シミュレータにより各種状況を想定した運転訓練を行っている。運転訓練の様子や操作内容を常時監視カメラで撮影し、運転操作の把握に利用している。

運転時の操作を把握するために訓練員に身体モデルを当てはめ、操作内容を記録する方式が提案されている[68]。同様な方式として、腕,足,腰,頭等の間接を持つ身体モデルを画像内の人物に当てはめ、動作を正確にCG化する画像処理研究が数多く報告されている[55][66][67][84][129]。それらの研究では身体モデルと人物の初期合わせや追跡途中のずれ

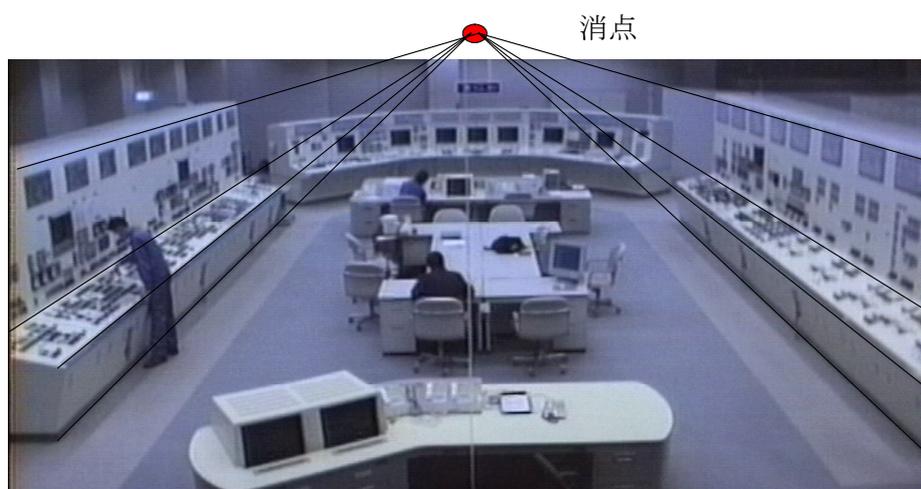


図 4-1 対象としている画像例と画像の消点

補正が課題となる。また、これらの方式は身体モデルの当てはめができるよう人物を拡大撮影する必要があるが、対象とする既存監視カメラの画像では訓練員の身長が画像内で 80 画素程度まで小さくなるため、身体モデルを当てはめづらい画像となっている。

監視カメラは発電所運転室の全景が写るように設置されており、訓練員の動線（動きの軌跡）と操作概要の把握に主眼が置かれている。本章では、そのような監視カメラで撮影した訓練画像から背景差分による画像処理で訓練員を検出し、機械学習で姿勢を認識した後、その認識結果を身体 CG で確認する方式を提案する。特に、訓練員の姿勢は常に変化するため機械学習による認識が困難な対象の一つである。そこで、発電所運転に特有の 4 種類の特定姿勢を使う認識方式を示す。提案方式は、画像ごとに訓練員の姿勢を 4 種類に分類するため、身体モデル当てはめ方式で必要な身体モデルと訓練員の初期対応付けや追跡途中のずれ補正を必要としない利点がある。

4.2 節で解析対象とする監視環境と監視対象を示す。4.3 節で画像特性と姿勢の制約を取り入れた姿勢認識方式を示す。4.4 節で約 2 万 1 千枚の画像を使った実験結果により、安定した姿勢認識が可能であることを示す。提案方式での姿勢認識には人物の上半身画像を利用したが、比較のために下半身画像を利用した認識結果を 4.5 節の考察で示す。4.6 節で本章の総括を示す。

## 4.2 監視環境と監視対象

### (1) 監視環境

一定照明の下、固定監視カメラで室内全体を撮影する発電所運転訓練画像を本章の対象とする。特に提案方式の検証では図 4-1 に示す発電所運転操作盤（左側、中央、右側）の左側操作盤の訓練員姿勢を実験に利用する。なお、運転訓練時の操作盤担当者は基本的に 1 名で、固定カメラのフォーカスやホワイトバランス等も一定である。操作盤の上段と中

段には主に監視用計器が設置され、下段には操作ボタンやスイッチ類が配置されている。電話は足元に配置され、受話器を取るときや置くときに体をかがめる必要がある。

## (2) 監視対象とする動作と姿勢

訓練員が操作盤の前で行う主な動作は、状態変化等の「監視」、スイッチ等の「操作」、機器の指示値読取等の「指差し確認」、または「電話」、「報告」、「移動」の6通りに大分類できる[107]。

特に、移動状態と静止状態で分類すると、表 4-1 に示すように移動状態は「監視」あるいは「指差し確認」のいずれかとなり、静止状態は「監視」、「操作」、「指差し確認」、「電話」のいずれかとなる。「電話」は話しながら「監視」や「操作」をする場合もあるが、ここでは電話を取る/置く動きだけを「電話」とみなす。なお、図 4-1 の中央に座る人物への「報告」は特定の動作がなく、音声解析を導入しないと判断できないため画像処理による姿勢認識の対象から除外する。

そのため、画像で認識すべき訓練動作は「監視」「操作」「指差し確認」「電話」であり、各動作の姿勢が提案方式の認識対象となる。なお、姿勢認識に加え訓練員の動線（動きの軌跡）が得られれば、訓練概要を4姿勢の身体CGで可視化できる。

表 4-1 訓練員の状態と姿勢

| 状態   | 姿勢                      |
|------|-------------------------|
| 移動状態 | 観察<br>指差し確認             |
| 静止状態 | 監視<br>操作<br>指差し確認<br>電話 |

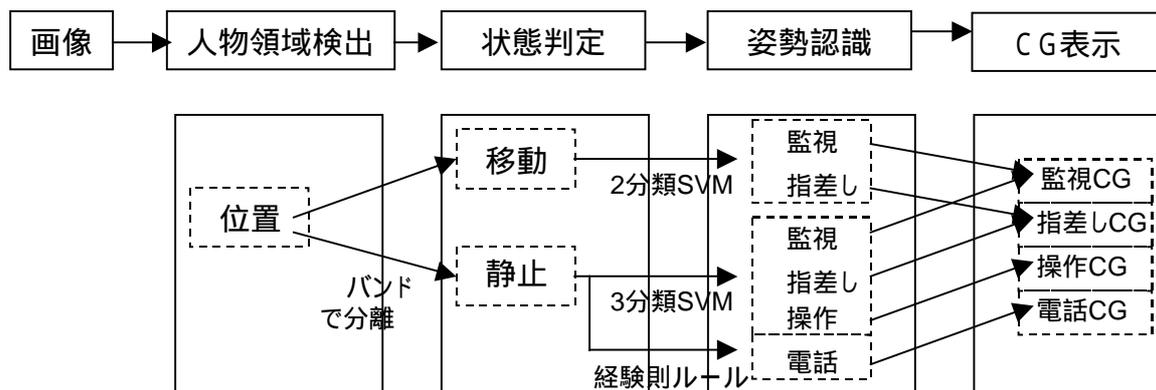


図 4-2 提案方式での姿勢認識処理手順

### 4.3 姿勢認識システムの概要

姿勢認識処理の概要を図4-2に示す。提案方式は、初めに背景差分を使い連続画像から人物領域を抽出し、次に人物領域の位置変化から動線を検出し動きの状態を判定する。さらに人物領域の画像と動きの状態を使い姿勢を認識し、その結果を身体CGに置換して確認表示する手順となっている。下記に、「人物領域検出」処理、「状態判定」処理、「姿勢認識」処理の詳細を示す。

#### 4.3.1 人物領域検出

人物領域検出処理は、はじめに差分処理により画像内の変化領域を検出し、その変化領域の雑音を除去した後、ラベリングを行い変化領域の塊を人物領域として検出する。下記の(1)~(3)に各処理の詳細を示す。

##### (1) 背景差分による変化検出

処理の初めに人物が写りこんでいない連続画像の平均画像を背景画像 $I_b$ として作成する。室内照明が一定でも蛍光灯のちらつきにより物体の縁などで輝度値にばらつきが出るため、その様子を背景画像の標準偏差画像 $I_d$ として作成する。画像内の人物領域を検出するため、連続画像 $I_i$ から背景画像 $I_b$ を下記の計算により差引き、その結果を差分画像 $B_i$ とする。

$$B_i = \begin{cases} |I_i - I_b| & \text{if } |I_i - I_b| > I_d \\ 0 & \text{else} \end{cases}$$

その後、差分画像 $B_i$ の画素値が閾値  $T$  より大きい値を人物領域候補画像 $B_i'$ として検出する。なお、本実験では閾値  $T$  を30の定数とした。



図 4-3 人物領域の検出結果例

## (2) 雑音除去

人物領域候補画像 $B_i'$ から孤立画素などの雑音を除去し、さらに検出できていない小領域を補完する。本実験では検出された画像領域の縁を2画素削り、その後2画素足しこむMorphologicalフィルタリング[75]により、数画素程度の塊を雑音として除去する。

## (3) ラベリングと統合

ラベリング処理により画素領域の塊ごとに番号を付与する。ただし、雑音除去後の画像では、頭部や手足等の領域が体と分離している場合が多いため、近傍領域 $L$ 以内で近接している画素領域を人物領域として統合する。この近傍領域の閾値 $L$ を一律に固定すると余分な雑音を人物領域として統合してしまうため、下記の式(4-1)にて $L$ を決定する。

透視投影効果が成り立つ画像で、レンズの焦点距離を $f$ とすると、カメラ座標系で $(x, y, z)$ にある点の像は投影面上の $(fx/z, fy/z)$ に投影される。任意の点を通り単位ベクトル $(m_1, m_2, m_3)$ に沿う直線上の点をパラメータ $s$ を用いて $(x+sm_1, y+sm_2, z+sm_3)$ と表現すると、この投影は投影面上で

$$\left( f \frac{x + sm_1}{z + sm_3}, f \frac{y + sm_2}{z + sm_3} \right)$$

となる。さらに $s \rightarrow \infty$ とすると $(fm_1/m_3, fm_2/m_3)$ に収束する。この原理から平行線の延長線は消点と呼ばれる1点に収束する(図4-1上部の点)。図4-1の画像中央に近い左側操作盤入り口で、人物の高さが約80画素になるため、重心から半分にあたる40画素を消点からの距離に応じて式(4-1)で閾値 $L$ を変化させ、その範囲にある画素塊を1つに統合する。

$$L = \frac{80}{2} \times \frac{|\mathbf{P}_2| \cos \theta_2}{|\mathbf{P}_1| \cos \theta_1} \quad (4-1)$$

ここで $\mathbf{P}_1, \theta_1$ は消点から左側操作盤入り口までのベクトルと、そのベクトルが横軸となす角度である。同様に、 $\mathbf{P}_2, \theta_2$ は消点から新たに得られた画素塊までのベクトルと、そのベクトルが横軸となす角度である。

以上の(1)~(3)の処理で得られた人物領域例を図4-3に示す。図4-3の上段は一塊の人物領域として検出した例であり、下段は検出領域の原画像と人物のシルエット、ならびにその2値画像を示している。

### 4.3.2 状態判定

はじめに、連続画像内での人物領域の位置変化を動線として検出する。ただし、人物領域の検出結果は背景が黒の場所で頭の髪部分が欠落したり、画像ノイズを人物領域として検出したりと様々な未検出や誤検出が発生する。図4-4は連続した3,000画像から検出した人物領域を正規化したサイズであり、集団から離れた場所にプロットされている部分は検出誤りである。そこで連続した $n_j$ 枚分の画像から人物領域サイズの移動平均を求め、標準偏差の $d_1$ 倍未満の変化を許容し、それ以上を検出誤りとして排除する。なお、本実験で

は 1,118 画像を使った予備実験で良好な結果を得た  $n_1=20$ 、 $d_1=1.5$  を利用した。

次に、予備実験画像から人物領域の検出誤りを除去した後の動線を X 軸と Y 軸ごとに時系列展開した結果を図 4-5 に示す。訓練員の動作は表 4-1 のように移動状態と静止状態に区分できる。そこで、図 4-5 に示した X 軸と Y 軸ごとの時系列の標準偏差を使い移動状態と静止状態を分離する。その分離は上述と同様に、連続した  $n_2$  枚分の画像の移動平均を求め、標準偏差が  $d_2$  以上を移動状態、 $d_2$  未満を静止状態とした。なお、本実験では図 4-5 のグラフを移動状態 598 画像、静止状態 520 画像に分割した  $n_2=20$ 、 $d_2=4$  を定数として使用した。

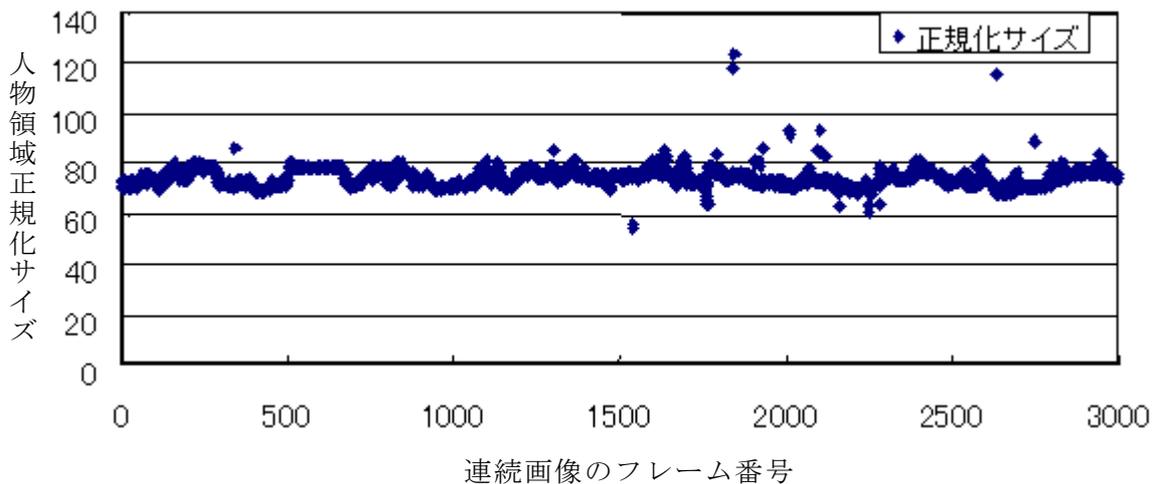


図 4-4 検出した人物領域のサイズを正規化した結果

### 座標

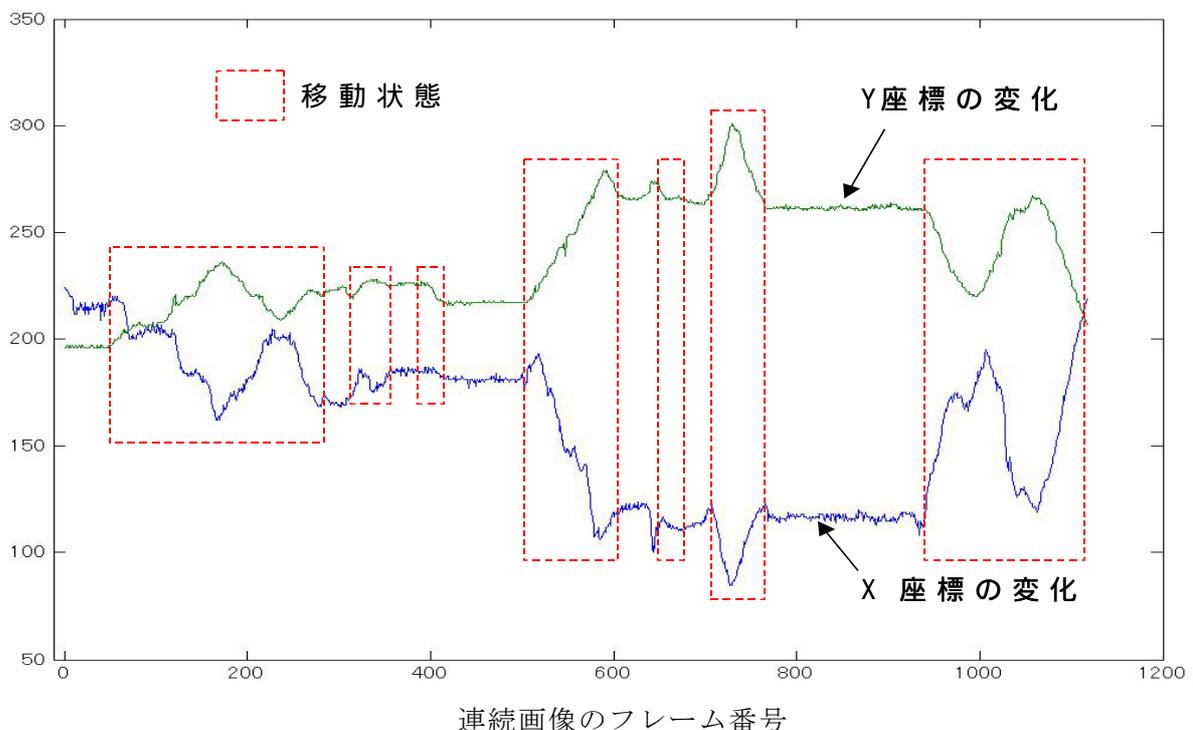


図 4-5 訓練員の移動状態と静止状態の分類

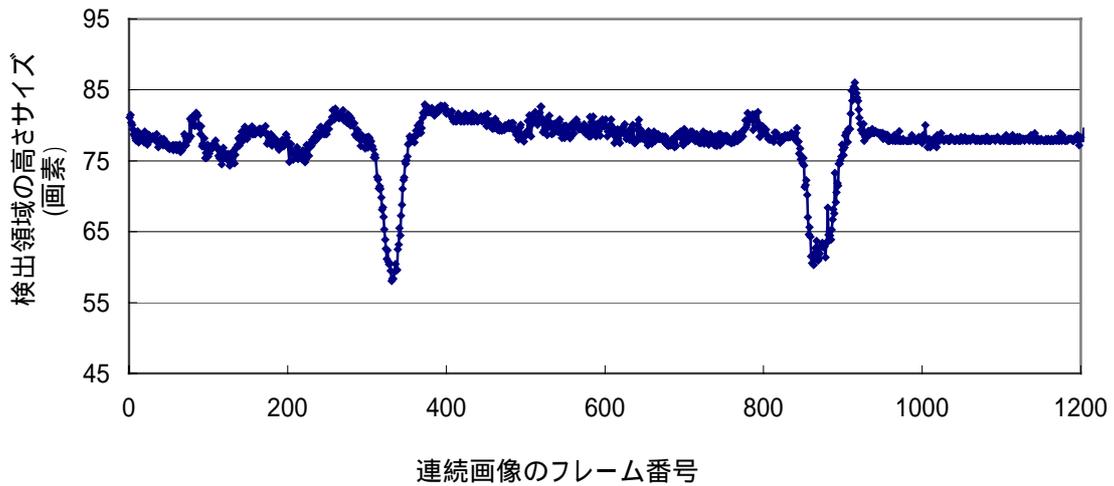


図 4-6 電話と他の姿勢の分離

表 4-2 正規化した姿勢の学習用画像例

| 姿勢                | 事例画像 |
|-------------------|------|
| 監視<br>(562 画像)    |      |
| 操作<br>(265 画像)    |      |
| 指差し確認<br>(108 画像) |      |

### 4.3.3 動的な切替による姿勢認識

訓練員の姿勢を認識するには、表 4-1 に示す「監視」「操作」「指差し確認」「電話」姿勢を認識する必要がある。特に、操作盤の足元に電話が設置されているため、「電話」姿勢

は図 4-6a に示すようにそのサイズが明らかに他の姿勢と異なる。そこで、「電話」と残りの姿勢を分けて認識ルールを作成した。

### (1) 電話姿勢の認識ルール

消点からの距離に応じて人物領域サイズを正規化し、人物の高さ変化を時系列で表示した結果が図 4-6 のグラフである。電話を取るときにかがむため、図 4-6 のグラフでV字部分が電話を取る姿勢である。そこで、正規化した人物領域サイズが 70 画素(閾値 H)以下とその前後 5 画像分を経験則で「電話」姿勢とした。

### (2) 3 姿勢の認識ルール

残りの 3 姿勢認識は経験則でのルール化が困難である。そこで、図 4-6 に示した予備実験用の連続画像から検出した人物領域画像を、目視で「監視」「操作」「指差し確認」の 3 姿勢に分類し、その後、機械学習により姿勢認識ルールを作成した。なお、学習画像は姿勢があいまいなものを除外した結果、「監視」562 画像、「操作」265 画像、「指差し確認」108 画像となった。人物領域を 32×32 画像に正規化した学習画像例を表 4-2 に示す。

教師付き機械学習用のアルゴリズムとして多様な方式が提案されているが、本論文では代表的な学習方法であり前章の人物認識実験で良好な結果を示した SVM を利用する。また、カーネルにはガウシアンカーネルを使用した。さらに、SVM の姿勢認識特徴には 32×32 画素へ正規化した人物領域を背景部分 0 と人物部分 1 に 2 値化し、その上半身に相当する画像の上半分 512 画素(=32×32/2)をそのまま形状特徴として利用した。

