博士論文

高安全知能コンピューティングシステム
に関する研究

石巻専修大学大学院理工学研究科博士後期課程物質機能工学専攻

DM19－0101A　高 田 健 一

指導教員　工藤すばる 教授

目 次

第１章
緒　言

情報システムの利用が広く市民生活に浸透し，高度情報化社会が進展していく中で ICT は私たちの生活にとってなくてはならないものとなり，暮しの様々な場面で活用されている．このような社会状況の中にあって，私たちの生活環境で発生する犯罪，急病，事故などの様々な危険事象に対して，人間の安全を支援するためのプラットフォームとして共通に適用可能な高安全知能コンピューティングシステムの実現が望まれている [1].

たとえば，犯罪行為への備えとして，近年，防犯カメラ等が広く公共の場に設置されてきてはいるが，これらのカメラの映像は犯行後の犯人検挙の手がかりや証拠として用いられ，犯罪行為を未然に防止するために用いられることはほとんどない．防犯カメラ等の設置には強盗，窃盗，通り魔的事件などの犯罪行為をある程度抑止するための効果があると考えられるが，それらの発生を十分に防ぐことは難しいのが現状である．このような犯罪行為に対し，これまで以上に高度な ICT を活用した不審人物の検出などの防犯・監視の対策が求められている．

また，高齢化社会の進展に伴い医療・健康分野においても，ICT は遠隔医療，診断支援など，地域医療や介護とも幅広く関連している．特に独居老人

の介護などにおいては，対象者の日常の様子や体調不良，命にかかわるような急病などによる状態の変化をタイムリーに見守るためのシステムの開発が必要となってくる．

交通事故防止にも ICT は必要不可欠である．交通事故の原因の 9 割以上は人間の認知・判断・操作のミスによりものであり，運転免許所有者の高齢化に伴い，逆走などの交通違反や車の操作ミスなどの人的要因による交通事故は増加傾向にある．このため歩行者やドライバーの安全を確保するためのシステムの構築が必要となってきている．

また，地球温暖化の影響と考えられる記録的な豪雨や台風による洪水や土砂災害などが近年増加しており，危険を感じたら自治体の避難勧告等を待たず速やかに避難することが重要となってきている．自然災害においては、不測の事態も想定されることから，必ずしもあらかじめ定められた避難場所等へ避難することが適切でない場合もあり，自らの判断で自宅などでの垂直避難等の検討も必要となる．同様に，大地震によって発生する津波に対しても速やかな避難行動をとることが最も重要である．地震による倒壊家屋，火災などの発生や道路の渋滞などにより，事前に検討した避難経路を通行できない場合や土地勘のない旅先などでの避難も想定され，ICT を用いた迅速な避難の支援が必要となる．

以上のような，犯罪，急病，事故，自然災害などの様々な危険事象に対し，それぞれ個別 にその検出や行動計画の決定手法などを開発しようとした場合，膨大な期間，経費，労力などのコストが発生してしまい，容易にこのようなシステムを導入することが難しくなってしまう．

本研究では，このような課題を解決し様々な危険事象に共通に適用可能な高安全知能コンピューティングシステムを実現するためのプラットフォーム

を提案している. このようなシステムの体系的な構成については報告がなされていないため, まず, 高安全知能コンピューティングシステムのプラットフォームの構成を検討する必要がある. ここで言うプラットフォームは, 対象とする危険事象に合わせ, その検出などを行うための学習データや安全行動の内容などを更新するだけでシステムを様々な危険事象に適用可能とするための共通基盤である. 本システムを構成するのは, 画像情報から危険事象を検出するための知能情報処理, 危険事象の発生が予測可能な場合の危険予測処理及びこれらの処理結果に基づき安全を確保するための行動計画を決定する知能情報処理である. このように, 検出, (予測), 行動計画により構成されたシステムは, たとえば, 犯罪行為を検出しその対策を実行したり, あるいは犯罪行為の発生を予測し未然防止対策を講じたり, また, 急病人が発生した際の救護活動を支援するなど, 対象とする危険事象が変更されたとても, その構成を変更することなく様々な危険事象に対し共通に適用が可能となる.