### (3) 動作状態による認識結果の切替

SVM は 2 分類問題用の機械学習方式なので、3 分類以上は SVM を組み合わせて使う必要がある。ただし、学習データが不十分な場合、分類クラス数が増えるに従い途中の誤分類の影響を受けて全体の認識率が低下する。そこで、表 4-1 より移動状態が 2 分類問題、静止状態が 3 分類問題となることから、誤認識を減らすため SVM の姿勢認識結果を式(4-2)で動的に切替て利用する。

$$C_2^i = \text{“移動”} \quad \text{then} \quad P^i = P_2^i \quad \text{else} \quad P^i = P_3^i \quad (4-2)$$

ここで、移動状態と静止状態の判定結果が  $C_2^i = \{\text{移動、静止}\}$ 、SVM の 2 姿勢認識結果が  $P_2^i = \{\text{監視、指差し確認}\}$ 、3 姿勢認識結果が  $P_3^i = \{\text{監視、操作、指差し確認}\}$  であり、最終認識結果が  $P^i$  である。ただし、 $i$  は連続画像の画像番号である。なお、静止状態 ( $C_2^i = \text{“静止”}$ ) では経験則ルールによる「電話」姿勢の認識結果を式(4-2)の結果よりも優先した。

最終的に姿勢認識結果を整えるため、同じ姿勢が  $N$  画像以上続かない姿勢を前後の姿勢で置換する経験則を付加した。下記では  $N$  を補正パラメータと呼ぶ。

## 4.4 認識実験

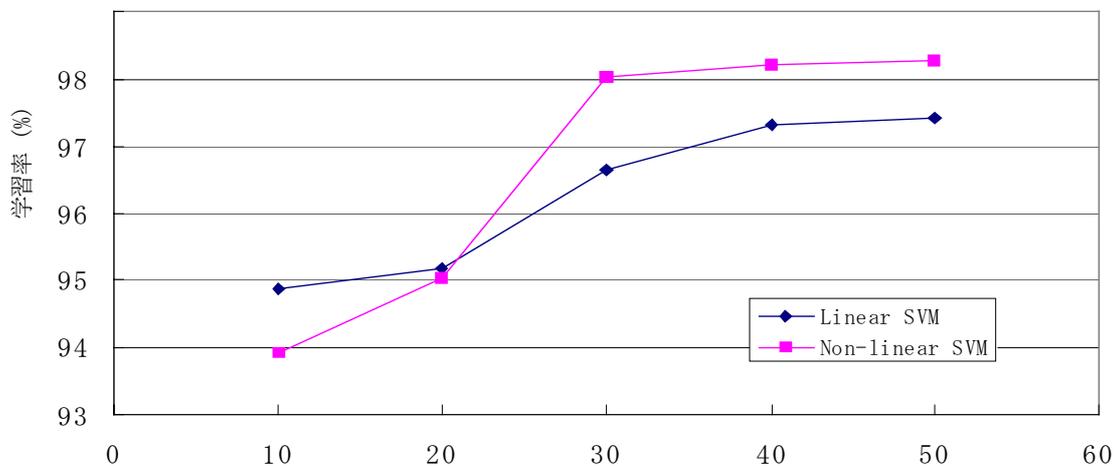
はじめに、機械学習で作成した認識ルールの検証実験を示す。次に、動線検出と姿勢認識の検証実験を示す。さらに、2万枚の画像による実験結果を示す。本実験では人物の上半身画像を SVM の姿勢認識に利用したが、考察では下半身画像を利用した場合の認識率を示す。なお、本実験では補正パラメータ  $N$  だけを変更し、他のパラメータは前述の値を用いている。

### 4.4.1 3 姿勢の認識ルールの検証

SVM は利用するカーネルや画像特徴で結果が大きく変動する。そこで、はじめに提案方式で利用するガウシアンカーネルの妥当性を調べた。表 4-2 に示した姿勢の目視分類画像 935 枚の 10%~50% を学習し、残りを認識した結果が図 4-7 である。ガウシアンカーネルは 50% の画像学習で約 98% の認識率であり、学習数が少ない 10% でも約 94% の認識率を示した。また、代表的な線形カーネルを使った同様の実験結果を図 4-7 に重ねて示す。この実験で、ガウシアンカーネルで姿勢が分類できること、学習数の増加に伴いガウシアンカーネルの認識率が線形カーネルを上回ることを確認した。

次に、SVM に利用した形状特徴の妥当性を調べるため、人物領域のサイズ特徴と色特徴を用いた比較実験を行った。人物領域のサイズ特徴として、人物領域の横幅と高さの比率、人物領域を囲む長方形内での人物領域の占める割合、検出位置に応じて遠近法で線形補完した人物領域サイズ（人物領域を囲む長方形の対角線）の 3 次元特徴を準備した。一方、色特徴として人物検出領域の上半身部分の色を検出し、 $rg$  空間上 ( $r = R/(R+G+B)$ 、 $g = G/(R+G+B)$ ) で得られるヒストグラム 1,024 次元特徴 (=32×32) を準備した。

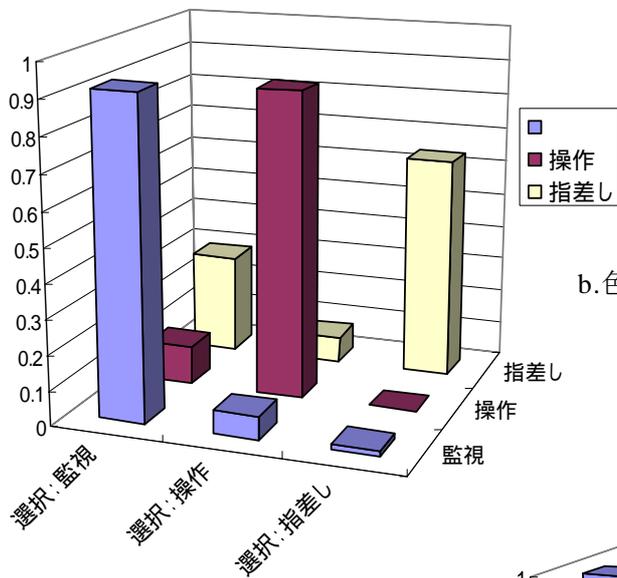
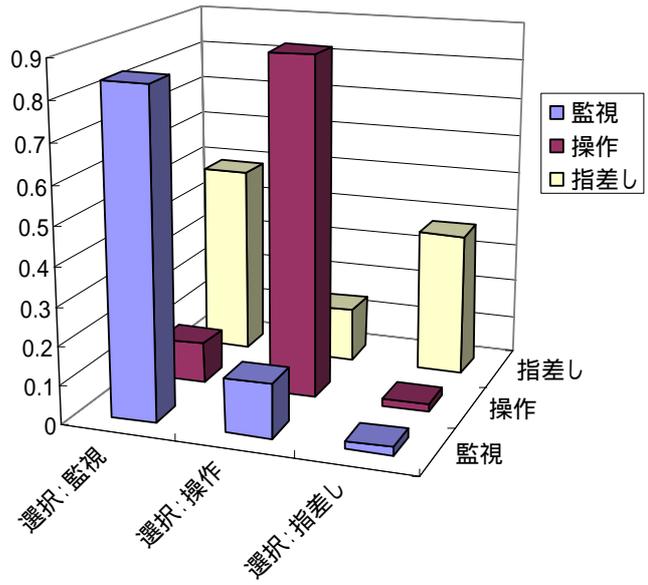
サイズ特徴の実験結果が図 4-8a であり、学習画像全体の認識率は 79.4% (743/935) であった。また、色特徴による実験結果が図 4-8b であり、認識率は 88.5% (828/935) であった。一方、本方式で利用する人物領域の形状特徴の認識率は 99.7% (934/935) と、他の特徴よりも高い値を示すことが確認できた(図 4-8c 参照)。



935 画像から学習に利用し画像の割合 (%)

図 4-7 学習数に伴う認識結果の推移

a. サイズ特徴による認識結果



b. 色特徴による認識結果

c. 形状特徴による認識結果

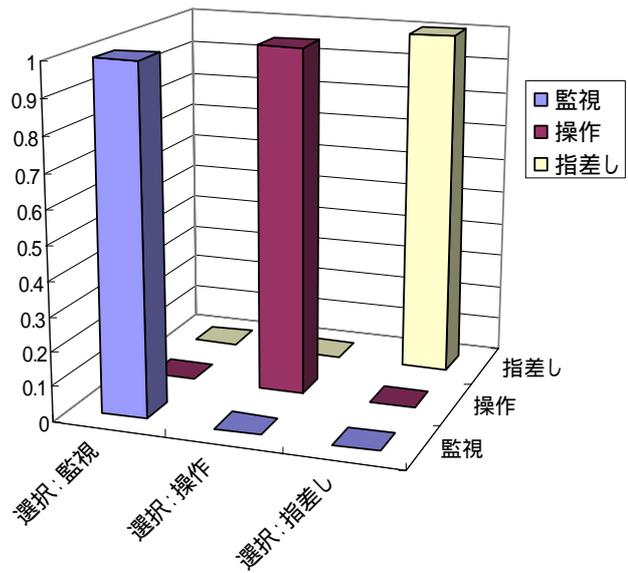


図 4-8 特徴による認識結果の比較

#### 4.4.2 動線検出と姿勢認識実験

学習画像の被写体となった人物の新たな訓練画像と、別人物の新たな訓練画像を使い、動線検出と姿勢認識を行った。

##### (1) 同一人物の実験結果

学習画像と同一人物の新たな撮影画像(340枚:30画像/秒)から、訓練員の姿勢を認識し、その結果をCGに置換した例が図4-9である。図4-9bの足元の線が動線検出結果を示し、身体CGが姿勢認識結果を示している。本実験で姿勢の置換用に準備した5種類のCG画像を図4-10に示す。なお、この実験画像に「電話」姿勢は含まれていない。

訓練画像の3姿勢の認識結果 $P^i_3$ が図4-11aであり、移動状態と静止状態を区別した結果 $C^i_2$ が図4-11bである。また、移動状態時は「監視」か「指差し確認」であるため、全画像を「監視」と「指差し確認」の2姿勢に分類した結果 $P^i_2$ が図4-11cである。図4-11a~cに示す結果 $P^i_3$ 、 $C^i_2$ 、 $P^i_2$ を用いる式(4-2)の姿勢認識結果 $P^i$ に、補正パラメータ $N=4$ として求めた最終結果を図4-11dに示す。図4-11dの上部に目視による姿勢の分類結果を示すが、最終的な自動認識結果は訓練員の姿勢を正しく認識していることが分かる。



a. オリジナル画像



b. 姿勢認識と軌跡検出の結果

図4-9 提案方式による姿勢認識結果例



a. 指差し確認      b. 監視      c. 操作      d. 歩行（移動）

図 4-10 認識結果を表示するために準備した姿勢 CG

## （ 2 ） 別人物の実験結果

学習画像と別人物の連続画像（600 枚:15 画像/秒）の姿勢認識結果を図 4-12 に示す。式 (4-2) と補正パラメータ  $N=4$  で認識結果の自動補正が行われた画像は全体で 32 画像であった。別人物の姿勢認識結果も同一人物と同様に姿勢を正しく認識し、動線検出も移動軌跡を正しく検出できた。

### 4 . 4 . 3 2 万枚の画像による姿勢認識

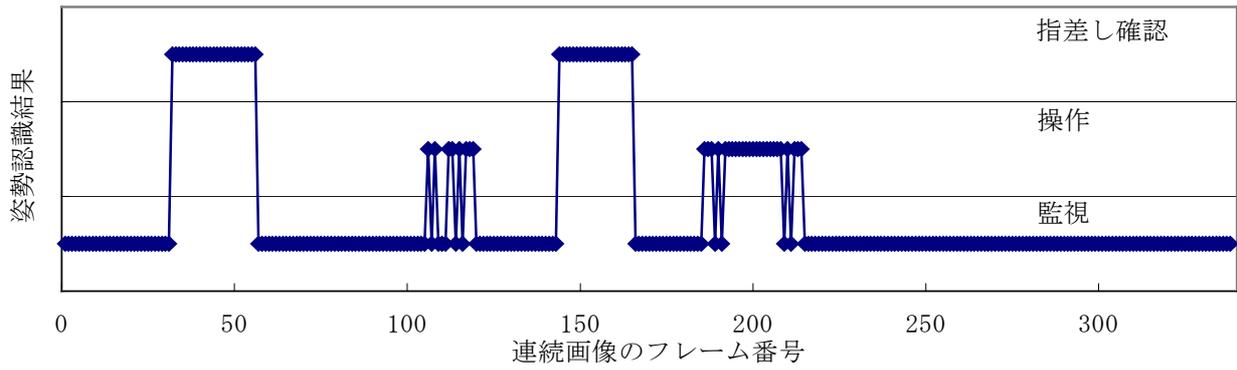
人物の向きや姿勢のばらつき、個人ごとの姿勢の違いに対する認識特性を調べるため、学習画像と同一人物ならびに別人物の連続画像を新たに各 1 万枚(連続画像の読み込み:15 画像/秒)準備し、検証実験を行った。新たに準備した 2 万枚は、1 時間 30 分の録画画像中、左側操作盤で訓練が実施された合計約 22 分である。なお、この実験画像には「電話」姿勢と、電話で話しながら「監視」や「操作」をする姿勢が含まれる。

前述の実験では詳細な結果を調べるため補正パラメータ  $N=4$  としたが、この実験では  $N=10$  とした。また、結果の判定は画像を目視検査した。なお、姿勢間の移行部分で目視区分が曖昧になるのを防ぐため、自動認識結果の姿勢変化時の前後 5 画像を下記の集計から除外した。

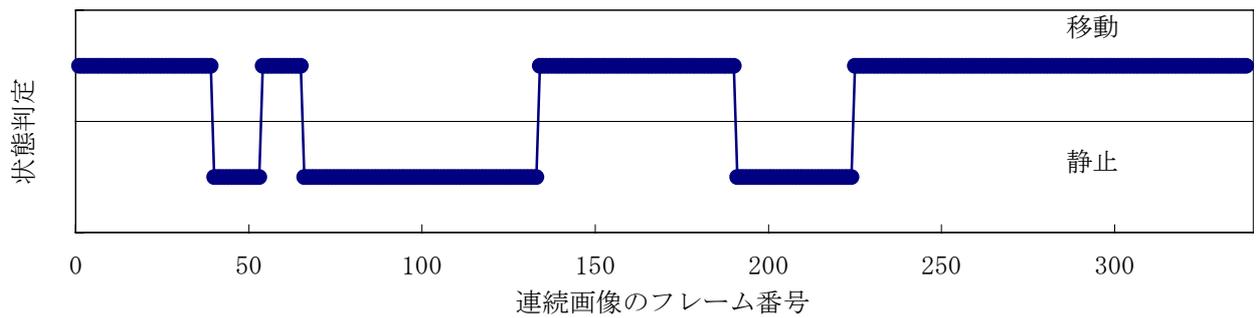
## （ 1 ） 同一人物の実験結果

学習画像と同一人物の姿勢認識率は 99.2% であり、その内の「監視」姿勢 7,688 画像に対し 7,645 画像を正しく認識した。「監視」姿勢が全体の約 8 割を占めるため、「監視」から「操作」「指差し確認」「電話」に移行した回数と各姿勢画像数の合計を表 4-3 に示す。

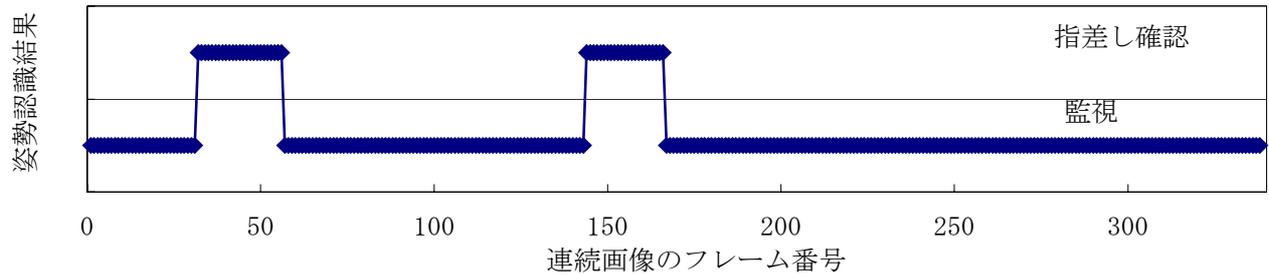
同一人物では認識率 99.2% と高い精度が得られたことから、提案方式が同一人物の向きや姿勢のばらつきに対応できることが分かる。なお、姿勢の誤認識は「操作」姿勢で上部計器の「監視」をするなど、2 つの動作が混合されたことが原因であった。また、未認識は「指差し確認」の腕の部分が検出できず「監視」として誤認識されていた。



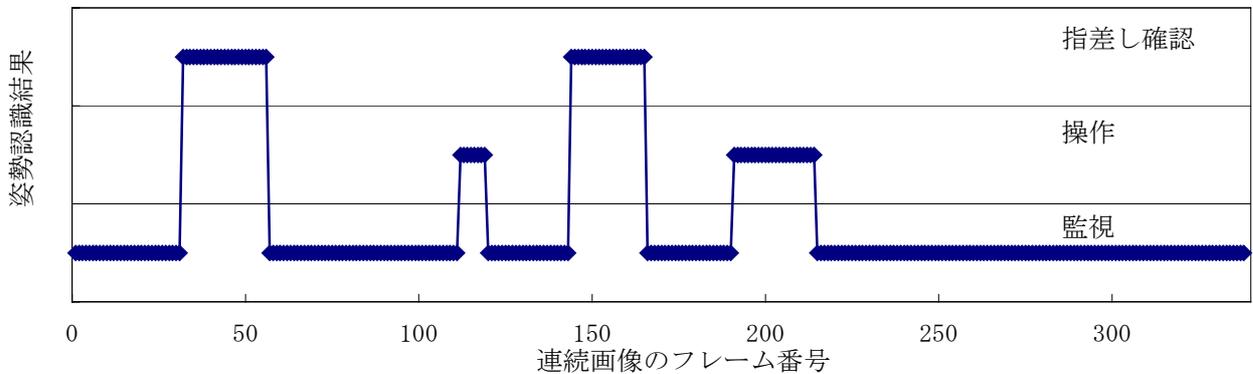
a. SVM による 3 姿勢認識結果



b. 移動状態と静止状態の識別結果



c. SVM による 2 姿勢認識結果



d. 最終的な姿勢認識結果

図 4-11 学習画像と同一人物の姿勢の認識結果

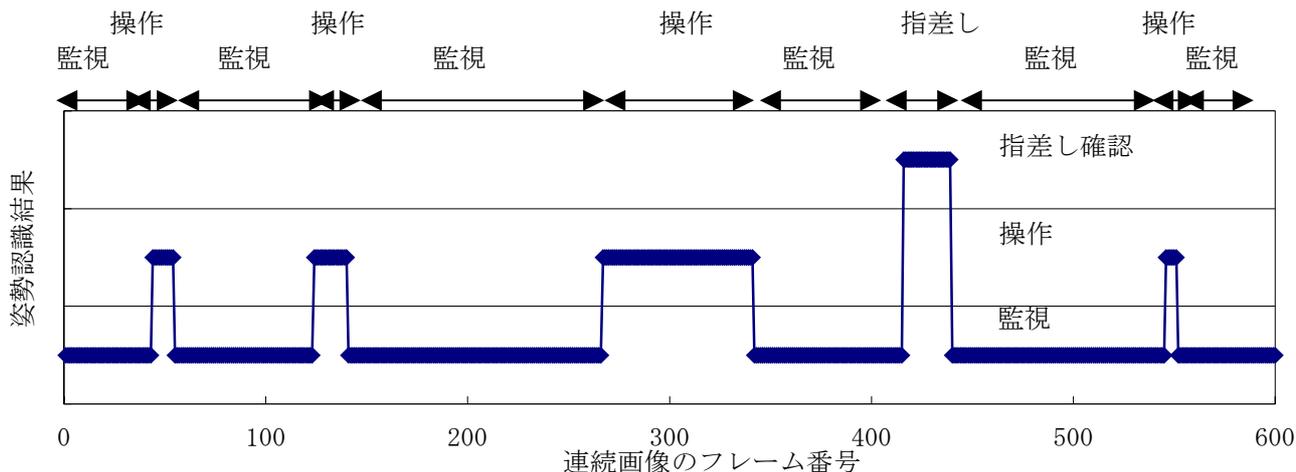


図 4-12 別人物の姿勢の認識結果

表 4-3 学習画像と同一人物の姿勢認識結果

監視から他の姿勢（操作、指差し確認、電話）に移行した回数

| 姿勢            | 正解                | 誤り             | 未検出            |
|---------------|-------------------|----------------|----------------|
| 操作<br>(合計)    | 10 回<br>(1216 画像) | 1 回<br>(43 画像) | 0              |
| 指差し確認<br>(合計) | 13 回<br>(267 画像)  | 0              | 1 回<br>(28 画像) |
| 電話<br>(合計)    | 9 回<br>(161 画像)   | 0              | 0              |

表 4-4 学習画像と別人物の姿勢認識結果

監視から他の姿勢（操作、指差し確認、電話）に移行した回数

| 姿勢            | 正解                | 誤り             | 未検出             |
|---------------|-------------------|----------------|-----------------|
| 操作<br>(合計)    | 14 回<br>(947 画像)  | 3 回<br>(99 画像) | 4 回<br>(165 画像) |
| 指差し確認<br>(合計) | 24 回<br>(1687 画像) | 0              | 2 回<br>(124 画像) |
| 電話<br>(合計)    | 4 回<br>(143 画像)   | 1<br>(21 画像)   | 0               |

## (2) 別人物の実験結果

別人物の姿勢認識率は 95% となり、「監視」姿勢は 6,063 画像中 5,933 画像を正解した。表 4-3 と同様に「監視」から「操作」「指差し確認」「電話」に移行した回数と各姿勢画像数の合計を表 4-4 に示す。

別人物の認識率は同一人物より劣るが、依然として高い認識率 95%を示しており、人物ごとの姿勢の違いに対応できていることが分かる。なお、姿勢の誤/未認識は、「操作」姿勢のように片手を操作盤に置きながら反対の腕で上部の計器に向かい「指差し確認」をしたり、電話を受ける際に外側に足を延ばして反動をつけたりするなど、想定外の個人的な姿勢特性に起因していた。

個人的な姿勢特性の影響を減らす方法として、学習に用いる画像を 1 人だけではなく多人数の姿勢を使う方法が挙げられる。ただし、姿勢の誤/未認識は人物領域の検出誤りからも影響を受けているため、背景や影による雑音で検出ミスが起きにくい人物領域検出方式も今後検討が必要である。

### **(3) 実用性**

現在、教官が監視カメラ画像を見ながら、訓練員の位置と姿勢から運転内容を把握している。提案方式では画像内の訓練員の位置を正確に記録し、その姿勢を高い精度（本実験結果では 95%以上）で認識できるため、教官が画像を見続けることなく特定の動作だけを選び出し警報を出すことができる。また、姿勢認識結果を修正し操作内容を CG 化することで、訓練員のプライバシーに配慮したビデオ教材を作成することができる。その認識修正も 4 種類の姿勢を選択するだけである。そのため、提案方式は運転監視だけでなく教材作成用の道具としても実用的であると言える。

### **(4) パラメータ設定**

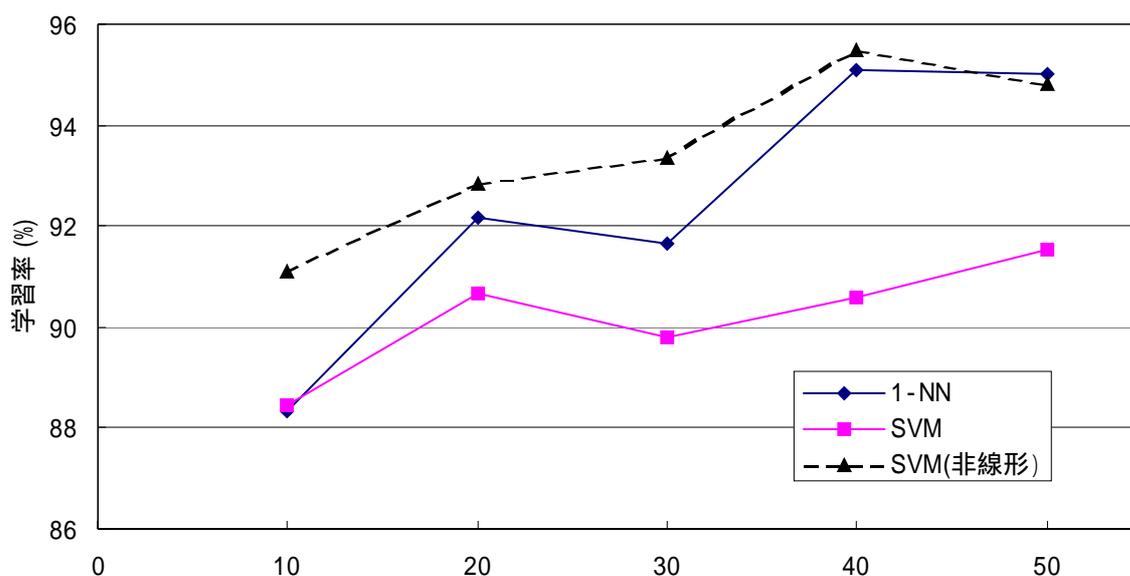
本実験対象の監視カメラで撮影した既存訓練画像ライブラリに対し、本実験で示した成果を直接使用できる。ただし、カメラの設置箇所が変更されている訓練画像ライブラリでは、4.3 節で示したパラメータの調整が必要となる。さらに、カメラの大きな移動により姿勢が表 4-2 の学習画像と比べ明らかに異なる場合や、異種発電所での監視カメラに用いる場合、姿勢ごとにサンプル画像を集め SVM による学習を再度行う必要がある。また、本説明では画像内の左側操作盤を使い説明を行ったが、右側操作盤でも同様に各姿勢の学習と各種パラメータの設定を行う必要がある。いずれの場合も監視カメラごとに一度設定すれば対処可能である。

## **4.5 考察**

### **4.5.1 下半身画像による姿勢認識**

上記の実験では、検出した人物の上半身画像により、「監視」「指差し確認」「操作」の 3 姿勢を認識した。そこで、本研究では下半身でも 3 姿勢の認識が可能かどうかを、4.4.1 で利用した 935 画像を使い学習率を調べる実験を行った。その結果が図 4-13 である。図 4-13 には SVM との比較のために 1-NN の結果も示している。

図 4-13 より学習に利用する画像数が増えるに従い 1-NN の精度が上がる様子が分かる。図 4-13 から下半身画像でも姿勢を認識できることが分かるが、図 4-7 に示した上半身による最大学習率



933 画像から学習に利用し画像の割合 (%)  
 図 4-13 下半身画像による 3 姿勢の学習結果

98%に比べ学習率が低いことが確認できた。

## 4.5.2 提案方式の限界と精度向上方法

### (1) 理論的な限界

提案方式は機械学習を使用しているため、人物検出の考察で述べたように未認識と誤認識の発生が避けられない。例えば、操作盤に手をかけながら計器を監視する状態など、想定外の姿勢をした場合に未認識あるいは誤認識が発生する。このような場合、「操作」と「監視」のどちらを正解とするか人でも判断が難しい。

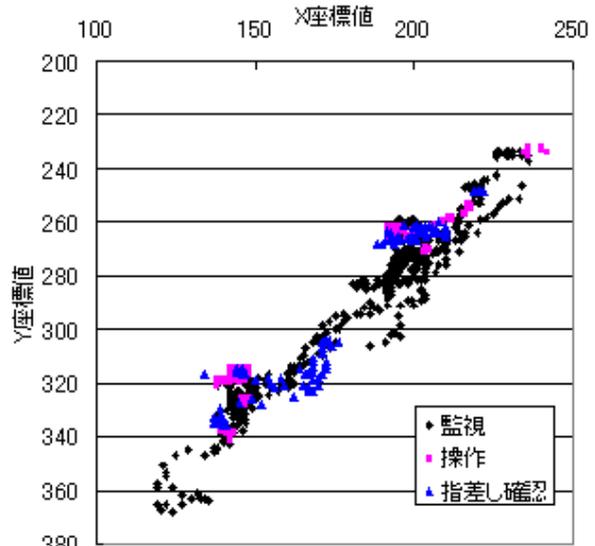
### (2) 対策

このような未認識と誤認識に対し、典型的な発電所操作手順を参考に誤りを補正する方式などが有効である。例えば、本実験で得られた姿勢認識結果を、検出位置でプロットした結果が図 4-14 であり、操作や指差し確認する場所が決まっていることが分かる。図 4-14a が画像例であり図 4-14b が訓練員の画像内での認識姿勢である。図 4-14c は検出した人物サイズを正規化して、画面の Y 軸での位置で認識姿勢をプロットしたグラフである。図 4-14 のように、運転内容により操作する部分や指差し確認する部分が決まっているばかりでなく、正しい操作手順もマニュアル化されているため、訓練シナリオにありえない単発的な姿勢を排除することで認識率を上げることが可能である。

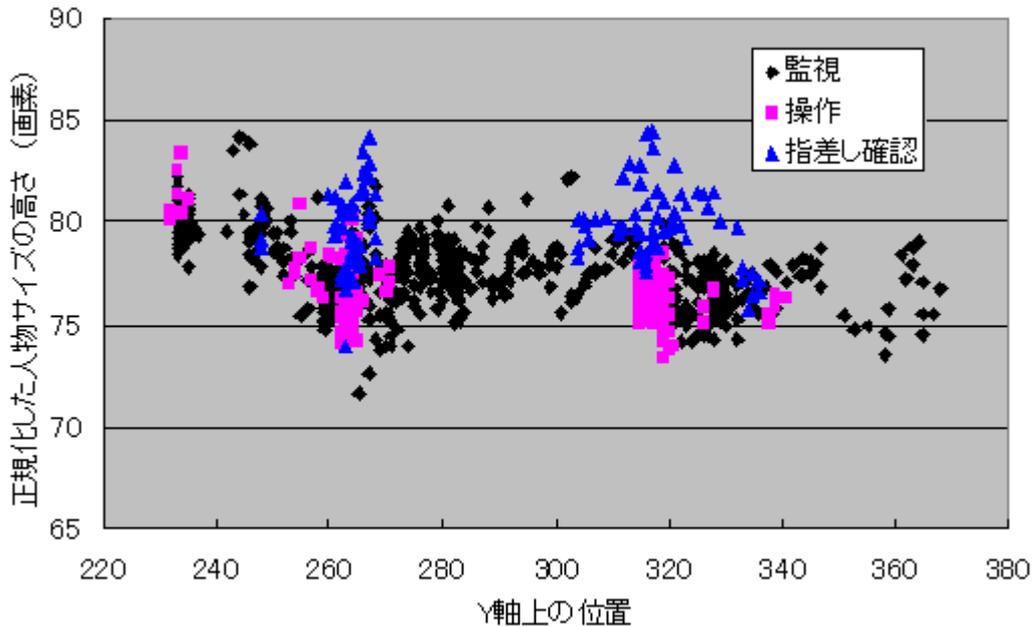
提案方式を使い初期の姿勢が分かれば、その姿勢を参考に人物に身体モデルを当てはめることが容易になる。そこで、既存の身体モデル当てはめ方式と組み合わせるシステム化により、未/誤認識を補正する方法も考えられる。



a. 操作盤の前を切出した画像



b. 操作盤の前での姿勢認識結果



c. 操作盤前での人物の Y 軸上での位置と正規化した人物サイズ

図 4-14 姿勢認識結果を検出位置に合わせて表示した結果

## 4.6 本章のまとめ

本章では 3 章の人物検出と同様の枠組みで、原子力発電所運転訓練を撮影している監視カメラの画像から、訓練員の操作姿勢を認識する方式を示した。

従来の画像処理方式では常に変化する訓練員の姿勢を認識することが困難であった。そこで、原子力発電所運転に特有な 4 種類の姿勢に特化し、動きの状態での姿勢認識ルールを切替る姿勢認識方式を提案した。提案方式により、訓練員の向きや姿勢のばらつきに影響

されることなく安定に姿勢を認識できることを実験で示した。ただし、機械学習を利用しているため、未認識や誤認識の発生は避けられない。そのため、典型的な操作手順の参照、あるいは身体モデルの当てはめ方式との組み合わせなどにより、認識精度向上を図る基礎研究が今後も必要である。

提案方式は、発電所運転の正しい操作手順を作業員の姿勢と位置とともに登録することで、通常の運転から逸脱した運転操作時に警報を出す画像監視システムの作成を可能にする研究成果である。また、おおまかな姿勢が分かる程度で役に立つスーパーなどのマーケティング調査や犯罪パターンの解析などにも応用できる可能性がある。

## 第5章

### 複数の既知人物の個人特定

#### 5.1 はじめに

3章では人物を実時間で検出する方式を示し、4章では検出した人物の動作を把握するための姿勢認識方式を示した。将来的に、目視監視を補助あるいは代行する画像監視システムには、監視員のように既知の人物や一度検出した人物を認識する能力が必要である。そこで、本章では監視カメラ画像から検出した全身画像を使い、人物を認識する方式を議論する。

画像から顔を検出し、人物を認識する研究成果が数多く報告されている。しかし、監視カメラで検出した全身画像から個人を認識する課題は、認識に利用できる安定した特徴が決められないため研究として取り扱いが難しく、近年、歩き方から人物を認識する研究[56][92][93]が始められた段階である。未だ、監視カメラで検出した人物の全身画像から、3章の検出や4章の姿勢認識と同様の枠組みで個人を認識する研究成果が報告されていない。そこで、本章では人物の全身画像を使い、動きによる前処理と機械学習による後処理で人物を認識する方式の実現を目指す。

この困難な課題に対し、室内の限られた環境の中で、既知の人物を検出し認識する問題を設定し、3章や4章と同様の枠組みで対処する方式を提案する。提案方式の認識処理で利用するSVMは2分類問題の機械学習であり、多人数を分類するにはSVMを組み合わせる必要がある。SVMの多クラス化に関して、幾つかの多段戦略法が提案されている[88][94][95][96]が、理論と実用の面から優劣をつけがたい状態である。そこで、SVMの多段戦略を実験結果により比較検討し、人物認識の実現可能性を議論する。

本章では3.2節で監視環境と監視対象を示し、3.3節で機械学習を利用した人物認識方式を提案する。3.4節では、はじめに4人の認識実験結果を示し、次に8人の被験者を16日間撮影した画像を用いた認識実験結果を示す。3.5節の考察では、提案方式の理論的な限界とその対策を示す。また、静止画像データベースCOIL[97]の100物体（7,200画像）による定量的な実験結果を示す。3.6節で本章のまとめを示す。

#### 5.2 監視環境と監視対象

本章では、研究室のコーヒーマシンを利用する研究室メンバーの認識問題を取上げる[98][99]。監視カメラはコーヒーマシンの前にあり、コーヒーマシンと監視カメラの距離は約4.5m、監視カメラの高さは約2mである。背景や照明はほぼ一定で、カメラのフォーカスも固定である。被験者はコーヒーマシンを使うために訪れるだけで、コーヒーマシンの利

用方法や姿勢などに関して何も制約を設けないものとする。なお、実験では人物の服装は一日を通して、ほぼ同じであると仮定している。また、監視カメラの画像は640×480画素とする。

なお、コーヒーマシンの前で起きる画像変化は、コーヒーマシンを利用する人物が起こす変化だけである。そのため、背景差分やフレーム間差分で得られる変化は、人物の動作で発生した変化であると見なせる。

### 5.3 認識システムの概要

3章や4章と同様に、前処理と認識の2つのモジュールで人物認識処理を構成している。図5-1に処理概要を示す。監視カメラの画像から背景差分等で人物領域を検出し、認識に用いる形状や色特徴を抽出する。その後、抽出した特徴を使い人物を認識する。

#### 5.3.1 前処理

前処理では、人物領域の検出と、検出した人物領域から特徴を計算する2種類の処理を行う。

##### (1) 人物領域の検出

人物領域の検出は2.2節で示した画像差分を利用する。まず、人物を検出するために、連続する2画像  $f_i(x,y)$ ,  $f_{i-1}(x,y)$  のフレーム間差分画像  $f_{d1}(x,y)$  を作成する。

$$f_{d1}(x,y) = |f_i(x,y) - f_{i-1}(x,y)| \quad (\text{但し } f_{d1} < T_1 \text{ なら } f_{d1} = 0)$$

人物が移動している場合、このフレーム間差分画像  $f_{d1}$  により移動方向に人物の輪郭が出現する(3章図3-18参照)。次に、人物領域を検出するため、フレーム間差分画像  $f_{d1}$  内で輪郭にはさまれた領域を人物領域画像  $f_{d2}$  として検出する。さらに、背景画像  $f_b$  と最新画像  $f_i$  により背景差分画像  $f_{d3}(x,y)$  を作成する。

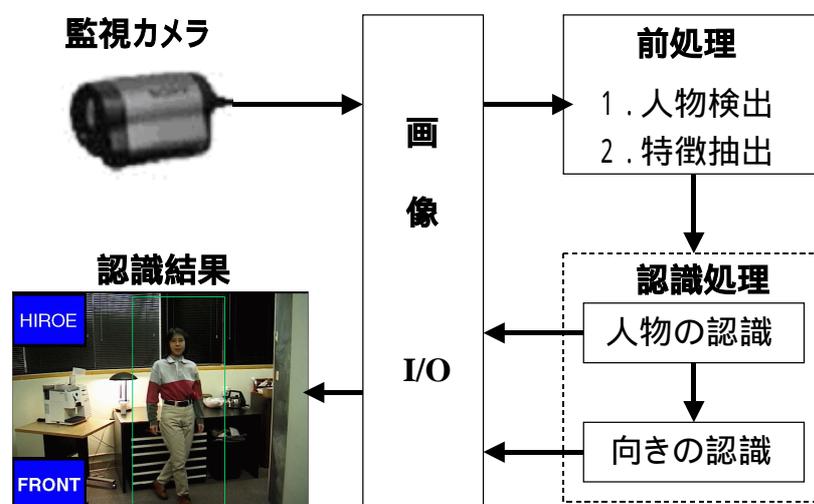
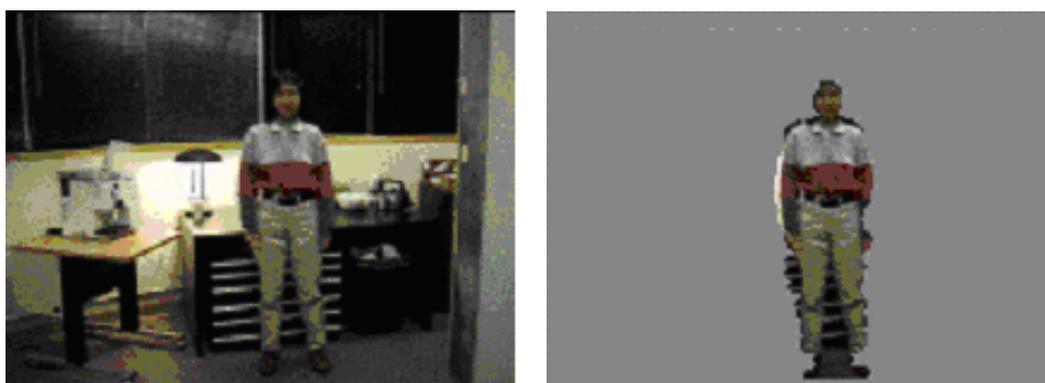


図 5-1 人物認識処理の概要



現画像

人物領域の検出結果

図 5-2 移動人物の検出例

$$f_{d3}(x,y) = |f_i(x,y) - f_b(x,y)| \quad (\text{但し } f_{d3} < T_2 \text{ なら } f_{d3} = 0)$$

この2枚の人物領域画像 $f_{d2}$ と背景差分画像 $f_{d3}$ の変化領域 ( $f_{d2} \neq 0$  or  $f_{d3} \neq 0$ ) を人物領域の検出画像 $f_o$ として作成する。ただし、検出画像 $f_o$ の人物領域画素数の合計が閾値 $T_o$ を超えない場合、画像内に人物がないものと見なし最新画像 $f_i$ を背景画像 $f_b$ の算出に利用する。実験では変化を含まない5画像分の平均を背景画像 $f_b$ として算出し、常に背景画像 $f_b$ を更新している。また、実験では閾値 $T_1$ ,  $T_2$ ,  $T_o$ を一定値とした。図5-2に最新画像 $f_i$ と人物領域の検出画像 $f_o$ の一例を示す。

## (2) 特徴抽出

SVMの認識処理では、物体の色、輪郭、シルエットの重心、テクスチャ、局所自己相関など、各種の特徴を使用できる。画像処理による物体認識では、認識に用いる特徴が認識精度に大きな影響をあたえるため、これまでに多くの実験結果[59][88][100][101][102]が報告されている。しかし、今回の課題にどの特徴が役立つか不明なため、色、形状、サイズ、局所自己相関等の下記の4種類の特徴を経験則で選んだ。いずれも(1)で作成した人物領域から下記の特徴を計算するものとする。

### ① 色ヒストグラム (RGB : 96次元)

人物領域のRGB値を色ごとにヒストグラム化した色特徴を準備する。各色を32分割し、合計96次元 (=32bin×3色) のヒストグラム特徴を画像から算出する。

### ② 正規化rg (Normalized rg : 1,024次元)

RGBの値から $r=R/(R+B+G)$ ,  $g=G/(R+B+G)$ を計算し、 $r$ 軸と $g$ 軸の2次元空間でのヒストグラムを求める。 $r$ 軸と $g$ 軸の座標をそれぞれ32分割し、合計1,024次元(=32×32)の特徴を画像から算出する。