次に, 本システムの各知能情報処理について述べる. 危険事象検出については, 画像情報が最も重要な情報であることに着目し, 監視カメラ等からの画像情報を用いることとした. このような画像情報を用いた検出手法には, 暴力行為, 人物の転倒, 交通事故や異常走行, 工場での火災, プールでの溺者などのための様々な手法が報告されている[2, 3, 4, 5, 6]. しかしながら, これらの手法はある特定の事象のみを取り上げ, その危険事象が発生していない通常の状態に固有の特徴量を抽出・学習し, 通常の状態から大きく逸脱しているものを危険であると検出するという手法が取られている. したがって, これらの手法のほとんどは, 設定した危険事象の検出に限定されるため, 汎用的に本システムのプラットフォームとして活用することが困難で

ある．そこで，本システムではこの問題を解決するため，シーン認識の手法を提案している．この手法では，深層学習（ディープラーニング）の 1 つである畳み込みニューラルネットワーク(CNN：Convolutional Neural Network)を画像の特徴抽出器として利用し，抽出した特徴量をサポートベクトルマシン(SVM：Support Vector Machine)で学習する．これにより，静止画による高精度のシーンの認識が可能となることを明らかにする[7]．しかしながら，このシーン認識では，たとえばコンビニ・銀行で発生する強盗，路上での転倒などのある特定の 1 種類のシーン（単一シーン）の発生の有無を認識するためのものであり，複数のシーンを同時に認識する（多クラス分類）ことには対応していない．このような複数種類のシーンが発生する可能性のある生活空間においても，高精度の多クラス分類を可能とする手法として識別器選択方式を提案する[7]．更に，心理学で用いられる感情価（快／不快）の考え方を導入することにより，学習コストを軽減し，この多クラス分類の汎用性の向上を実現させる手法を提案する．

また，これらのシーン認識は静止画を用いたものであり，暴力行為など人間やその他の物体の動きにより，はじめて不審・異常な行動，危険な動作であると判断可能となるシーンを認識できないという問題が発生する．これを解決するためには，動画を用いてシーン認識を行う必要があり，高精度な静止画シーン認識手法を動画に対応したものへ拡張することによりこれを可能とする手法を提案する[8]．提案手法では，動画のフレーム(静止画)に RGB 化したオプティカルフローの画像を重畳した画像（重畳画像）を用いることにより，静止画シーン認識の手法を活用し少ない計算量で動画シーン認識を可能とする．更に，動画シーン認識結果を決定する過程に隠れマルコフモデルを導入し，重畳画像の識別結果の系列から真に危険事象が発生している状

態なのか否かを推定することにより, シーン認識の精度を向上させる手法を提案する.

危険事象の発生を事前に検知するための危険予測の処理については, 危険事象発生前にそれと因果関係があると推測される事象が発生している画像(前兆シーン) を学習することにより, シーン認識の手法を用いて前兆シーンを認識可能である. したがって, 前兆シーン認識におけるその確からしさは, 危険事象発生の可能性を示すものとなりうる. しかしながら, 本システムの危険予測においては, 前兆シーンが非常に高い確率で危険事象の発生を示唆しているという知見が得られている場合を除いては, 前兆シーンが認識されたからと言って必ずしも危険事象が発生するとは限らない, ということを仮定している. たとえば, 店内をうろうろし落ち着きがないように見える人物が何らかの犯罪行為を実行するとは限らないし, また, しゃがみこんでいる人物が必ずしも急病を発症するとは限らない. それらの人物はただ商品を探しているだけかもしれないのである. したがって, 前兆シーン認識の精度を向上させても, 危険予測の精度を必ずしも向上させることにはならない場合がある. この問題を解決するため, 新たにシーン認識と統計データ分析を組み合わせた危険予測の手法を提案する. 提案手法では, シーン認識のクラス事後確率と犯罪発生などに関わる統計データに基づき, 前兆シーンの発生時点（時間, 場所, 天候など）における危険事象発生確率から, リアルタイムでの危険予測を行うことが可能となる.