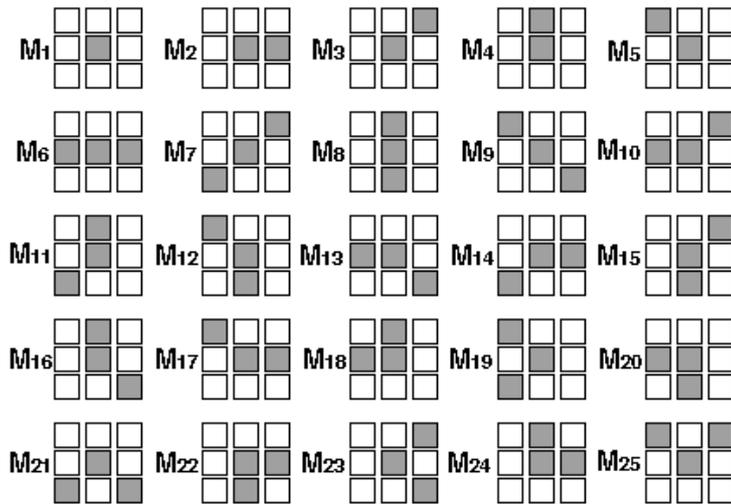


図 5-3 局所自己相関の局所パターン[103]

③ 色ヒストグラム+人物のサイズ (RGB+Global Shape : 136次元)

人物の身長や横幅なども人物認識に使えるのではないかと考え、①のRGBのヒストグラムに人物領域の縦と横の画素数ヒストグラムを加えた。背景を除去した人物領域画像を縦10分割/横30分割し、縦横独立に画素数のヒストグラムを求める。その結果に①のRGBのヒストグラムをあわせ、合計136次元(=32×3+10+30)の特徴を算出する。

④ 局所自己相関 (Local Shape : 75次元)

人物領域の色の局所自己相関を特徴として計算する。この局所自己相関の計算には、図5-3に示す局所パターン[103]を利用する。ただし、脳研究のモデル化の成果に視覚野の方向を感知する方向受容野群の働きがMAXオペレーションで説明できるとの仮説[104]が提案されているため、2次の相関(図5-3でのパターンM<sub>6</sub>~M<sub>25</sub>)を方向受容野と見立て、最大値を示す相関だけを特徴に加算した。本実験では画像から合計75次元(25×3 : 25パターンに3種類の色)の特徴を抽出する。

ここで図5-3のパターンをM<sup>i</sup>, i=1, ..., 25, とし、V<sub>k</sub>を3×3画素とすると、最初の5つは単純な線形加算  $\sum_k V_k \cdot M_i$  だが、残りは非線形加算  $F = \sum_k C(k, i)$  とする。ここでC(k, i)は、

$$C(k, i) = \begin{cases} V_k \cdot M^i & \text{if } V_k \cdot M^i = \max_j (V_k \cdot M^j) \\ 0 & \text{その他} \end{cases}$$

であり、相関の最大値だけを検出する関数である。さらに、人の視覚特性を参考に、R+G-B, R-GとR+Gの3チャネル[105]に対して上記相関を計算した。

### 5.3.2 SVMの多段戦略

SVMは2クラス分類問題用の学習方式であり、複数の対象を分類するにはSVMを組み合わせて利用する必要がある。SVMは入力画像を $x$ 、学習画像を $x_i$ とし、入力画像と学習画像の距離を計算するカーネル関数 $K$ を使い、

$$f(x) = \sum_{i=1}^m d_i K(x, x_i)$$

と表記できる。 $d$ は重み係数である。画像を $n$ クラスに分類するにはSVMを組み合わせて $k$ 個の関数

$$f_k(x) = \sum_{i=1}^m d_{ki} K(x, x_i)$$

を準備する必要がある。学習で準備する関数の数 $k$ や各関数 $f_k(x)$ の組み合わせ方にバリエーションがあり、その理論的な優劣は議論されている段階である[115][127]。その中でも最も代表的な方式として、1対他方式(One vs. All others)と2クラス対方式がある。

1対他方式は、クラス $k$ に属す学習画像を $x_i^k$ とすると、全てのトレーニングパターン、

$$\{x_1^1, \dots, x_{l_1}^1, \dots, x_1^n, \dots, x_{l_n}^n\}$$

を、

$$\begin{array}{l} 1. \{x_1^1, x_2^1, \dots, x_{l_1}^1\} vs \{x_1^2, x_2^2, \dots, x_{l_2}^2, \dots, x_1^n, x_2^n, \dots, x_{l_n}^n\} \\ \vdots \\ n. \{x_1^n, x_2^n, \dots, x_{l_n}^n\} vs \{x_1^1, x_2^1, \dots, x_{l_1}^1, \dots, x_1^{n-1}, x_2^{n-1}, \dots, x_{l_{n-1}}^{n-1}\} \end{array}$$

と分割し、SVMにより $n$ 個の関数 $f_1 \sim f_n$ を事前に作成する。その後、新たな入力画像 $x$ を、

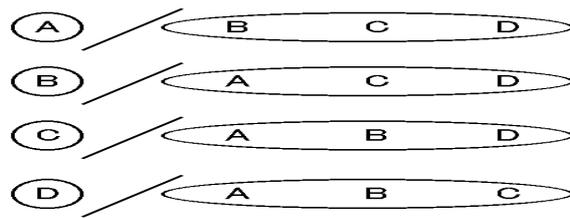
$$m = \arg \max \{f_1(x), \dots, f_n(x)\}$$

の最大を示すクラスに分類する戦略である。複数のクラスが選ばれた場合、クラスの選択回数や識別面からの距離などが最終判定の基準に使われている。図5-4aに4クラスの分類方式を示す。

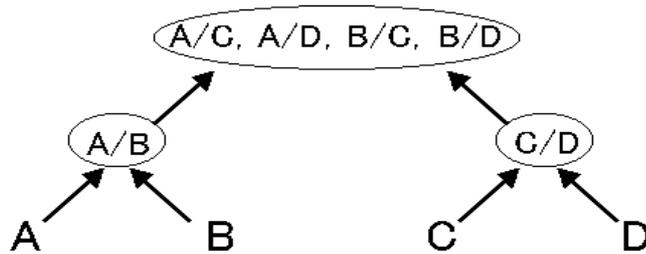
一方、2クラス対方式の場合、2クラス $i$ と $j$ の学習画像

$$\{x_1^i, x_2^i, \dots, x_{l_i}^i\} vs \{x_1^j, x_2^j, \dots, x_{l_j}^j\}$$

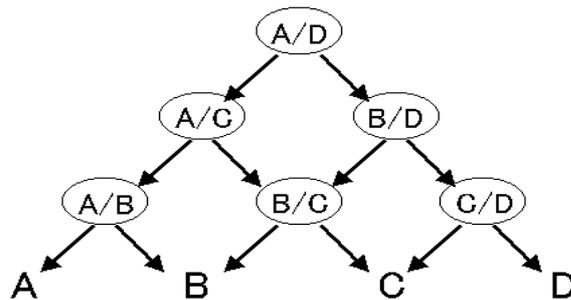
を分類する関数 $f_k$ をクラス対 $(i, j)$ ごとに作成し、新たな入力画像 $x$ を関数 $f_k(x)$ を組み合わせる戦略である。関数 $f_k(x)$ の組み合わせ方として、図5-4bに示すようなテニストーナメント方式や図5-4cに示す決定木を使う方式がある。本章では図5-4b, cの図の書き方から便宜上ボトムアップ方式、トップダウン方式と呼んで区別する(詳細は2.5.4節を参照)。



a. 1対他による比較方式



b. ボトムアップによる比較方式



c. トップダウンによる比較方式

図5-4 SVMの多段戦略による多クラス分類

## 5.4 既知人物の認識と向きの認識実験

本節では、2種類のデータによる実験結果を示す。実験では監視カメラのフォーカスやホワイトバランス、室内の照明等は一定で、背景も実験期間中はほぼ一定であった。

はじめに、図5-5に示す4人の認識実験と、図5-6に示す各人の4方向の向き（前、後、左、右）の認識実験を示す。4人の認識実験では、複数の画像特徴を3種類の多クラスSVMで認識した結果を示す。なお、他の手法との比較のため、広く使われているk-NN (*k*-Nearest Neighbor)による認識結果をあわせて示す。次に、4人の認識実験で優秀な結果を示した画像特徴を使い、16日間に渡り撮影した8人分の1,127画像による認識実験結果を示す。なお、被験者は、コーヒーマシンの利用頻度が高い上位8人である。



図 5-5 4人のシルエット検出例



図 5-6 4方向のシルエット検出例

#### 5.4.1 実験1：人物の認識と向きの認識

実験では4人の被験者を4方向（前後左右）から各40枚撮影した画像を、人物の学習画像とした。さらに、各方向の画像から10枚ずつ選び、被験者ごとに40画像を向きの学習画像とした。つまり、図5-5と図5-6に示すような人物検出結果を、人物ごとに160枚（4人の合計640枚）準備し、人物と向きを別々に機械学習した。

機械学習では、1つの多クラスSVMを人物認識用として準備し、4つの多クラスSVMを人物ごとの4方向の認識用として準備した。そのため、合計5通りの多クラスSVMを準備した。

新たに撮影した4人の合計418画像に、人物認識と向きの認識を行った。図5-7に実験結果の一例を示す。図5-7の各画面の左上ウィンドウが人物の認識結果(名前)であり、左下ウィンドウが向きの認識結果である。また、人物を囲む長方形の枠はシステムが捕らえた人物位置である。システムは始めに人物を認識し、次ぎに向きを認識する段階的な処理を行っている。

図5-8に多クラスSVMの各多段戦略と実験に利用した4種類の画像特徴による4人の認識結果を示す。図5-9に向きの認識結果を示す。なお、他方式との比較として図5-8と図5-9には $k$ -NNによる認識結果を示した。この実験で使用したSVMはいずれも線形カーネルである。

図5-8に示す人物の認識結果は、局所的な形状特徴を利用するよりも色特徴を利用した結果が高い認識率を示している。特に、正規化rg特徴(1,024次元)が、認識方式に関係なく100%の認識率を示している。RGBの色特徴に身長や横幅を加えるよりも、RGB(96次元)だけのほうが高い認識結果を示している。当初、色特徴に身長や横幅を加えることで

認識精度が向上することを想定していた。しかし、実際は被験者の動きにより、監視カメラと被験者の距離が常に変化することが原因となり、認識精度を落とす結果となった。一方、局所自己相関を使った場合、SVM は 90%後半の高い認識率を示しているが、 $k$ -NN では 80%半ばとなっている。これは SVM の汎化能力により得られた差であると考えられる。実験では処理速度の関係から線形カーネルを利用したが、データを高次元空間に展開する非線形カーネルを利用することで、認識精度を上げることができると考えられる。

図 5-9 に示す向き認識の結果は、正規化 rg 特徴 (1,024 次元) が最も高い認識率を示しているが、図 5-8 の人物認識に比べ認識精度が落ちていることが分かる。人物認識と同様、RGB の色特徴に身長と横幅を加えた特徴が最も低い認識率を示している。一方、局所自己相関特徴は、正規化 rg 特徴の約 1/13 の特徴量にも関わらず高い認識率を示す結果となった。

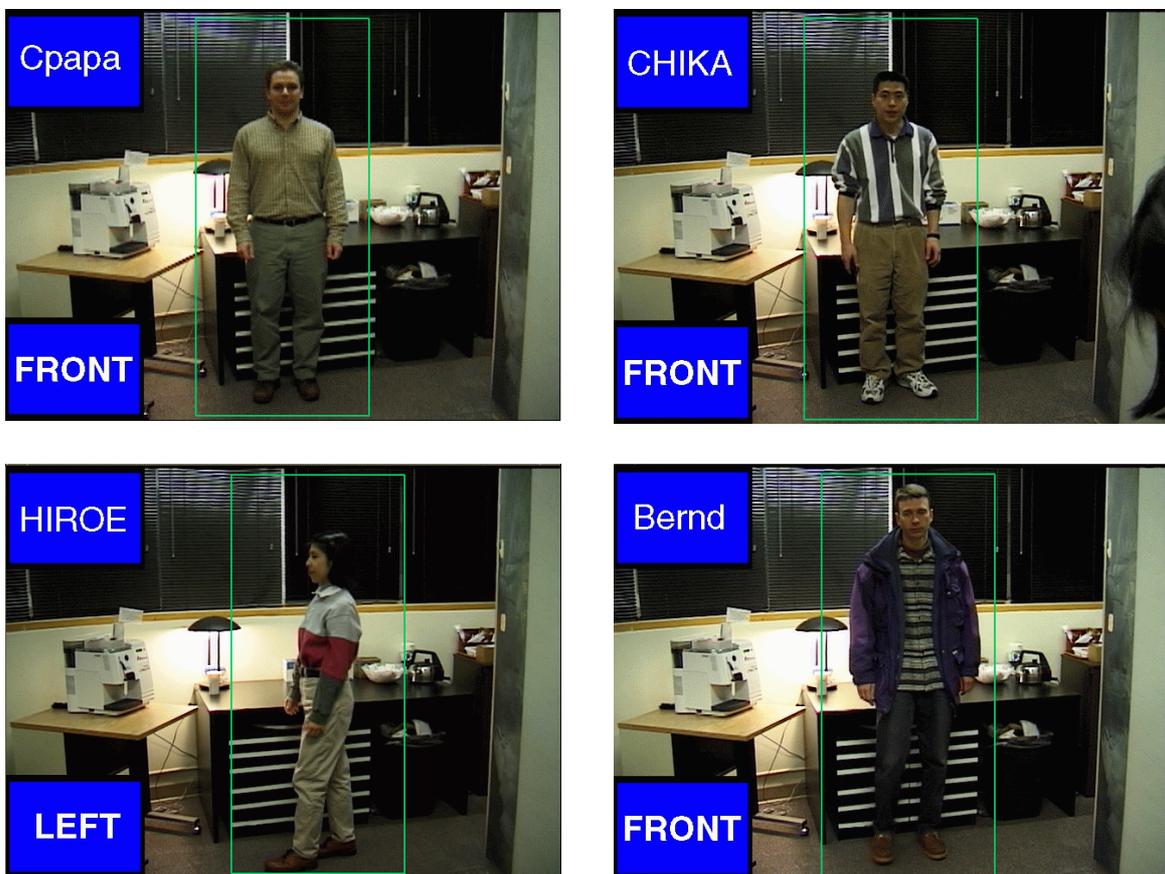


図5-7 服装による人物と向き認識の実験結果例

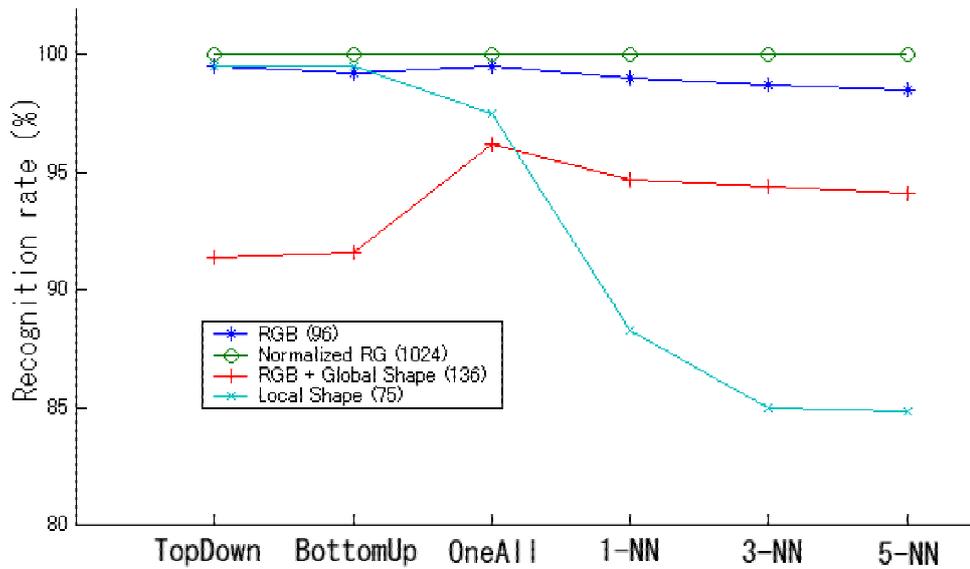


図5-8 4種類の特徴を使った人物認識の実験結果

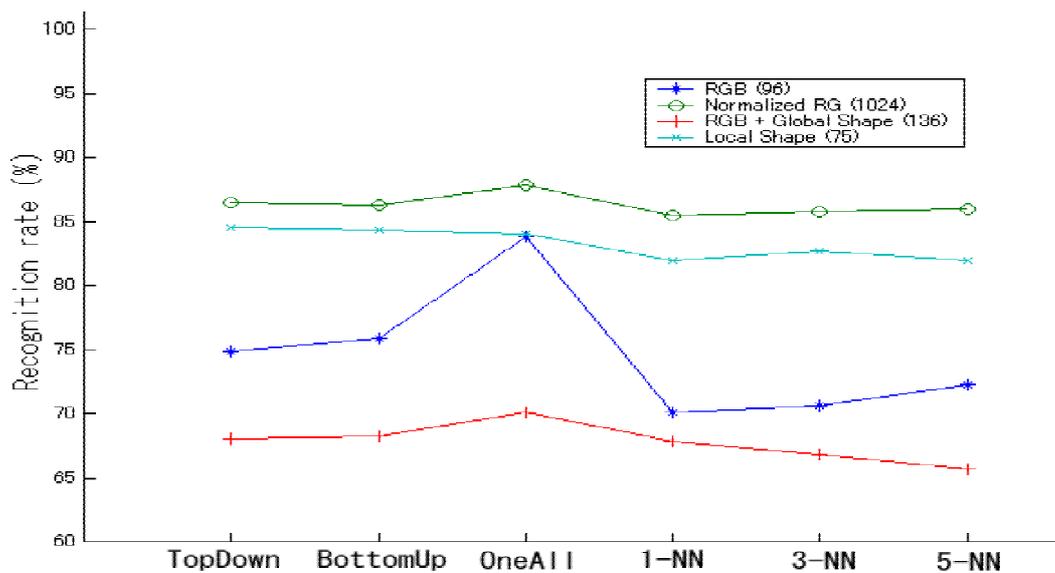


図5-9 4種類の特徴を使った向き認識の実験結果

#### 5.4.2 実験2：16日間の人物認識

図 5-8 と図 5-9 で正規化 rg 特徴が最も高い認識結果を示したため、今度は 16 日間に渡り向きを考慮せず人物を自動的に撮影した 1,127 画像（最も撮影回数の多い上位 8 人）を準備し、SVM と 1-NN による認識実験を行った。実験用に準備した 8 人の画像例を図 5-11 に示す。得られた 1,127 画像を、学習用画像（トレーニング用）と認識実験用画像（テスト用）に 9 対 1, 5 対 1, 1 対 1, 1 対 5 に分け、認識実験を行った結果が図 5-10 である。

認識結果は 1-NN がいずれの線形の SVM よりもわずかに高い認識率を示したが、非線形カーネルを SVM に利用すると 1-NN に比べ 4~5% 程度認識率が向上した。実験では実時

間処理に重点を置いているため、計算量のかからない線形カーネルを SVM に利用したが、計算速度を気にする必要がない録画ビデオの解析等では、非線形カーネルが有効である。一方、SVM はトレーニング用の画像数が少なくなると、2クラス対方式（トップダウン方式とボトムアップ方式）よりも1対他方式が高い認識率を示す結果となった。この理由は、2クラス対方式の各学習で使う画像数が、1対他方式で使う画像数よりも少ないことに起因していると考えられる(詳細は 5.5.5 参照)。

図 5-10 の結果より、同日に撮影した画像であれば監視領域に現れた人物を識別できる可能性があることが分かった。そこで、さらに 15 日間分の画像を学習し、残り 1 日分を認識した実験結果を表 5-1 に示す。線形カーネルの SVM よりも k-NN のほうが 4%~7%ほど高い認識率を示したが、非線形カーネルを利用すると k-NN とほぼ同じ認識率を示した。なお、SVM の多段戦略では、残り 1 日の画像数が少ないことから 2クラス対方式よりも1対他方式が高い認識率を示している。

表 5-1 の結果は同日に撮影した画像を利用した図 5-10 の結果に比べ低い認識率だが、いずれの方式も 50%前後の認識率が得られている。撮影日が異なると、着ている服が変わるため低い認識率を予想していたが、偶然確率よりも高い 50%程の認識率が得られた。これは、個々人の服装の好み等を反映した結果であると考えられる。