危険事象検出, あるいは危険予測に基づく行動計画については, 身体的損傷や経済損失などの被害をできるだけ小さくするための行動を決定するという重要な役割を果たすものであり, 対応すべき危険事象に対して, リアルタイムで取得された利用場所, 設置環境及びユーザ固有の条件に適用的に合わ

せた安全確保の支援を実現する方法について提案する.提案手法ではその知能処理を階層化することにより,様々な危険事象に適応可能なプラットフォームとして機能させることができる.行動計画の第1階層では,グローバルな安全行動を選択し,第2階層ではこれを実現するための詳細行動を決定する.また,第1階層で選択される行動候補には,あらかじめ優先順位付けを行うことにより,それらの行動が実行された場合の被害を未然に防止する,あるいは被害を最小限にくい止める効果(行動の効果)を十分に上げることが可能となる.しかしながら,この優先順位付けにおいて問題となるのは,行動の効果に関する事例データが存在しない,あるいは非常に少なく,これらのデータから客観的にそれらの優先順位を決定できない場合である.この課題に対しては,特に,危険予測に基づいて選択される行動候補の優先順位付けの効果に関する事例データが存在しないことから,階層分析法を用いて,システム設計者が主観的に危険予測結果を考慮して優先順位付けする手法を提案する.

また,特に,第1階層において,危険回避のための最も典型的な行動である「避難」が選択された場合,その詳細行動である避難経路の探索問題は非常に重要な課題となる.この課題に対しては,シーン認識により得られる,時々刻々変化する経路環境情報に基づき,リアルタイムで最善の避難経路を探索・決定する手法を提案する.

本論文は,以上の内容を取りまとめたものであり,以下に示す6章により構成される.

第1章は緒言であり,本研究の背景,目的及び概要を述べたものである.

第 2 章では，高安全知能コンピューティングシステムにより，様々な危険事象に対して共通に利用できるプラットフォームを実現するためのシステムの構成とその適用範囲について明らかにする．

第 3 章では，危険事象検出のための CNN と SVM を組合せたシーンの認識手法を提案し，評価実験により，高精度の静止画シーン認識が可能となることを明らかにする．また，識別器選択方式による多クラス分類の手法と感情価を用いた多クラス分類の手法を提案し，評価実験によりこの手法が単一シーンの認識と同程度の認識精度で多クラス分類が可能となることを明らかにする．

第 4 章では，静止画シーン認識の手法をベースとした動画シーン認識の手法を提案し，静止画では認識が難しい動きのあるシーンを高精度に認識可能であることを評価実験により明らかにする．

第 5 章では，前兆シーンの認識の結果と統計データの分析結果を組み合わせた危険予測の手法を提案する．これにより，リアルタイムで精度の高い危険予測が可能となることを，適用例をとおして明らかにする．

第 6 章では，行動計画における階層化された知能処理の過程を示し，階層分析法を用いた行動候補の優先順位付けが有用な手法であることを明らかにする．また，詳細行動決定問題の典型例として，シーン認識に基づいた避難経路探索問題を取り上げ，適用例によりその有効性を明らかにする．

第 7 章は結言である．

以上，本論文の企図するところを概説した．

第２章
危険事象に対する安全性を実現するための知能情報処理

2.1　まえがき

人間は，危険事象に遭遇した際，パニック状態になったり，その危険性を過小評価したり，冷静な判断が不可能となってしまう場合が多々あることが指摘されている[9，10]．このような状況下において，高安全知能コンピューティングシステムは，安全を確保するために人間がとるべき最善の行動を知能情報処理によって支援することを目的としている．また，様々な危険事象に対して共通に利用できるプラットフォームをこの高安全知能コンピューティングシステムにより実現することも重要な課題である．本章では，様々な危険事象に対して，人間の安全性を確保するための行動支援が可能となるような高安全知能コンピューティングシステムの構成等を提案し，その知能情報処理の考え方について述べる．