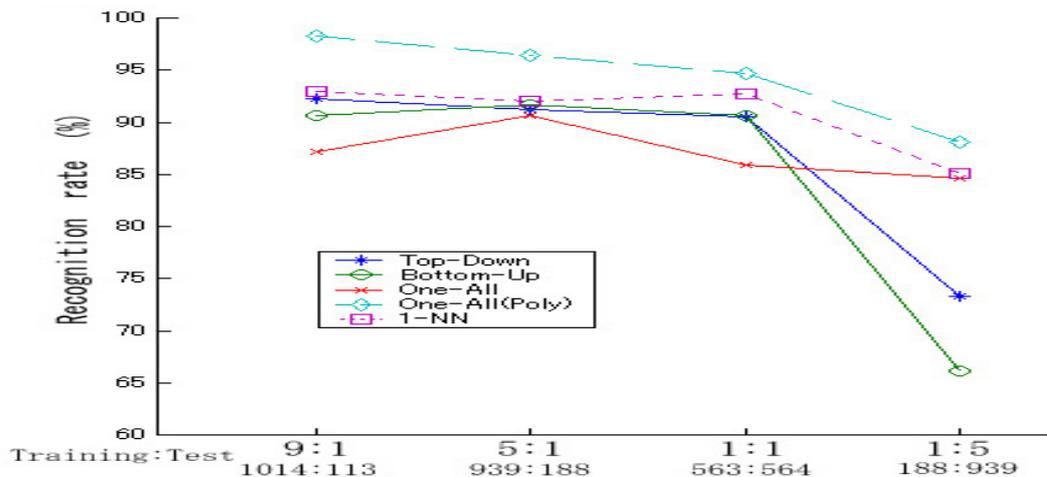


図 5-10 8 人の 16 日間分の撮影画像を利用した認識実験結果

表 5-1 15 日間の画像を学習し残り 1 日分を認識した結果

| 種類      | SVM    |        |                      | k-NN |      |      |
|---------|--------|--------|----------------------|------|------|------|
|         | トップダウン | ボトムアップ | 1 対他                 | 1-NN | 3-NN | 5-NN |
| 認識率 (%) | 45.9   | 45.9   | 49.2<br>(poly: 52.9) | 53.3 | 50.0 | 50.8 |

(15日間1,005画像、残り1日122画像)



図 5-11 実験に利用した 8 人の画像例

## 5.5 考察

### 5.5.1 SVMの多段戦略による認識

学習画像数が充分確保できている状態では、SVMの多クラス戦略はいずれも優劣を付けがたい結果となった。ただし、学習画像数が少ない場合(図5-10,表5-1)、1対他方式が2クラス対方式(ボトムアップ方式、トップダウン方式)よりも高い認識率を示している。全クラスの学習画像を使う1対他方式に比べ、2クラス対方式は2クラスの画像だけで学習を行うため、学習画像数が少ない場合1対他方式が有利になると考えられる。例えば、クラス数が10で各クラスの学習画像が10枚とすると、1対他方式では10画像対90画像の学習となる。一方、2クラス対方式は10画像対10画像の学習であり、1対他方式に比べ明らかに判別空間内でのデータが素になる。そのため、データが密な1対他方式のほうが望ましい境界の近くに識別面が作成される可能性が高い。

しかし、実用面からすると、多くのクラスと大量の学習画像が得られている場合、1対多方式は全ての画像を学習に毎回使うため、2クラス対方式に比べ学習時間が大幅にかかる問題がある。特に、既存の学習結果に新たなクラスを追加する場合、1対他方式は全てを学習しなおさなければならないが、2クラス対方式は新たなクラスと既存のクラス間で学習を行うだけで、過去に実施した学習結果をそのまま利用できる利点がある。

2クラス対方式としてボトムアップ方式とトップダウン方式を比較したが、本実験結果では両者の優位な差は出なかった。むしろ、SVMの多段戦略よりも、どのような画像特徴を利用するかが、認識率に大きな影響を与える実験結果となっている。そこで、次項では複数のパスで1つのクラス選択にたどり着くトップダウン型の戦略を取上げ、最も認識精度が高い正規化rgを使い、変形しない物体を用いた定量的な実験結果を示す。

### 5.5.2 物体の認識特性

前節の実験に用いた人物画像は、姿勢や服装が常に変化しているため、一定の評価が難しい。そこで、形状不変の人形や物体を撮影した画像、ならびに100物体の画像をデータベース化したコロンビア大学のCOIL(Columbia Object Database Library)[97]を用いた認識実験結果を示す。いずれも、前節の人物認識実験で良好な結果を得た正規化rg特徴(1,024次元特徴)を、トップダウン型の多クラスSVMで認識した結果である。

#### (1) 形状不変物体の認識率と処理速度

はじめに、固定カメラであらかじめ5つの物体を手にとり各100枚(5物体で合計500枚)のサンプル画像を撮影し、SVMの学習画像とした(図5-12参照)。認識実験用に学習画像とは別に新たに各400枚(5物体で合計2,000枚)の画像を撮影し、物体認識を行った。実験で利用した学習画像とテスト画像の一例を図5-12に示す。また、システムの出力結果例を図5-13に示す。図5-13では、各画像内の左下文字列が認識した物体名を示している。



図 5-12 形状不変な物体領域を検出した画像例

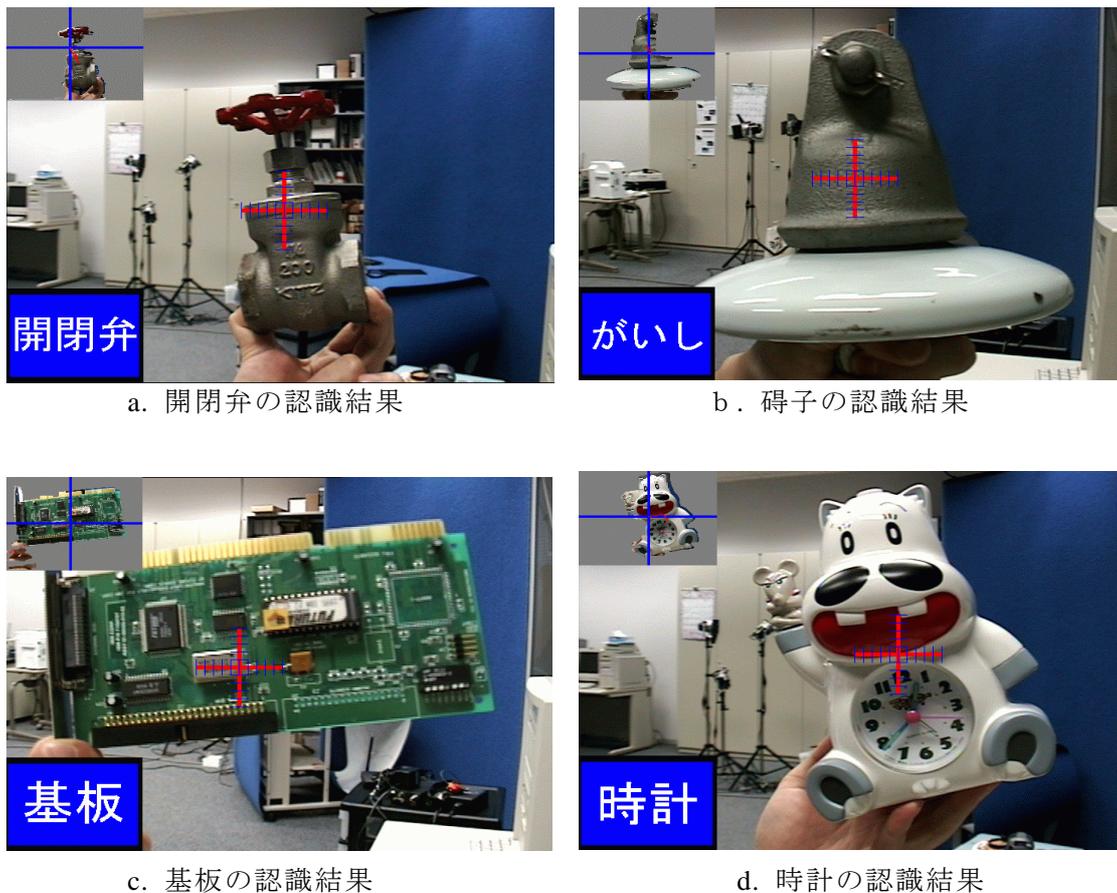


図 5-13 形状が変形しない物体認識の結果例

SVMによる5物体の認識率は100%であった。試作システムの動作速度は2台の計算機SGI-O2,R1200,270MHzを並列で使い、1台の計算機で物体検出と認識処理を行い、残りの1台で物体の回転等を計測した[135]。SVMによる1画像の認識速度は約0.06秒であった。

## (2) 物体認識特性

提案方式の物体認識特性を調べるため、コロンビア大学の画像ライブラリCOIL100の7,200画像を使い認識実験を行った。COILは物体を5度刻みに回転しながら撮影した100物体の画像(合計7,200枚 =  $100 \times 360 \text{度} / 5 \text{度}$ )が収録されている。COILの中には、色や形状が似た物体も多く含まれている。図5-14にCOILの画像例を示す。

各物体の画像数の5%, 10%, 20%, 30%, 40%をSVMで学習し、残りの画像で認識実験を行った。また、比較用に $k$ -NNと2次の多項式カーネルによるSVMの認識実験を行った。実験結果を図5-15に示す。なお、 $k$ -NNは $k=1,3,5$ を実験したが $k=1$ の結果が $k=3,5$ の結果を常に上回った。この1-NNよりもSVMの認識率は常に上であった。また、学習データが少ない場合、多項式カーネルを用いたSVMよりも線形SVMの方が良い認識率を示すが、学習データが増えるに従い認識率が入れ変わる結果となった。これは、ある程度サンプル数が確保できる場合、高次空間にデータを写像し分類する方式が有効であることを示している。

全画像の5%(400枚)を学習し、残り6,800枚を認識する処理速度は、線形のSVMが6分50秒、1-NNが35分9秒であった。使用した計算機はSGI-O2,R1200,270MHzである。

### 5.5.3 向きの認識特性

上記の考察では、形状不変物体を使い認識特性を調べたが、ここでは同様に形状不変物体を使い、向きの認識特性を調べた実験結果を示す。

#### (1) 向きの認識

図5-16に時計人形の向きを認識した実験結果例を示す。この実験は、人物の向き認識実験と同様に、時計人形を4方向(前後左右)から各100枚撮影し、SVMによる学習を行った。前項と同様、認識には正規化rg特徴とトップダウン型多クラスSVMを利用した。

新たに4方向から撮影した各400枚(合計1,600枚)を使い、向きの認識実験を行った。図5-16の各画像の左下ウインドウはSVMが判定した物体の向きを示している。SVMによる向きの認識率は97%であった。

#### (2) 向きの認識特性

向き認識の特性を調べるため、図5-16で利用した時計人形に似ている物体(「招き猫」)をCOILから選び、SVMにより向き認識実験を行った。COIL中の「招き猫」の画像数は72枚(=360度/5度)である。



図 5-14 データベース COIL の画像例[97]

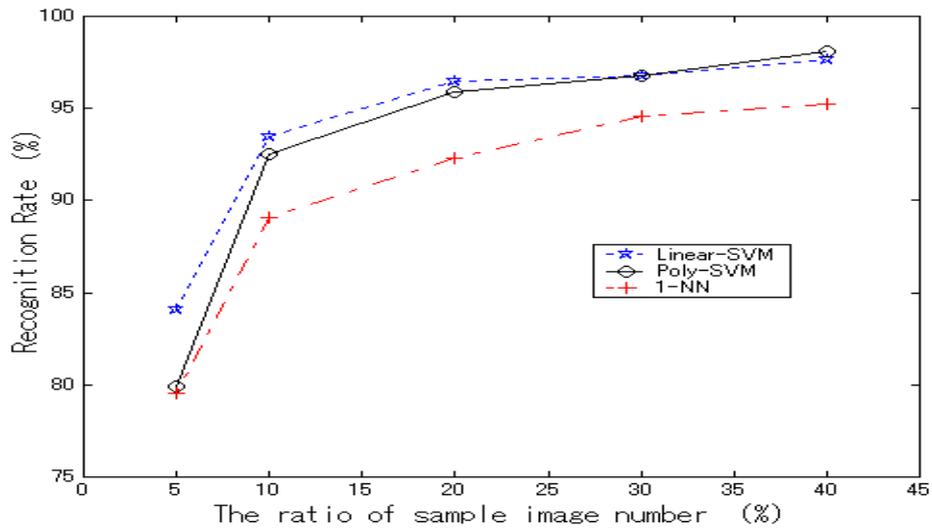
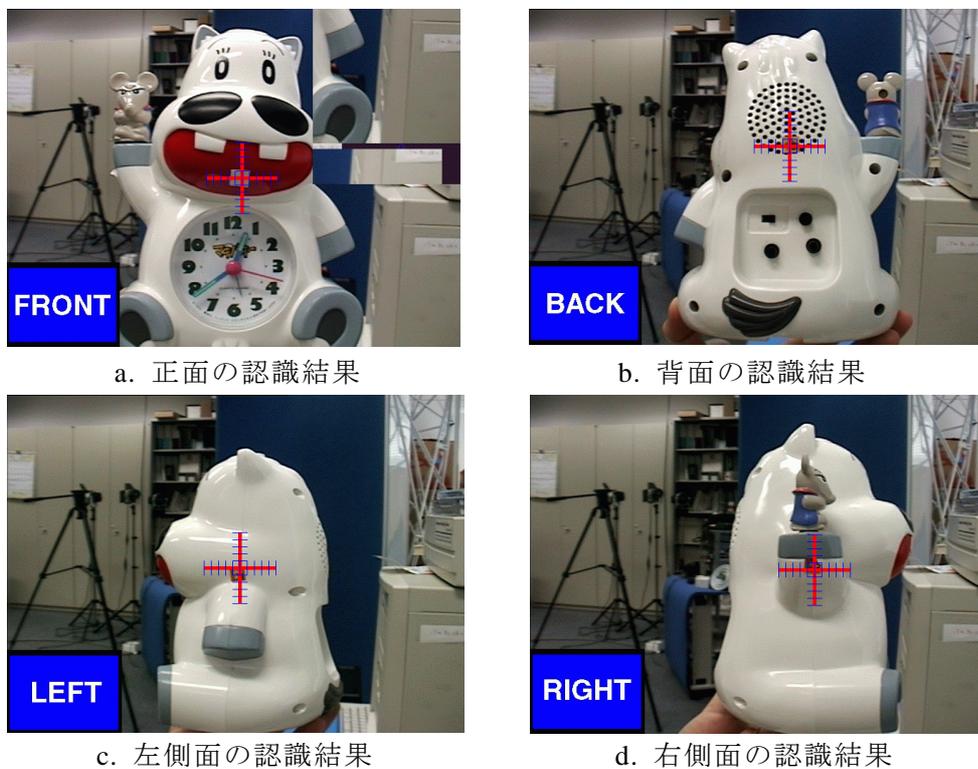


図 5-15 学習サンプル数と物体認識率 (使用画像: 7200 枚)



a. 正面の認識結果

b. 背面の認識結果

c. 左側面の認識結果

d. 右側面の認識結果

図 5-16 形状不変物体の向きの認識結果例

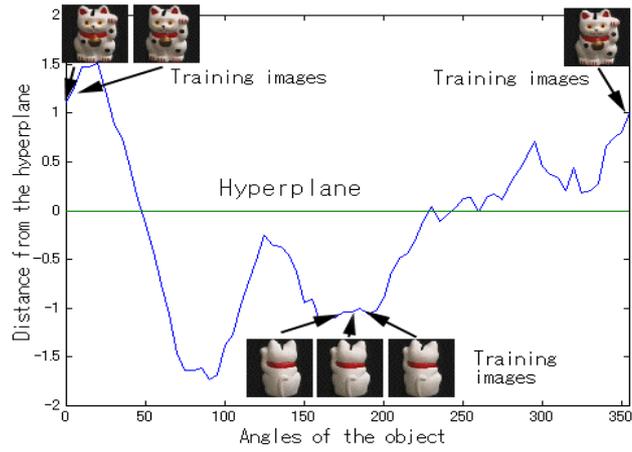


図 5-17 SVM による前面・背面の分類結果

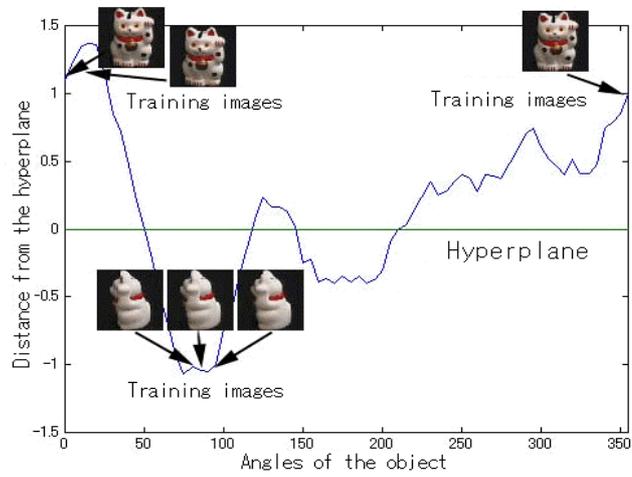


図 5-18 SVM による前面・左側面の分類結果

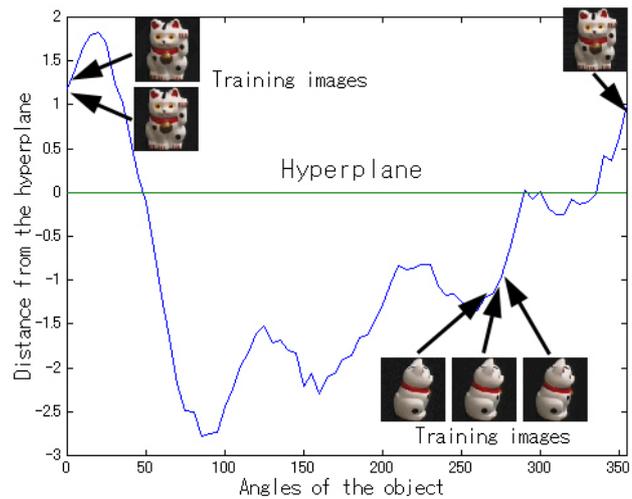


図 5-19 SVM による前面・右側面の分類結果

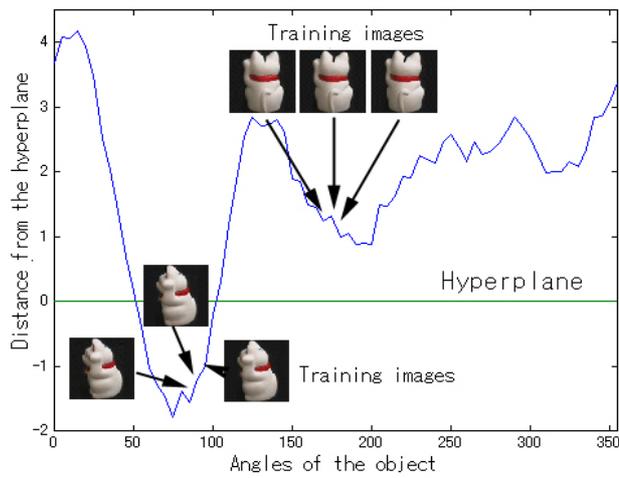


図 5-20 SVM による背面・左側面の分類結果

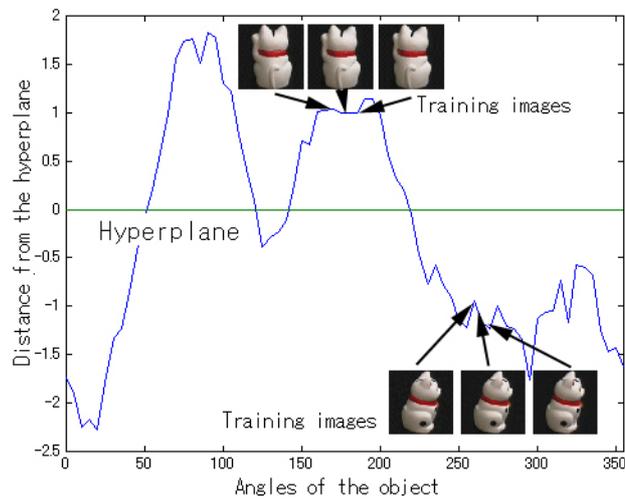


図 5-21 SVM による背面・右側面の分類結果

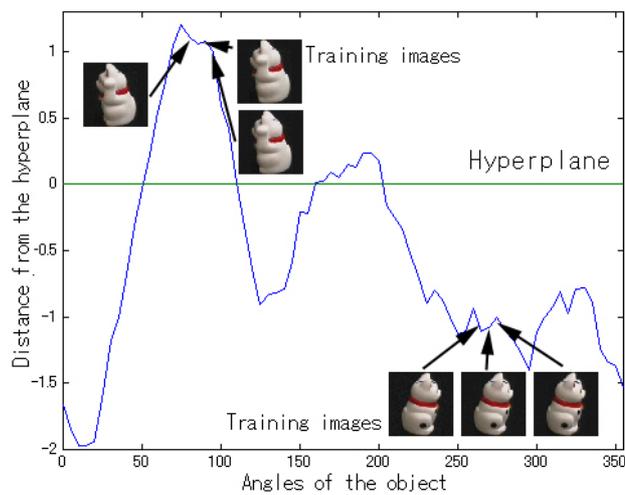


図 5-22 SVM による左側面・右側面の分類結果

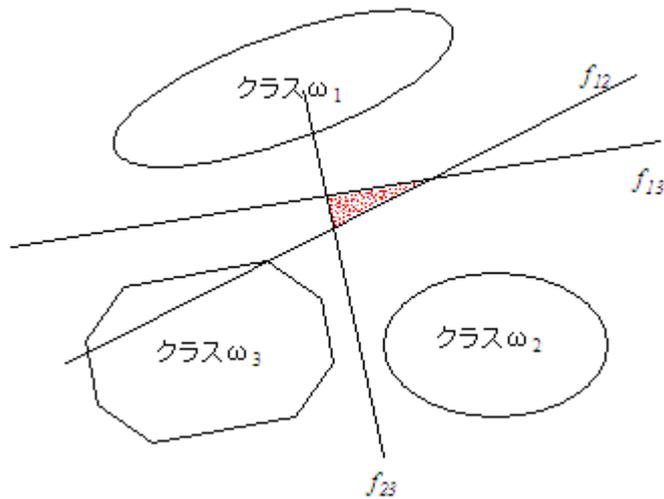


図 5-23 2 クラス比較の順番で結果が変わる領域

前後左右の各 3 枚の画像を SVM で学習させ、残りの角度を SVM がどのように判定するかを調べた結果が図 5-17～図 5-22 である。正面の 3 画像（355 度,0 度,5 度）と背面の 3 画像（175 度,180 度,185 度）を学習させ、残りの画像と各クラスの識別面までの距離を撮影角度ごとにプロットした結果が図 5-17 である。正面(355 度,0 度,5 度)と左側面(85 度,90 度,95 度)及び右側面(265 度,270 度,275 度)に対する実験結果が図 5-18 と図 5-19, 同様に背面と左右側面の特性を調べた結果が図 5-20 と図 5-21、左右側面を調べた結果が図 5-22 である。

SVM の認識では、図 5-17～図 5-22 の 0 ラインの上か下かで、どちらのクラスに画像が属するかを判定している。いずれの結果も、学習画像の近傍角度で正しく分類できる様子が分かる。特に、図 5-20 では左側面の認識範囲が最も狭く、その範囲が 55 度～100 度であることが分かる。

本実験により、人形の左右のわずかな違いを認識できることが分かった。人物の左右や前後の向きは、今回実験に利用した人形よりも明確な違いがあるため、5.4 節の人物の向き認識実験で 87% 程度の精度が得られたものと考えられる。

#### 5.5.4 提案方式の限界と対策

提案した人物認識方式での技術的な限界と、その限界に対する対策案を下記に示す。

##### (1) 認識の限界

- ・ 誤認識

物体認識に利用した SVM による 2 クラス対方式は、比較の順番により認識結果が変わる不安定さがある。例えば、図 5-23 に示すような 3 クラス  $\{\omega_1, \omega_2, \omega_3\}$  の各クラス対の識別関数を  $f_{12}(x)$ ,  $f_{23}(x)$ ,  $f_{13}(x)$  で示すと、網掛けをした 3 角形領域は識別関数の適用順番で結

果が異なる。例えば、網掛け部分を  $f_{12}(x), f_{13}(x)$  の順番で認識すると  $\omega_3$  となり、 $f_{23}(x), f_{12}(x)$  の順番で認識すると  $\omega_1$  となる。この問題に対し、各領域の重なりをなくす分類方式[133]が多数提案されているが、理論的な議論が続いている段階である。この問題以外にも、SVMによる機械学習は期待損失と経験損失のトレードオフが判断の基準となっているため、想定外の画像に対し誤認識や未認識が発生する可能性を含んでいる。

- ・新たな人物の認識と登録

本論文で利用している SVM は教示付き画像を機械学習している。そのため、すでに学習が済んでいる人物に対応できるが、新たな人物に対応できない問題がある。現状では、新たな人物が現れたかどうかの判定を目視で行い、定期的に学習を繰り返す必要がある。

## (2) 対策

- ・誤認識

認識に機械学習を利用しているため、誤認識の発生は避けられない。そのため、全身画像以外の特徴を機械学習した結果と組み合わせ、認識精度向上を図る必要がある。例えば、顔認識と組み合わせたり、歩き方による人物認識と組み合わせたりすることで、人物認識の精度向上を図る必要がある。システム化の観点から、既存の物理的なセンサと画像処理の組み合わせで誤認識を低減する方式も考えられる。例えば、社員証などに無線 IC タグを埋込み、特定箇所を通過した人物を無線 IC タグで認証し、通過時の画像が普段の全身画像と一致しない場合、社員証の持ち主を確認するように警備員へ警報を出す方式などが考えられる。

- ・新たな人物の区別と登録

現在のところ新たな人物の追加はマンマシーン・インタフェースを使い、監視員がシステムに教示する必要がある。この教示の自動化に向け、教師無し学習の活用、あるいは One-Class SVM [134] などにより各クラスから孤立している人物画像を見つける方法などを今後研究する必要がある。または、システム化の観点から、社員証に無線 IC タグなどの物理的なセンサを取り付け、社員ごとに画像を自動蓄積し、ある程度画像が収集できた段階で自動的に学習を行う方式なども考えられる。

## 5.6 本章のまとめ

監視カメラを利用した遠隔からの個人認証は、指紋・虹彩・静脈等の個人特有の特徴を利用することができないため非常に難しい問題である。この困難な問題に対し、本章では監視カメラ画像から検出した人物の全身画像を使い、既知人物を認識する方式を研究した。特に、研究室内の限られた環境のなかで、限られた人物を検出し認識する問題を検討した。

多人数を認識するため SVM の多段戦略を検討し、特定の 8 人を 16 日間撮影した監視画像を使い、監視カメラで一度撮影した同一日であれば約 88~98% の割合で個人を特定できたことを示した。さらに、人物の形状が常に変形するため、変形しない物体を撮影したコロンビア大学の画像データベース COIL100 などの画像を使い、検討方式の認識特性を定量的

に示した。また、SVMの多段戦略よりも、認識に利用する画像特徴に認識精度が大きく依存することを実験結果で示した。SVMの多段戦略の比較実験では、学習画像が少ない場合は1対他方式を利用すべきだが、学習画像が豊富な場合は学習時間がかからず追加学習が容易な2クラス対方式が実務面で有利であることを示した。

本章で示した実験は限定的なものだが、監視カメラで撮影した全身画像を使い人物を認識する機能の開発に貢献する成果を示している。今後、人物認識機能を持つ画像監視システムの開発に向け、認識精度向上を図る画像処理の基礎研究と、顔認識や各種センサとの組み合わせによるシステム化の2つの観点から、さらなる研究が不可欠である。

## 第6章

# 結 論

### 6.1 各章の総括と結論

電力施設などで設備の稼動状態の把握や、保守保全および警備の観点から、画像監視システムが数多く導入されている。しかし、これらの多くの監視カメラ画像を監視員が見続けることが不可能なため、目視を補助あるいは代行する画像処理が求められていることを1章で示した。特に、監視員が目視で行っている以下の機能を持つ画像処理の開発が望まれていることを示した。

- ・ 監視領域に入る人物を選択的に実時間で検出する機能
- ・ 検出人物の動作に応じて警報を出すための姿勢認識機能
- ・ 検出人物の個別認識機能

電気事業で利用している画像監視システムを例に、侵入者監視の自動化が最も要望の高い社会的ニーズであることを示した。また、監視対象になる頻度が高いがその全身画像による認識はほとんど着手されていないという技術的な理由で、画像処理による人物の検出と認識を対象として、上記3機能の実現方法を本論文で示した。特に、人物の検出や認識は、その形状などが不安定でパラメータ設定が困難なため、機械学習が不可欠である。ただし、機械学習を適応しただけでは実時間処理や安定した認識ができない。そこで、脳の視覚処理を参考に、動きによる人物領域の絞り込みを先に行い、その後SVMの機械学習を用いた人物の検出・認識を行う段階的な処理で、移動中の人物を検出し認識する画像処理の実現方式を示した。

各課題の検討用に監視環境と監視対象を限定した評価可能な実問題を設定し、実験結果により提案方式の有効性を示した。なお、本論文では移動体として最も一般的な監視対象である人物を想定したが、画像監視システムに与える学習画像を変更することで、人物以外の移動体にも拡張できる汎用的な方式となっている。

本章では6.1節で各章の総括を示した後、本研究で参考にした視覚処理と提案方式の関連を示す。6.2節で残された今後の課題と画像監視システムの将来像を示し、本論文を締めくくる。

#### 6.1.1 画像監視システムの必要性

1章では、電気事業等で利用している画像監視システムの構成と機能を示し、画像監視システムの中で自動化できていない部分が、監視員による目視監視であることを示した。画像処理の歴史的な考察から、監視員による目視を画像処理で代行することはシーン理解の困難な課題に相当し、現在も解決に向けた研究が続いていることを示した。さらに、監視員が目視で行っている主な処理が、監視領域に入る移動体の検出・認識・追跡であるこ

とを示した。その中でも中心的な処理である検出と認識用に提案されている従来の画像処理の概要とその技術課題を示した。

電気事業で利用している画像監視システムを具体例として取上げ、最も要望の高い社会的ニーズが侵入者監視であることを示した。さらに、画像監視システムによる監視現場を、屋内、電気所構内、電気所構外に分類し、画像処理による侵入者監視の難しさに差があることを示した。本論文では、その中でも屋内や電気所構内の画像監視システムに対し、社会的ニーズが高い人物を対象に、目視監視を補助あるいは代行する3機能（①人物の実時間検出、②検出人物の全身画像による姿勢認識、③検出人物の全身画像による人物認識）の実現を目指すことを示した。

人物の検出と認識は、人物形状などが不安定でパラメータ設定が困難なため、機械学習が不可欠である。そこで、機械学習として2分類問題用の識別器であるSVMを使い、監視領域に入る人物を検出し認識する方針を示した。ただし、機械学習を適用しただけでは実時間処理や安定した認識ができないため、脳の視覚処理を参考に、動き解析と機械学習を組み合わせたアプローチを採用することを示した。具体的には、動きによる人物領域の絞り込みを先に行い、その後、SVMによる人物の検出・認識を行う段階的な処理で、人物の実時間検出と認識を可能にする統一的な枠組みである。

### 6.1.2 関連する画像処理研究

2章では画像監視システムに活用できる従来の画像処理方式に関して、人物の検出と認識の観点から技術概要とその課題を示した。また、検出と認識の両方に利用できる汎用的な機械学習を解説した。さらに、現在の画像処理よりも汎用的な監視機能を実現している脳の視覚野の構造と機能を概説した。

人物の検出では機械学習による汎用的な方式が提案されているが、現在の市販パソコンを使い実時間で動作しない課題があることを示した。検出した人物の動作を把握するための姿勢認識では身体モデルを当てはめる研究が進んでいるが、人物の様々な姿勢を定義することが困難なため、人物検出と同様の機械学習の枠組みで姿勢認識を行う研究成果が少ないことを示した。さらに、人物認識では、顔や歩き方を利用した方式が提案されているが、全身画像を利用した人物認識が行われていないことを示した。

人物の検出や認識に既存の機械学習が共通した技術として使えるため、教師付き機械学習の概要を説明し、2分類問題用のSVMやAdaBoostなどを解説した。また、人の脳細胞の処理速度は現在市販されている計算機よりも遅いが、100億個以上も存在するといわれている脳細胞の構造化と機能分化により、既存の画像処理による人物検出よりも高速な処理を実現しているため、画像処理の設計に参考となる視覚野の構造と機能を示した。特に、脳では動きによる移動体領域の絞り込みを先に行い、その後、移動体領域の認識を行っている可能性が高いことを示し、動きによる絞り込みと、その絞り込んだ領域の認識処理を行うことで、画像処理の高速化と高精度化が図れる可能性があることを示した。

### 6.1.3 人物の実時間検出

SVM を使い監視画像から人物を検出する方式は既に提案されている。しかし、従来の

SVMによる人物検出方式は画像内から人物を選択的に検出できるものの、実時間(30 画像/秒)で処理ができなという弱点があった。そこで3章では、視覚処理を参考に、動きによる人物領域の絞込みの後に SVM の人物検出を使い、実時間で人物を検出する方式を考案した。動きによる人物領域の絞込み処理として、固定カメラ用にライン型前処理を、巡回カメラ用に側抑制型前処理を考案した。

ライン型前処理は、監視領域内に監視ラインを設定し、監視ライン上で動きの速い部分に優先順位をつけ、優先順位の上位を機械学習の探索範囲とする方式である。一度人物ではないと判定した部分は、一定時間のあいだ人物検出の探索範囲から除外する方式とし、処理の高速化を図った。一方、側抑制型前処理は、移動中の人物が周囲の背景と異なる動きをすることに着目し、周囲と異なる動き部分を機械学習の探索範囲とする方式である。この処理は視覚野での動き処理経路（特に MT 野）を参考に、周囲と同じ方向に動く部分の選択を抑え、周囲と異なる方向に動く部分を優先して選択する方式としたものである。

評価実験では、考案した前処理と従来の機械学習による人物検出方式を使い、実時間で人物検出ができることを示した。従来の機械学習による人物検出は、単独で実時間処理ができないが、考案方式はその障害を解決している。そのため、考案方式は画像監視システムの監視機能向上に直接寄与する成果となっている。

#### 6.1.4 検出人物の姿勢認識

画像内の人物の動作を把握するためには、検出した人物の姿勢を認識する必要がある。全身画像に身体モデルを当てはめる研究成果が数多く報告されているが、様々な姿勢を定義することが難しいため3章の人物検出と同様な枠組みで姿勢認識を行う画像監視システムは存在していない。そこで、4章では検出した全身画像を用い機械学習で姿勢を認識する方式を考案した。

人物は様々な姿勢をするため姿勢の認識精度を議論することが難しいが、発電所運転操作のような基本姿勢が定義できる監視環境であれば、姿勢認識精度を評価できる。そこで、発電所の運転訓練を撮影した画像を使い、姿勢を認識する方式を考案した。動きと操作姿勢の間に相関があるため、考案方式は発電所運転特有の4種類の姿勢を SVM で学習し、人物の動きに合わせて4姿勢の学習結果を動的に切替て利用する動的切替方式とした。

評価実験では、監視カメラで撮影した1時間30分の画像で95%の姿勢認識率が得られ、考案方式により人物の向きや姿勢のばらつきなどを吸収した姿勢認識ができることを示した。この姿勢認識方式は、通常の運転から逸脱した運転操作時などに警報を出す画像監視システムの作成を可能にする成果であり、人物の動作に応じた警報を出す画像監視システムの実現に寄与する成果である。

#### 6.1.5 検出人物の認識

画像処理による顔、指紋、虹彩、静脈などを利用した個人認証に比べ、監視画像から検出した人物は形状や服装がたえず変化し、撮影する高さや角度により見え方が大きく異なるため、人物を特定する安定した特徴を得にくい。そのため、監視員が目視で行っている人物認識を、従来の画像処理方式で実現することが難しく、近年、歩き方で人物を認識する

研究が始まった段階である。しかし未だ、全身画像を使い SVM などの機械学習を用いた人物認識研究は行われていない。そこで、5章では、3章や4章と同様の枠組みで、監視カメラで検出した全身画像から個人を認識する問題を議論した。

具体的な問題として、研究室内のコーヒーマシンを利用する研究室メンバーの認識問題を設定し、監視カメラ画像から検出した全身画像を使い個人を特定する方式を機械学習で実現した。ただし、SVM は2分類問題用の認識方式なので、多人数を認識するには SVM を組み合わせて利用する必要がある。そこで、本論文では1対他方式と2クラス対方式による SVM の組み合わせ戦略を、複数の画像特徴を使い比較検討した。

評価実験では研究室内に出入する既知人物8人の16日間の全身画像(約1,120画像)を使い、SVM による人物認識結果を示し、同一撮影日の全身画像を学習すると95%~98%程度の認識率が得られたことを示した。また、100物体が写る7,200画像による認識実験を行い、より定量的な検証結果を示した。さらに、識別人数が多く学習画像が豊富な場合、SVM による多人数の認識には、学習時間がかからず追加学習が容易な2クラス対方式による SVM の多段戦略が実務面で有利であることを示した。本実験結果は限定的なものだが、監視カメラで撮影した全身画像から、一度撮影した人物を個別に特定する機能の開発に貢献する成果である。

### 6.1.6 視覚処理との関連

脳の視覚処理は、白黒画像で高速に動きを処理する経路と、カラー画像で物体認識を行う経路に分かれている。動き処理経路は認識処理経路よりも処理が速く、2経路の主要な階層で接続がある。例えば、動き処理経路で周囲と異なる動き領域を検出する領野(MT野)と、認識処理経路で物体認識に繋がる領野(V4)の間で接続がある。本論文では目視機能を補助あるいは代行する画像処理の実現に、このような視覚野の2経路の処理とその接続を参考とした。

3章から5章で示した人物検出方式、姿勢認識方式、個人認識方式は、いずれも画像内の動きにより人物領域を絞り込む前処理と、絞り込んだ領域を機械学習で認識する後処理の2段階処理で実現している。ただし、各考案方式で前処理と後処理のウェイトが異なる。図6-1に各処理での前処理と後処理のウェイトの違いを概念図で示す。人物検出では、実時間の動作を確保するため動き処理を優先している。姿勢認識では人物の動きにより機械学習を切替えているため、両処理の役割は同程度である。人物の個人認識では、機械学習を多段に組み合わせて多人数を認識する必要があり、認識処理に大きなウェイトがか

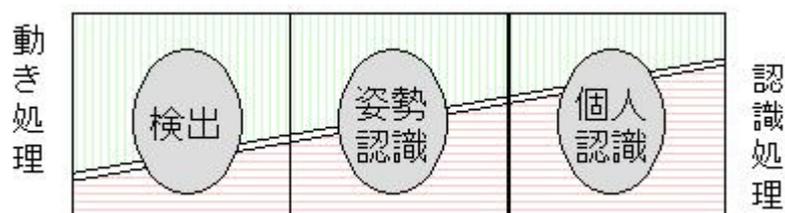


図 6-1 動き処理と認識処理の比率

かっている。

図6-1に示すように処理のウェイトは異なるが、動きによる前処理と機械学習による後処理で高速な人物検出、姿勢認識、個人認識を行う画像処理が開発できることを本論文は示している。つまり、視覚野の動き処理経路でのMT野と認識処理経路のV4野の接続を参考にした処理で、従来困難であった画像処理を実現したことになる。

## 6.2 今後の課題と展望

本論文では画像監視システムのための画像処理として、①人物の実時間検出、②姿勢認識、③人物認識を、動き処理による絞込みと機械学習による検出/認識の組み合わせで実現する方式を示した。画像監視システムの中で唯一自動化できていない目視監視を補助あるいは代行する画像処理の実現に向け、本論文は一つの方向性を示す成果であると考えている。特に、人物の実時間検出は、既設の画像監視システムに直接導入できる成果である。姿勢認識は、限定的な課題でその有効性を示したが、今後、提案方式を個別の実問題に適用しながらより汎用的な方式を目指す必要がある。人物認識は、個人を認識できる可能性を示した初期段階であり、今後、人物の自動登録方法などを含めた基礎研究を行う必要がある。本節では残された課題と今後の展望を示す。

### 6.2.1 今後の課題

#### (1) 人物の検出

機械学習では事前に学習できない人物画像や背景画像に対し、誤検出や未検出を起こす可能性がある。そのような問題に対し、監視員がインタフェースを使いインタラクティブに教示を行い、再学習を繰り返す必要がある。ただし、教示を繰り返しても誤報と侵入者の見逃しを完全に無くすことは現時点で理論的に不可能である。そのため、画像監視システムの作成には、誤報低減と侵入者見逃しのトレードオフを常に念頭においたシステム設計が必要である。

例えば、原子力発電所などでは、誤報の低減よりも侵入者見逃しの低減に力を入れるべきであり、赤外センサが変化を検知した場合、画像内変化がわずかであっても警備員に警報を出す必要がある。一方、社員しか通ることがなく大事故にいたらない設備であれば、誤報低減を目指した設計を行う必要がある。

画像監視システムは24時間動作することが多く、夜間の照明、霧、雨、雪などで誤報や侵入者の見逃しを起こす可能性がある。例えば、霧などで画像内の人物が見えない場合、画像処理が不可能となる。そのような場合、霧の中でも使えるミリ波センサなどを主とする侵入者監視に切替る必要がある。そのため、画像監視システムの設計にあたっては、画像処理をシステムから切り離す基準などの検討も今後必要である。

#### (2) 姿勢認識

本論文で述べた姿勢認識は、画像内の動き処理と機械学習で代表的な姿勢が認識できる

ことを示した段階である。室内や電気所構内を監視する一般的な画像監視システムでは、動作から不審者かどうかを判断する姿勢認識があればよく、提案方式で対応できると考えられる。しかし、より詳細な姿勢認識を必要とする監視には、提案方式で代表的な姿勢を認識した後に、従来の身体モデルを人物に当てはめる処理が必要である。

また、提案方式は認識に機械学習を利用しているため、誤認識が発生する可能性がある。誤認識に対して、シーン中での前後の姿勢認識結果の整合性を検証し、誤認識を訂正する機能の開発が今後必要である。

### (3) 個人認識

上記と同様、認識に機械学習を利用しているため、誤認識の発生が理論的に避けられない。今後、システム化の観点から誤認識を少しでも少なくする方法を検討する必要がある。例えば、監視カメラの画像が顔認識可能な解像度であれば顔認識と組み合わせる、あるいは歩き方による人物認識と組み合わせるなど、他の人物特徴を利用した認識方式との組み合わせにより精度向上を図る必要がある。また、物理的なセンサとの組み合わせで、精度向上を図ることも必要である。例えば、社員証に無線 IC タグを埋め込み、特定箇所を通過した人物を無線 IC タグで確認し、その後、提案方式で個人認識を行い、問題があれば警備員へ確認を促す 2 重チェックの個人認証などが考えられる。