2.2　高安全知能コンピューティングシステムが対象とする危険事象

高安全知能コンピューティングシステムは，利用場所，設置環境及びシステムユーザ固有の条件に合わせて，人間ができるだけ安全に様々な危険事象

に対処できるよう，その行動を計画し，提示等を行うことによりユーザを支援するシステムである．

したがって，本システムが対象とする危険事象は，表 2.1 に例示するような，強盗などの犯罪行為，急病人等の発生など，ユーザが自らシステムの支援を受け対処可能な危険事象である．また，危険事象を検出するために，画像認識の分野で特に目覚しい成果を挙げているディープラーニングに着目し，また，レーダー信号，GPS による位置情報，赤外線センサーなどの他の情報ソースと比較して，危険事象検出には重要な情報が含まれている画像情報を用いていることとした．したがって，本システムでは監視カメラやスマートフォン等からの映像により検出が可能な危険事象を対象としている．ただし，津波などの自然災害の発生の検出については，本システムが用いる画像情報のみでは困難な場合もあることから，インターネットからの情報，たとえば，気象庁などから発信される津波到達時間などの情報を取得することによりユーザを支援する必要がある．

また，これらの危険事象が確実に発生するとは限らないが，その発生の可能性を示す前兆現象であると考えられる事象が存在し，それを画像情報として取得可能な場合は，それらも本システムが対象とする事象に含むものとする．

表 2.1　高安全知能コンピューティングシステムが対象とする危険事象とそのシーン認識対象及び行動支援内容の例

危険事象	シーン認識対象	行動支援内容
万引きなどの窃盗	犯行前の不審者の行動，犯行	不審者，犯人への対応 防犯設備（防犯ベル，警報ランプなど）の稼働，通報など
強盗，暴力行為，通り魔など	犯行前の不審者の行動，犯行	不審者，犯人への対応 防犯設備（防犯ベル，警報ランプなど）の稼働，通報，避難支援など
テロ	犯行前のテロリストの行動， テロ行為	通報，避難支援など
急病，けが	急病人等の状況	容態確認，応急処置，通報，急病予知など （大規模な危険事象発生に伴うけが人等発生の場合は，避難支援）
建物火災	通路の状況	避難支援，（一般住宅の場合は，消火設備稼働，通報なども含む）など
津波，洪水などの災害	道路の状況	避難支援，避難所情報の取得など
交通事故，危険運転	車両，歩行者， 運転手の状況	歩行者や運転手への警告，事故回避行動支援，通報など
危険生物	クマ，イノシシなどの出没	追い払い，警報，通報など

その他の対象とする危険事象としては，水難事故，転落（高所など）事故などが考えられる．

2.3　高安全知能コンピューティングシステムのプラットフォームの構成

高安全知能コンピューティングシステムでは，リアルワールドからの情報から危険事象を検出し，また，可能な場合には危険予測を行い，これらの情報に基づき行動計画において人間の安全な行動を支援するシステムである．図 2.1 に本システムのプラットフォームの構成を示す．

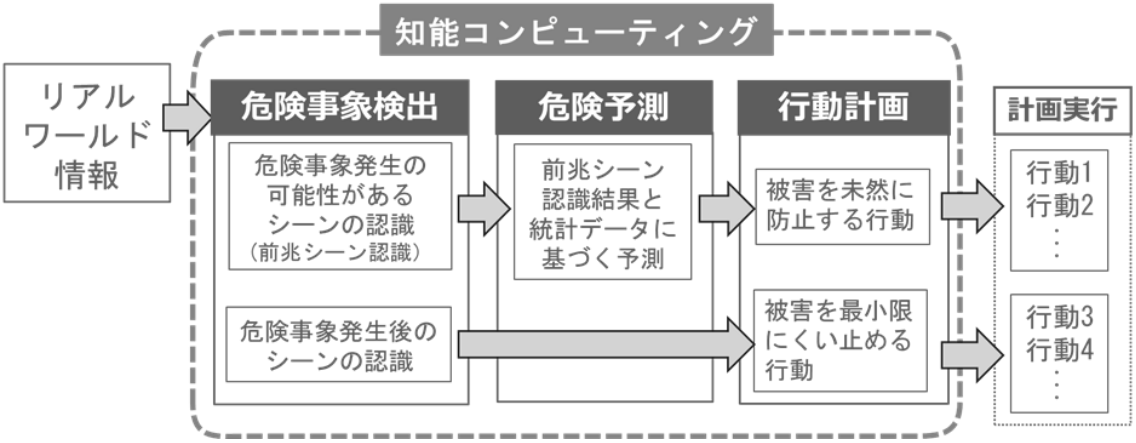


図 2.1　高安全知能コンピューティングシステムのプラットフォームの構成

本システムでは，図 2.1 のとおり，まず，システムが設置された環境において危険事象が発生しているかどうかを検出しなければならない．そのために監視カメラ等からの画像情報を用いる．したがって，危険事象の発生を見落とさないためには，それらの場面（シーン）の画像を高精度に認識する必要があり，これを実現するのが図 2.2 に示したシーン認識であり，その詳細は第 3，4 章で述べる．また，危険事象が発生する可能性がある場合に起こる事象の存在が分かっていて，画像情報として取得可能な場合には，そのシーン（以下，前兆シーンと呼ぶ）の認識も可能であることから，それも危険事象検出に含むものとする．すなわち，危険事象検出は，シーン認識に基づ

き行うということである.表 2.1 に危険事象を検出するためのシーン認識の対象を示したが，たとえば，コンビニ強盗について言えば，その前兆シーンは，店の外から店内をうかがう様子や不審人物の入店，店内をウロウロする様子などが考えられる．また，コンビニ強盗が発生してしまった場合には，その犯行シーンが認識対象となる．

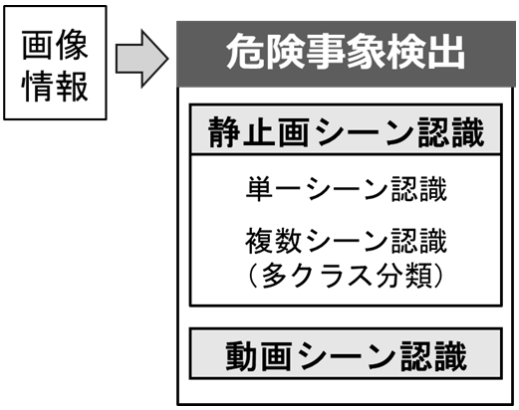


図 2.2　危険事象検出のための知能情報処理（シーン認識）

なお，これらのシーン認識結果については，前兆シーンの認識の場合はその確率（クラス事後確率）を出力とし，危険事象が発生してしまっているシーンの認識の場合については，その発生の有無（バイナリ）を出力としている．

このように,危険事象検出において得られる前兆シーンの認識結果などの情報から,危険事象発生の可能性を予測することを危険予測と呼ぶこととする．この危険予測については，第 5 章で提案する手法を用いる．この手法では,過去に発生した危険事象に関係する統計データの分析結果を予測に生かすことが可能であることから，前兆シーンの認識結果だけではなく，これらを組み合わせることにより,リアルタイムで精度の高い危険予測が可能となる．たとえば，警察署に隣接したスーパーマーケットにおいて，これまで一度も強盗が発生したことがない場合（統計データ），ある時不審者に見える