現在のところ提案方式は、新たな人物の追加学習にマンマシーン・インタフェースを使い、監視員がシステムに画像を教示する必要がある。今後、教師無し学習などの導入により、新たな人物の自動登録方法を研究する必要がある。

### (4) 視覚処理

永年に渡る画像処理の研究成果により、対象を限定することで目視を凌駕する画像処理が実現できている。しかし、画像監視システムのように一般的なシーン理解が必要な課題では、現在の計算機をもってしても目視と同等の画像処理が実現できていない。今後、画像監視システム用の画像処理の高度化に向け、視覚野の構造と機能を参考にできる部分が多々あるものと考えられる。逆に、画像処理のシステム化の観点から、生理学的に解明が難しい個別視覚野の結合の役割等を明確化できる可能性がある。

その意味では、本論文は視覚野の動き処理経路の **MT** 野と認識処理経路の **V4** 野の接続関係を参考に、動き処理経路から認識処理経路へ信号を伝達することで、工学的に役立つ画像処理が実現できることを示している。今後、**V4** 野から **MT** 野への逆方向の情報伝達の意味づけと工学的活用方法をシステム化の観点から検討する必要がある。また、2 章の関連研究(2.6.4 節)で示したが、人は無意識にサッケードと呼ばれる眼球運動を行っており、画像内の重要な部分に目を向けている。このような無意識の選択的注意は、人物の全身画像の学習等を必要としないため、機械学習の未認識や誤認識の課題を補う処理としての活用が期待できる。今後、このような選択的注意などの機能を画像処理で実現できれば、目視監視を代行する画像処理の実現に寄与するだけでなく、生理学的に解明が難しい個別視覚野の結合の役割を明確化することに寄与できる可能性がある。

## 6.2.2 監視システムの今後

本論文では、画像監視システム内で唯一自動化できていない部分が目視監視であることを示し、目視監視を補助あるいは代行する画像処理の実現方法を研究した。

目視に代わる画像処理の実現は、画像処理の黎明期から議論されてきたシーン理解の難しい課題であり、今後もさらなる基礎研究が不可欠である。一方、近年、無線 IC タグなどのデバイスや様々なセンサが急速に普及し始めている。そのため、システム化の観点から、画像処理結果と各種センサの反応を同列のデータとして扱い、変化を機械学習し侵入者を検知する研究などが進むものと予想できる。今後、従来の赤外センサに加えミリ波センサや無線 IC タグなどの各種センサと監視カメラの画像を同列に扱う監視システムの研究が進めば、目視を上回る監視が実現できる可能性がある。その際には、画像処理も大きな監視システムの構成部品に過ぎなくなるが、構成部品が信頼性の高い情報を提供するニーズは今後も変わらないであろう。

本論文では、電気事業用の大規模な画像監視システムを対象に議論したが、考案した画像処理は家庭内のホームセキュリティなどにも適した技術である。家庭内に設置するカメラであれば、顔や個々人の体型（体形）なども活用でき、信頼性の高い人物認識を実現できる可能性がある。このような技術は、高齢化と核家族化が進んでいる国内で、遠隔地にいる家族の安全を守る画像監視システムとしても活用が見込まれる。

## 参考文献

1. 下中直也編集: 世界大百科事典, 平凡社, 1988.
2. 高木幹雄, 下田陽久監修: 新編画像解析ハンドブック, 東京大学出版, 2004.
3. 情報処理学会編: 情報処理ハンドブック, 情報処理学会, 1995.
4. 長尾真: 画像認識論, コロナ社, 1983.
5. C.R. Brice, and C.L. Fennema: Scene Analysis Using Regions, Artificial Intelligence, Vol.1, No.3, pp.205-226, 1970.
6. T. Kanade: Picture Processing by Computer Complex and Recognition of Human Faces, Technical Report, Kyoto University, Department of Information Science, 1973.
7. 工業技術院監修: 画像処理サブルーチンパッケージ SPIDER Users' Manual, 情報技術コンソーシアム, 1982.
8. 上野晴樹: 自律型人型サービスロボットの研究開発, 日本ロボット工業会機関誌「ロボット」, No.109, 1996.
9. P.J. Phillips, P.J. Rauss, and S.Z. Der: FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results, Army Research Lab Technical Report 995, 1996.
10. 高橋学: 監視カメラ画像処理技術の展開, 第 11 回画像センシングシンポジウム論文集, pp.35-40, 2005.
11. 北海道新聞: 増殖する眼, 犯罪抑止効果は不明確, 4 月 20 日, 2005.
12. 国土交通省: 災害情報河川のリアルタイム画像, <http://www.mlit.go.jp/river/movie.html>
13. 佐藤真一: 大規模映像アーカイブに基づく映像キーワードの抽出と応用に関する研究, 科学研究補助金研究成果報告書, 課題番号 15500082, 2005.
14. D. Marr: VISION, W.H. Freeman and Company, 1982.
15. T. Poggio, V. Torre, and C. Koch: Computational Vision and Regularization Theory, Nature, 317, pp.314-319, 1985.
16. B.K.P. Horn: Robot Vision, The MIT Press, 1986.
17. S. Zeki: A Vision of the Brain, Blackwell Scientific Publications, 1993.
18. J.G. Nicholls, A.R. Martin, and B.G. Wallace: From Neuron to Brain, Sinauer Associates, 1992.
19. B.A. Wandell: Foundations of Vision, Sinauer Associates, 1995.
20. 川人光男: 脳の計算理論, 産業図書, 1996.
21. 読売新聞: 安全大国の幻想, 10 月 26 日, 2005.
22. S. Huwer, and H. Niemann: Adaptive Change Detection for Real-Time Surveillance Applications, Proceedings of the IEEE International workshop on Visual Surveillance, pp. 37-46, 2000.
23. B. Kim: Multi-Source Human Identification, Massachusetts Institute of Technology Master Thesis, 2003.
24. T. Horprasert, D. Harwood, and L.S. Davis: A Statistical Approach for Real-time Robust

- Background Subtraction and Shadow Detection, Proceedings of the IEEE International Conference on Computer Vision FRAME-RATE Workshop, 1999.
25. 羽下哲司, 藤原秀人, 鷺見和彦: 首振り,ズームカメラを用いたトラッキング型侵入監視システム, 電子情報通信学会技術報告, PRMU99-67, pp.23-30, 1999.
  26. 藤吉弘亙, 小村剛史, 矢入江口郁子, 香山健太郎, 吉水宏: 歩行者 ITS のためのフレーム間差分による移動体検出法とその評価, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol.45, No.SIG13(CVIM 10), pp.11-20, 2004.
  27. 高藤政雄, 北村忠明, 小林芳樹: 空間微分および差分処理を用いた車両抽出法, 電子情報通信学会論文誌, Vol.J80-D-II, No.11, pp.2976-2985, 1997.
  28. A. Elgammal, D. Harwood, and L.S. Davis: Nonparametric Model for Background Subtraction, Proceedings of the European Conference on Computer Vision, pp.751-767, 2000.
  29. 鈴木美彦: 画像処理システム,画像処理方法,画像処理プログラム, 特許出願番号:特願2002-160281, 特開 2004-7246, 2004.
  30. C. Stauffer, and W.E.L. Grimson: Adaptive Background Mixture Models for Real-time Tracking, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.246-252, 1999.
  31. 大津展之: 判別および最小2乗基準に基づく自動しきい値選定法, 電子情報通信学会論文誌, Vol.J63-D-II, No.4, pp.349-356, 1980.
  32. 納由美子: 2値化レベル決定装置, 特許出願番号:特願平3-74263, 特開平4-286093, 1992.
  33. G. L. Foresti: A Real-Time System for Video Surveillance on Unattended Outdoor Environments, IEEE Transaction on Circuits and Systems for Video Technology, Vol.8, No.6, pp.697-704, 1998.
  34. N. M. Oliver, B. Rosario, and A. P. Pentland: A Bayesian Computer Vision System for Modeling Human Interactions, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.831-843, 2000.
  35. 長屋茂喜, 宮武孝文, 藤田武洋, 伊藤渡, 上田博唯: 時間相関型背景判定法による移動物体検出, 電子情報通信学会論文誌, Vol.J79-D-II, No.4, pp.568-576, 1996.
  36. 松山隆司, 和田俊和,波部斉,棚橋和也: 照明変化に頑健な背景差分, 電子情報通信学会論文誌, Vol. J84-D-II, No.10, pp.2201-2211, 2001.
  37. 佐藤雄隆, 金子俊一, 五十嵐悟: 周辺増分符号相関画像に基づくロバスト物体検出及び分離, 電子情報通信学会論文誌, Vol. J84-D-II, No.12, pp.2585-2594, 2001.
  38. 佐伯規夫, 鈴木健志, 岩田清: 画像認識による侵入者検出方式, FUJITSU General Technical Journal, No.1, 1999.
  39. 嶺直道, 八木康史, 谷内田正彦: 空間微分画像と差分画像の結合による変化領域検出法, 電子情報通信学会論文誌, Vol. J77-D-II, No.3, pp.631-634, 1994.
  40. 倉爪亮, 由井俊太郎, 辻徳生, 岩下友美, 原健二, 長谷川勉: Fast Level Set Method の提案とビデオ画像の移動物体のリアルタイム追跡, 情報処理学会論文誌, Vol.44, No.8, pp.2244-2254, 2003.
  41. 関真規人, 和田俊和, 藤原秀人, 鷺見和彦: 背景変化の共起性に基づく背景差分, 情報

- 処理学会論文誌コンピュータビジョンとイメージメディア, Vol.44, No. SIG 5 (CVIM6), pp.54-63, 2003.
42. 羽下哲司, 鷺見和彦, 八木康史: 変化領域内の動きの時空間特徴に着目した屋外情景における歩行者の検出, 電子情報通信学会論文誌, Vol.J87-DII, No.5, pp.1104-1111, 2004.
  43. A. Talukder, S. Goldberg, L. Matthies, and A. Ansar: Real-time Detection of Moving Objects in a Dynamic Scene from Moving Robotic Vehicles, Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vol.2, pp.1308-1313, 2003.
  44. J.Y.A. Wang, and E.H. Adelson: Representing Moving Images with Layers, IEEE Transactions on Image Processing, Vol.3, No.5, pp.625-638, 1994.
  45. A. Bab-Hadiashar, and D. Suter: Motion Segmentation Using Robust Statistics and Spatial Continuity, Proceedings of the International Workshop on Image Analysis and Information Fusion, pp.271-280, 1997.
  46. M. Irani, and P. Anandan: A Unified Approach to Moving Object Detection in 2D and 3D Scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20, No.6, pp. 577-589, 1998.
  47. M. Irani, B. Rousso, and S. Peleg: Computing Occluding and Transparent Motions, International Journal of Computer Vision, Vol.12, No.1, pp.5-16, 1994.
  48. 白井良明, 三浦純: 複雑背景における人の追跡, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol.43, No.SIG4, pp.33-42, 2002.
  49. S. Fejes, and L.S. Davis: Detection of Independent Motion Using Directional Motion Estimation, Computer Vision and Image Understanding, Vol.74, No.2, pp.101-120, 1999.
  50. W.B. Thompson: Combining Motion and Contrast for Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.2, No.6, pp.543-549, 1980.
  51. Y. Altunbasak, E.P. Eren, and M.A. Tekalp: Region-Based Parametric Motion Segmentation Using Color Information, Graphical Models and Image Processing, Vol.60, No.1, pp.13-23, 1998.
  52. R. Fablet, P. Bouyhemy, and M. Gelgon: Moving Object Detection in Color Image Sequences Using Region-level Graph Labeling, Proceedings of the IEEE International Conference on Image Processing, Vol.2, pp.939-943, 1999.
  53. M. Gelgon, and P. Bouyhemy: A Region-level Graph Labeling Approach to Motion-based Segmentation, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp.514-519, 1997.
  54. 上條俊介, 坂内正夫: 時空間 MRF モデルに基づく車両と歩行者の統一的トラッキング, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol.45, No.SIG13 (CVIM10), pp.54-63, 2004.
  55. D.M. Gavrilu: The Visual Analysis of Human Movement: A Survey, Computer Vision and Image Understanding, Vol.73, No.1, pp.82-98, 1999.
  56. S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, IEEE Transactions on Pattern

- Analysis and Machine Intelligence, Vol.27, No.2, pp.162-177, 2005.
57. V.N. Vapnik: Statistical Learning Theory, John Wiley and Sons Inc., New York, 1998.
  58. M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio: Pedestrian Detection Using Wavelet Templates, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.193-199, 1997.
  59. A. Mohan, C. Papageorgiou, and T. Poggio: Example-based Object Detection in Images by Components, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No.4, pp.349-361, 2001.
  60. C. Papageorgiou, M. Oren, and T. Poggio: A General Framework for Object Detection, Proceedings of the IEEE International Conference on Computer Vision, pp.555-562, 1998.
  61. P. Viola, M.J. Jones, and D. Snow: Detecting Pedestrians Using Patterns of Motion and Appearance, Proceedings of the IEEE International Conference on Computer Vision, Vol.2, pp.734-741, 2003.
  62. P. Viola and M. J. Jones: Rapid Object Detection using a Boosted Cascade of Simple Features, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.511-518, 2001.
  63. R.T. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa: A System for Video Surveillance and Monitoring, Technical Report: CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.
  64. C. Stauffer, and W.E.L. Grimson: Learning Patterns of Activity using Real-time Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.747-757, 2000.
  65. I. Haritaoglu, D. Harwood, and L.S. Davis:  $W^4$ : Real-Time Surveillance of People and Their Activities, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.809-830, 2000.
  66. H. Sidenbladh, M.J. Black and D.J. Fleet: Stochastic Tracking of 3D Human Figures Using 2D Image Motion, Proceedings of the European Conference on Computer Vision, pp.702-718, 2000.
  67. 山本正信: コンピュータビジョン技術による次世代モーションキャプチャ, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol.43, No. SIG4 (CVIM4), pp. 1-9, 2002.
  68. 伊藤貴志, 山本正信: 原子カプラント運転作業の動画像解析と VR 空間での再現, 新潟大学工学部情報工学科卒業論文, 1999.
  69. 中島慶人: 原子力発電所運転訓練時の訓練員の検出と姿勢認識, 電気学会論文誌, Vol.125-D, No.1, pp.60-66, 2005.
  70. C. Nakajima, M. Pontil, B. Heisele, and T. Poggio: Full-body Person Recognition System, Pattern Recognition, Vol.36, No.9, pp.1997-2006, 2003.
  71. 浅田稔: ダイナミックシーンの理解, 電子情報通信学会, 1994.
  72. N. Ohta: Image Movement Detection with Reliability Indices, IEICE Transactions on

- Information and Systems, Vol.E74, No.10, pp.3379-3388, 1991.
73. C. Nakajima: Extraction of Salient Apexes from an Image by Using The Function at the Primary Visual Cortex, Proceedings of the International Conference on Pattern Recognition, Vol.1, pp.720-724, 1998.
  74. S.T. Barnard, and W.B. Thompson: Disparity Analysis of Image, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.2, No.4, pp.333-340, 1980.
  75. S.M. Haynes, and R. Jain: Detection of Moving Edges, Computer Vision, Graphics and Image Processing, Vol.21, No.3, pp.345-367 1983.
  76. N. Ancona, and T. Poggio: Optical Flow from 1D Correlation: Application to a Simple Time-to-Crash Detector, Massachusetts Institute of Technology A.I. Memo, No.1375, 1993.
  77. J. Davis: Recognizing Movement using Motion Histograms, Massachusetts Institute of Technology Media Laboratory Perceptual Computing Section Technical Report, No.487, 1998.
  78. J. Davis: Hierarchical Motion History Images for Recognizing Human Motion, Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video, pp.39-46, 2001.
  79. B. Horn, and B. Schunck: Determining Optical Flow, Artificial Intelligence, Vol.17, pp. 185-203. 1981.
  80. B. Lucas, and T. Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision, Proceedings of the International Joint Conference on Artificial Intelligence, pp. 674-679. 1981.
  81. G. Bradski, and J. Davis: Motion Segmentation and Pose Recognition with Motion History Gradients, Machine Vision and Applications, Vol.13, pp.174-184 2002.
  82. B. Heisele, U. Kressel, and W. Ritter: Tracking non-rigid, Moving Objects Based on Color Cluster Flow Segmentation of Range and Intensity Image Sequences, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.257-260, 1997.
  83. R. Lienhart, and J. Maydt: An Extended Set of Haar-like Features for Rapid Object Detection, Proceedings of the IEEE International Conference on Image Processing, pp.900-903, 2002.
  84. C.R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland: Pfunder: Real-Time Tracking of the Human Body, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp.780-785, 1997.
  85. R.O. Duda, and P.E. Hart: Pattern Classification and Scene Analysis, Wiley, New York, 1973.
  86. 甘利俊一, 麻生英樹, 津田浩治, 村田昇: パターン認識と学習の統計学, 岩波書店, 2003.
  87. 鳥脇純一郎: 認識工学 パターン認識とその応用, コロナ社, 1993.
  88. M. Pontil, and A. Verri: Support Vector Machines for 3-D Object Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20, No.6, pp.637-646, 1998.
  89. B. Schölkopf, A. Smola, R. Williamson, and P. L. Bartlett: New Support Vector Algorithms, Neural Computation, Vol.12, pp.1207-1245, 2000.
  90. H. Saito: Hierarchical Neural Analysis of Optical Flow in the Macaque Visual Pathway, Brain

Mechanisms of Perception and Memory, Oxford University Press, pp.121-139, 1993.

91. C. Papageorgiou, and T. Poggio: A Trainable Object Detection System: Car Detection in Static Images, Massachusetts Institute of Technology AI Memo 1673, 1999.
92. M.S. Nixon, J.N. Carter, J.M. Nash, P.S. Huang, D. Cunado, and S.V. Stevenage: Automatic Gait Recognition, Proceedings of the IEE Colloquium on Motion Analysis and Tracking, pp. 1-6, 1999.
93. M.S. Nixon, J.N. Carter, M.G. Grant, L.G. Gordon, and J.B. Hayfron-Acquah: Automatic Recognition by Gait: Progress and Prospects, Sensor Review, Vol.23, No.4, pp.323-331, 2003.
94. J.C. Platt, N. Cristianini, and J. Shawe-Taylor: Large Margin Dags for Multiclass Classification, Advances in Neural Information Processing Systems, pp.547-553, 2000.
95. O. Chapelle, P. Haffner, and V. Vapnik: Support Vector Machines for Histogram-based Image Classification, IEEE Transactions on Neural Networks, Vol.10, No.5, pp.1055-1064, 1999.
96. A. Elisseeff, Y. Guermeur, and H. Paugam-Moisy: Margin Error and Generalization Capabilities of Multiclass Discriminant Systems, Technical Report NC2-TR-1999-051, NeuroCOLT2 Technical Report, 1999.
97. S.A. Nene, S.K. Nayar, and H. Murase: Columbia Object Image Library COIL-100, Department of Computer Science, Columbia University Technical Report CUCS-006-96, 1996.
98. C. Nakajima, M. Pontil, B. Heisele and T. Poggio: People Recognition in Image Sequences by Supervised Learning, Massachusetts Institute of Technology A.I Memo No.1688, CBCL Paper No.188, 2000.
99. C. Nakajima, M. Pontil, and T. Poggio: People Recognition and Pose Estimation in Image Sequences, Proceedings of the International Joint Conference on Neural Networks, Vol.4, pp. 189-194, 2000.
100. O. Chapelle, P. Haffner, and V. Vapnik: Support Vector Machines for Histogram Based Image Classification, IEEE Transactions on Neural Networks, Vol.10, No.5, pp.1055-1064, 1999.
101. C. Papageorgiou, and T. Poggio: Trainable Pedestrian Detection, Proceedings of the IEEE International Conference on Image Processing, pp.35-39, 1999.
102. 高橋圭子, 松浦正樹, 杉山岳弘, 阿部圭一: 人間による画像の色分類結果と領域分割結果に基づいた色空間の比較評価, 電子情報通信学会論文誌, Vol.J84-D-II, No7, pp.1378-1388, 2001.
103. 大津展之, 栗田多喜夫, 関田巖: パターン認識, 朝倉書店, 1996.
104. M. Riesenhuber, and T. Poggio: Hierarchical Models of Object Recognition in Cortex, Nature Neuroscience, Vol.2, No.11, pp.1019-1025, 1999.
105. 内川恵二: 色覚のメカニズム, 色を見る仕組み, 朝倉書店, 1998.
106. C. Nakajima: Posture Recognition of Nuclear Power Plant Operators by Supervised Learning, Proceedings of the IEEE International Conference on Image Processing, Vol.2, pp.877-880, 2004.
107. 高野研一, 野地邦男, 佐野敏明, 佐相邦英, 鈴木弘一: コンピュータグラフィックス CG

- を利用した運転訓練システムの開発その1:システム構成の確立と要素技術の開発, 電力中央研究所研究報告, S00003, 2001.
108. 松尾健一, 今村弘, 田中浩: 屋内における移動体検出への Malib の適用に関する検討, 電気学会通信研究会資料, CMN-04-08, pp.35-38, 2004.
  109. 松尾健一: C 言語オープンソースライブラリを利用した侵入監視プログラムの開発, 電子情報通信学会技術報告, CQ2004-83, pp.25-28, 2004.
  110. 中島慶人: 既設監視カメラを用いた高速な侵入者検知方式, 電力中央研究所報告, R04007, 2005.
  111. 中島慶人: 監視ライン上の歩行者を機械学習により実時間で検知する方式, 電気学会研究会資料, 通信研究会, pp.55-60, 2005.
  112. S. Kang, J. Paik, A. Koschan, B. Abidi, and M. A. Abidi: Real-time Video Tracking using PTZ Cameras, Proceedings of the SPIE International Conference on Quality Control by Artificial Vision, Vol.5132, pp.103-111, 2003.
  113. 高橋祐介, 亀井俊男: パン・チルト・ズームを用いた移動物体追尾システム, センシングフォーラム論文集, Vol.17, pp.55-60, 2000.
  114. 森田俊彦: 局所相関演算による動きの検知と追跡, 電子情報通信学会論文誌, Vol.J84-D-II, No.2, pp.299-309, 2001.
  115. R. Rifkin, and A. Klautau: In Defense of One-Vs-All Classification, Journal of Machine Learning Research, Vol.5, pp.101-141, 2004.
  116. M. Leo, P. Spagnolo, T. D'Orazio, and A. Distanti: Human Activity Recognition in Archaeological Sites by Hidden Markov Models, Proceedings of the IEEE Pacific Rim Conference on Multimedia, pp.1019-1026, 2004.
  117. D.H. Hubel: Eye, Brain, And Vision, Scientific American Library, No.22, 1988.
  118. 矢澤利弘, 中島慶人: 設備図面入力のための図面認識技術の現状調査, 電力中央研究所調査報告, Y89006, 1990.
  119. T. Poggio, and S. Smale: The Mathematics of Learning: Dealing with Data, Notices of the American Mathematical Society, Vol.50, No.5, pp.537-544, 2003.
  120. 米本和也: 固体撮像素子 CMOS の挑戦 CCD との競争・共存, 映像情報インダストリアル, Vol.32, No.6, pp.29-35, 2000.
  121. 電力用画像伝送システム調査専門委員会編: 画像伝送システムの現状と将来, 電力における適用事例, 電気学会技術報告, 第 771 号, 電気学会, 2000.
  122. 辻孝之, 長岡伸治: 夜間の歩行者認知支援システムの開発, 第 11 回画像センシングシンポジウム論文集, pp.387-390, 2005.
  123. 猪毛尾壮一郎, 伊藤憲彦: アナログメータ読取り技術の電力設備自動監視への適用検証, 画像電子学会第 216 回研究会予稿集, pp.81-87, 2005.
  124. T. Poggio, and E. Bizzi: Generalization in Vision and Motor Control, Nature, Vol.431, pp.768-774, 2004.
  125. A.L. Yarbus: Eye Movements and Vision, Plenum Press, 1967.
  126. L. Itti, C. Koch, and E. Niebur: A Model of Saliency-based Visual Attention for Rapid Scene

- Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp.1254-1259, 1998.
127. C.W. Hsu and C.J. Lin: A Comparison of Methods for Multi-class Support Vector Machines, IEEE Transactions on Neural Networks, Vol.13, pp.415-425, 2002.
  128. 瀬戸洋一: バイオメトリックセキュリティ認証技術の動向と展望, 情報処理学会誌, Vol.47, No.6, pp.571-576, 2006.
  129. K.M. Cheung, S. Baker, and T. Kanade: Shape-From-Silhouette Across Time: Part II: Applications to Human Modeling and Markerless Motion Tracking, International Journal of Computer Vision, Vol.63, No.3, pp.225-245, 2005.
  130. T. Poggio, S. Mukherjee, R. Rifkin, A. Rakhlin and A. Verri: "B," In Uncertainty in Geometric Computations, J. Winkler and M. Niranjana editors, Kluwer Academic Publishers, pp.131-141, 2002.
  131. C. Koch: Biophysics of Computation: Information Processing in Single Neurons, Oxford University Press: New York, New York, 1999.
  132. B. Heisele, T. Serre, S. Prntice, and T. Poggio: Hierarchical Classification and Feature Reduction for Fast Face Detection with Support Vector Machines, Pattern Recognition, Vol.36, No.9, pp.2007-2017, 2003.
  133. B. Schölkopf, C.J.C. Burges, and A.J. Smola: Advances in Kernel Methods, The MIT Press, 1999.
  134. B. Schölkopf, J. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson: Estimating the Support of a High-dimensional Distribution, Neural Computation, Vol.13, No.7, pp.1443-1472, 2001.
  135. 中島慶人, 伊藤憲彦: 画像処理による物体認識と位置姿勢判定, 電気学会論文誌, Vol.121-C, No.10, pp.1516-1523, 2001.
  136. 鳥脇純一郎: 3次元デジタル画像処理, 昭晃堂, 2002.
  137. 遠山和也, 福島邦彦: 視野の動きを抽出する神経回路モデル, 電子情報通信学会論文誌, Vol.J89-D, No.2, pp.271-280, 2006.

# 研究業績

## 学術論文 (全文査読付き)

### <主著>

1. 中島慶人：原子力発電所運転訓練時の訓練員の検出と姿勢認識，電気学会論文誌，Vol.125-D, No.1, pp.60-66, 2005.
2. Chikahito Nakajima, Massimiliano Ponti, Bernd Heisele, and Tomaso Poggio : Full-Body Person Recognition System, Pattern Recognition, Vol.36, pp.1997-2006, 2003.
3. 中島慶人, 伊藤憲彦：画像処理による物体認識と位置姿勢判定，電気学会論文誌，Vol.121-C, No.10, pp.1516-1523, 2001.
4. 中島慶人：視覚特性を活用した1枚の写真からの物体形状推定，情報処理学会論文誌，Vol.40, No.8, pp.3179-3187, 1999.
5. Chikahito Nakajima and Toshihiro Yazawa : A Recognition Method of Facility Drawings and Street Maps Utilizing the Facility Management Database, IEICE Transactions on Information and System, Vol.E79-D, No.5, pp.555-560, 1996.
6. 中島慶人, 佐藤真一, 白井良明, 上野晴樹：旋回中の監視カメラで撮影した画像からの高速な侵入者検知，電気学会論文誌投稿中.

### <共著>

1. 堤富士雄, 中島慶人：画像検索と機械学習を組み合わせたビデオ映像認識システム，電気学会論文誌，Vol.121-C, No.10, pp.1524-1529, 2001.

## 国際会議 (全文査読付き)

### <主著>

1. Chikahito Nakajima : Posture Recognition of Nuclear Power Plant Operators by Supervised Learning, Proceedings of the IEEE International Conference on Image Processing (ICIP 2004) , pp.877-880, 2004.
2. Chikahito Nakajima, and Norihiko Itoh : A Support System for Maintenance Training by Augmented Reality, Proceedings of the International Conference on Image Analysis and Processing (ICIAP 2003), pp.158- 163, 2003.
3. Chikahito Nakajima, and Massimiliano Pontil : Maintenance Training of Electric Power Facilities Using Object Recognition by SVM, Proceedings of the International Workshop on Pattern Recognition with Support Vector Machines (SVM 2002), pp.112-119, 2002.
4. Chikahito Nakajima, Norihiko Itoh, Massimiliano Pontil, and Tomaso Poggio : Object Recognition and Detection by a Combination of Support Vector Machine and Rotation Invariant Phase Only Correlation, Proceedings of the International Conference on Pattern Recognition (ICPR 2000), pp.787-790, 2000.
5. Chikahito Nakajima, Massimiliano Pontil, and Tomaso Poggio : People Recognition and Pose

Estimation in Image Sequences, Proceedings of the International Joint Conference Neural Networks (IEEE-INNS-ENNS IJCNN 2000), pp.189-194, 2000.

6. Chikahito Nakajima, Massimiliano Pontil, Bernd Heisele, and Tomaso Poggio : People Recognition in Image Sequences by Supervised Learning, Massachusetts Institute of Technology, A.I Memo No. 1688, CBCL Paper No. 188, 2000.
7. Chikahito Nakajima : Extraction of Salient Apexes from an Image by Using the Function at the Primary Visual Cortex, Proceedings of the International Conference on Pattern Recognition (ICPR 1998), pp.720-790, 1998.
8. Chikahito Nakajima, and Toshihiro Yazawa : Automatic Recognition of Facility Drawings and Street Maps Utilizing the Facility Management Database, Proceedings of the International Conference on Document Analysis and Recognition (ICDAR 1995), pp.516-519, 1995.

<共著>

1. Fujio Tsutsumi, and Chikahito Nakajima : Hybrid Approach of Video Indexing and Machine Learning for Rapid Indexing and Highly Precise Object Recognition, Proceedings of the IEEE International Conference on Image Processing (IEEE ICIP 2001), pp.645-648, 2001.
2. Toshihiro Yazawa, and Chikahito Nakajima : Recognizing the handwritten drawing of the distribution facilities utilizing the Master DB, Proceedings of the Expert Systems Applications to Power Systems (ESAP 1993), pp.636-641, 1993.

## 全国大会・研究会等

<主著>

1. Chikahito Nakajima, Hiromi Asano, Go Suzuki, and Masaru Noda : Position Detection of Oil Refinery Field Operations by Wearable and Surveillance Cameras, Proceedings of the International Symposium on Design, Operation and Control of Chemical Processes, pp.452-457, 2005.
2. 中島慶人 : 監視ライン上の歩行者を機械学習により実時間で検知する方式, 電気学会研究会資料 通信研究会, pp.55-60, 2005.
3. 中島慶人, 伊藤憲彦 : 複合現実による訓練業務支援, 電気学会研究会資料 通信研究会, pp.51-56, 2003.
4. 中島慶人 : 1次視覚野の働きに着目した画像内の特徴点抽出, 人工知能学会全国大会第11回論文集, pp.400-403, 1997.
5. 中島慶人 : 同心円型受容野の応答に着目した主観的輪郭形成の一手法, 電子情報通信学会研究会資料 ヒューマン情報処理研究会 HIP96-11, Vol.96, No.115, pp.75-80, 1996.
6. 中島慶人 : 都市計画図上での用地マスタを利用した配電設備位置自動補正, 電子情報通信学会 第6回機能図形システムシンポジウム論文集, pp.45-50, 1995.
7. 中島慶人, 矢澤利弘 : 配電設備図面の自動認識(2) —設備記号認識に基づく道路の認識—, 情報処理学会第45回全国大会論文集, pp.2-315-316, 1992.
8. 中島慶人, 矢澤利弘 : 架空配電設備図面の設備記号認識に基づく道路の自動認識, 電

- 子情報通信学会画像の認識・理解シンポジウム論文集, pp.2-183-190, 1992.
9. 中島慶人：既存設備マスタを利用した手書き架空配電設備図面自動入力 of の試み, 電子情報通信学会第3回機能図形情報シンポジウム論文集, pp.7-12, 1992.
  10. 中島慶人, 上野晴樹：アルゴリズムの知識に基づくプログラム理解システム, 情報処理学会第36回全国大会論文集論文集, pp.1415-1416, 1998.

<共著>

1. 上野剛, 中野幸夫, 中島慶人：オフィスビルにおけるエレベータ運転台数の変更による省エネルギー効果と便益の変化に関する分析, 第25回エネルギー資源学会研究会講演論文集, pp.69-72, 2006.
2. 上野剛, 中野幸夫, 中島慶人：オフィスビルにおけるエレベータの利用実態計測調査, 第22回エネルギーシステム・経済・環境コンファレンス論文集, pp.503-506, 2006.
3. 堤富士雄, 中島慶人：巡視点検ビデオを迅速にデータベース化, 電子情報通信学会技術報告, パターン認識・理解研究会, Vol.101, No.203, pp.1-8, 2001.
4. 伊藤憲彦, 中島慶人：位相限定相関法を用いた3次元物体の位置・姿勢推定方法, 情報処理学会研究報告, コンピュータビジョンとイメージメディア研究会, Vol.2000, No.82, pp.1-8, 2000.
5. 矢澤利弘, 中島慶人, 坂内廣蔵：配電設備図面入力における初心者支援, 人工知能学会知的教育システム研究会資料, SIG-IES-9602, pp.25-30, 1996.
6. 田中稔, 矢澤利弘, 中島慶人：配電図面の情報管理, 平成7年電気学会電子情報システム部門大会論文集, 1995.
7. 矢澤利弘, 中島慶人：配電設備図面の自動認識—支線の自動認識と効率的な修正方法—, 情報処理学会第49回全国大会論文集, pp.2-109-110, 1994.
8. 矢澤利弘, 中島慶人：配電設備図面の自動認識, 電子情報通信学会第5回機能図形情報シンポジウム論文集, pp.73-74, 1994.
9. 矢澤利弘, 中島慶人：配電設備図面の自動認識(1) —配電設備マスタによる設備記号の認識—, 情報処理学会第45回全国大会論文集, pp.2-313-314, 1992.
10. 上野晴樹, 中島慶人：プログラム理解による論理チェックシステム, 人工知能学会第1回全国大会論文集, pp.449-452, 1987.

## 解説等

<主著>

1. 中島慶人：Support Vector Machine による物体認識を利用した保守訓練支援の試み, オペレーションズリサーチ学会, Vol.46, No.5, pp.237-242, 2001.
2. 中島慶人：1枚の写真から立体画像を作成する試み, Bit, Vol.131, No.9, pp.32-36, 1999.
3. 中島慶人：1枚の写真からの3次元形状推定—視覚特性の活用—, 画像ラボ, Vol.10, No.6, pp.19-22, 1999.
4. 中島慶人：配電設備記号の位置自動補正, 画像ラボ, Vol.6, No.10, pp.42-45, 1995.

<共著>

1. 矢澤利弘, 中島慶人: 配電設備マスタを利用した配電設備図面の自動認識, 電気現場技, Vol.31, No.367, pp.40-43, 1992.

## 電力中央研究所研究報告

<主著>

- 1 中島慶人: 既設監視カメラを用いた高速な侵入者検知方式, 研究報告 R04007, 2005.
- 2 中島慶人, 伊藤憲彦: 簡易な複合現実環境生成システムによるシナリオに基づく訓練環境の構築, 研究報告 R02028, 2003.
- 3 中島慶人, 伊藤憲彦: 保守訓練支援用バーチャルリアリティシステムの簡易実現, 研究報告 R00029, 2001.
- 4 中島慶人, 伊藤憲彦: 画像処理によるリアルタイム物体認識と位置姿勢推定 -Support Vector Machine と Rotation Invariant Phase Only Correlation の利用, 研究報告 R00003, 2001.
- 5 中島慶人: 1枚の写真からの3次元形状復元方式, 研究報告 R97013, 1998.
- 6 中島慶人: 人間の視覚処理手順に習った画像の特徴抽出方法, 研究報告 R96006, 1997.
- 7 中島慶人: 地図情報システムにおける配電設備の自動位置決め方式, 依頼報告 R94501, 1995.
- 8 中島慶人, 矢澤利弘: 地形図上の配電ネットワーク図面の自動認識, 研究報告 Y92014, 1993.
- 9 中島慶人, 矢澤利弘: 設備図面の自動認識(1) -単純な回路図面の認識-, 研究報告 Y90022, 1991.

<共著>

- 1 上野剛, 中野幸夫, 中島慶人: エレベータの運転台数変更による省エネルギー効果と利用者便益の変化に関する定量的分析-オフィスビルにおける事例検討-, 研究報告 R06001, 2006.
- 2 伊藤憲彦, 中島慶人: 位相限定相関を用いた3次元物体の位置・姿勢検出法, 研究報告 R01001, 2001.
- 3 堤富士雄, 中島慶人: 巡視・点検映像のリアルタイムデータベース化手法(その3)-模様の特徴を用いた映像認識手法の高速・高精度化-, 研究報告 R00033, 2001.
- 4 渡部直人, 萩原豊, 伊藤憲彦, 中島慶人: バーチャルリアリティ技術による放射性物質輸送総合支援システムの概念検討, 調査報告 U99052, 2000.
- 5 楠淳, 坂内廣蔵, 石野隆一, 中島慶人: 携帯用情報端末を用いた現場業務支援システムの開発と評価, 研究報告 R97029, 1998.
- 6 楠淳, 石野隆一, 村田博士, 中島慶人: 現場業務支援へのマルチメディア技術の適用, 調査報告 R95024, 1996.
- 7 中島慶人: 地図情報システムにおける配電設備の自動位置決め方式, 依頼報告 R94501, 1995.

- 8 矢澤利弘, 中島慶人: 配電設備図面の自動認識, 依頼報告 R93502, 1994.
- 9 矢澤利弘, 中島慶人: 設備図面の自動認識 (2) - 配電設備マスタを利用した手描き線路図の理解 -, 研究報告 Y92001, 1992.
- 10 矢澤利弘, 中島慶人: 設備図面入力のための図面認識技術の現状調査, 調査報告 Y89006, 1990.

## その他

### <主著>

1. 中島慶人: 監視カメラを用いた高速な侵入者検知, 電力中央研究所研究年報, pp.104-105, 2006.
2. 中島慶人: 監視領域への接近者を実時間で検知する画像監視方式, 電力中央研究所研究発表会 - 情報通信部門予稿集, pp.73-76, 2005.
3. 中島慶人: 可視画像での照明変化と背景変化に対応した人物検知方式の提案, 研究調査資料 R03905, 電力中央研究所内資料, 2004.
4. 中島慶人, 伊藤憲彦: 複合現実環境の評価用訓練シナリオ, 研究調査資料 R01908, 電力中央研究所内資料, 2002.
5. 中島慶人: 保守訓練支援用バーチャルリアリティシステム, 電力中央研究所研究発表会 - 情報通信部門予稿集, pp.57-60, 2001.
6. 中島慶人: 動画像からの物体抽出と位置・姿勢推定, 電力中央研究所研究発表会 - 情報通信部門予稿集, pp.81-84, 2000.
7. 中島慶人, 伊藤憲彦: 連続画像からの3次元情報取得の予備的検討, 研究調査資料 R98920, 電力中央研究所内資料, 1999.
8. 中島慶人: 人の視覚処理に習った画像の特徴点抽出, 電力中央研究所研究発表会 - 情報通信部門予稿集, pp.115-122, 1997.
9. 中島慶人: 初期視覚の情報処理モデルに関する調査, 研究調査資料 R95906, 電力中央研究所内資料, 1996.
10. 中島慶人: アルゴリズムの知識に基づくプログラム理解に関する研究, 東京電機大学理工学研究科システム工学専攻修士論文, 1989.
11. 中島慶人: 知的プログラミング支援システム INTELLITUTOR の開発 - プログラムの意味理解についての基礎研究 -, 東京電機大学理工学経営工学科卒業論文, No.6130, 1987.

### <共著>

1. 伊藤憲彦, 中島慶人: 位相限定相関を用いた3次元物体の位置・姿勢検出における問題点, 研究調査資料 R00908, 電力中央研究所内資料, 2001.
2. 伊藤憲彦, 中島慶人: 複合現実感を実現する画像処理技術に関する調査, 研究調査資料 R97904, 電力中央研究所内資料, 1998.
3. 楠淳, 石野隆一, 中島慶人: 現場支援システムの開発 - 電力へのマルチメディア適用に向けて -, 研究調査資料 R96914, 電力中央研究所内資料, 1997.

4. 石野隆一, 中島慶人: 3次元画像計測技術の現状と研究課題, 研究調査資料 R95911, 電力中央研究所内資料, 1996.
5. 矢澤利弘, 中島慶人: 配電図面自動認識システム, 電力中央研究所研究発表会—情報通信部門予稿集, pp.84-85, 1993.
6. 矢澤利弘, 中島慶人: 配電設備マスタを利用した配電設備図面の自動認識, 情報処理研究, 電力中央研究所, No.19, pp.33-42, 1993.
7. 矢澤利弘, 中島慶人: 設備図面の自動認識—単純な回路図面の認識—, 情報処理研究, 電力中央研究所, No.18, pp.33-42, 1991.
8. 系統・情報通信研究分科会: 電力設備等に対するバーチャルリアリティ (VR) の適用化基礎研究, 中地域共同研究報告 N94101, 1995.
9. 上野晴樹, 中島慶人, 他2名: 知的プログラミング環境におけるプログラムの意味と理解に関する報告, 情報処理振興事業協会技術センター, 1988.

## 特許出願

1. 中島慶人: 画像処理による移動体候補の検出方法及び移動体候補からの移動体を検出する移動体検出方法, 移動体検出装置及び移動体検出プログラム, 特願 2006-005921, 2006.
2. 中島慶人: 画像処理による侵入物体検出装置および方法およびプログラム, 特願 2005-199990, 2005.
3. 渡部直人, 萩原豊, 中島慶人, 伊藤憲彦: 現場業務支援システム, 特開 2001-331216, 2001.
4. 中島慶人, 矢澤利弘, 田中稔, 伊庭博文: 設備図面からの電柱の支線方向及び支柱方向自動認識システム, 特開平 8-83334, 1996.
5. 矢澤利弘, 中島慶人, 田中稔, 伊庭博文: 設備図面からの電柱自動認識システム, 特開平 8-83333, 1996.