人物が認識（前兆シーン認識）されたとしても，その人物が今ここ（リアルタイム）で犯罪を実行する可能性は非常に低いと考えることは自然なことである．

この危険予測により，図 2.3 に示すとおり，危険事象が発生する危険性の度合いを得ることができる．

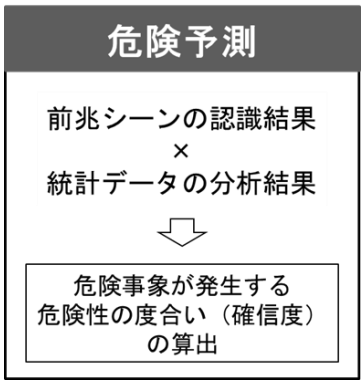


図 2.3　シーン認識に基づいた危険予測のための知能情報処理

行動計画においては，安全性をできるだけ考慮した行動を数理的に決定し，危険事象発生の未然防止策やユーザへの注意喚起，関係機関への通報，避難経路の指示など，被害を軽減・回避するための様々な支援を行う．表 2.1 の行動支援内容の欄に行動計画に基づいた各危険事象に対する支援内容の例を示している．この行動計画では，危険予測が可能な場合，例えば犯罪など人為的に発生する危険事象については，その発生を未然に防止する行動をユーザに提示することが可能である．しかしながら，洪水などの自然災害については，危険予測が可能であっても，災害そのものの発生を未然に防止する行動を取ることは困難であるため，被害を最小限にくい止めるための避難行動を提示することとなる．また，未然防止行動を取ったにもかかわらず，危険事象発生の防止に失敗した場合や前兆シーン認識が不可能で危険事象が発生してしまった場合には，同様に被害を最小限にくい止める行動を提示することとなる．この行動計画は，図 2.4 に示すとおり，階層化されており，第 1

階層では，優先順位付けされたグローバルな行動候補を選択し，第 2 階層ではこれを実現するための詳細行動を決定する．このように階層化するのは，各行動候補に対応する詳細行動が複数存在するため階層化しない場合，行動候補数が増大してしまうことを防ぐためである．これにより，行動計画を様々な危険事象に適応可能なプラットフォームとして機能させることが可能となる．その詳細は第 6 章で述べる．

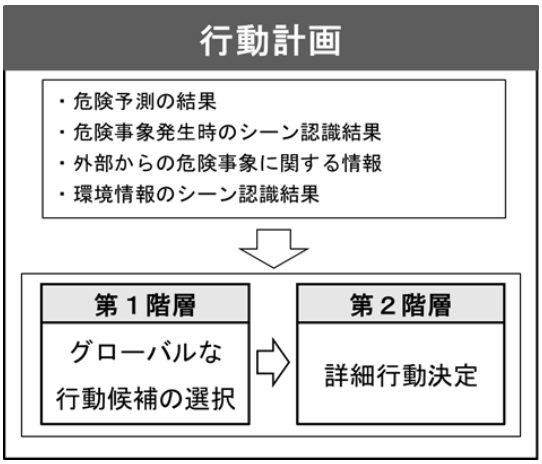


図 2.4　行動計画を決定するための知能情報処理

以上のように，高安全知能コンピューティングシステムは，危険事象検出，危険予測（ただし，危険予測が不可能な場合は，このプロセスは実行されない．），行動計画の各知能情報処理プロセスが連携して機能することにより，安全に様々な危険事象に対処できるようユーザを支援することが可能である．すなわち，危険事象の種類が変わっても，各プロセスで用いられるシーン認識などの知能情報処理は，汎用性が高くそれらの手法自体を変更する必要がないため，対象とする様々な危険事象に合わせて，シーン認識の学習用データを変更したり，行動計画を修正（組み換え）したりするだけで，共通にシ

ステムの一連のプロセスを活用できる．このことは，本システム自体もプラットフォームとして応用範囲を容易に拡充可能であることを示している．

2.4 むすび

高安全知能コンピューティングシステムが対象とする危険事象検出，危険予測及び行動計画の３つの知能情報処理によってシステムが構成されていることを示した．これにより，強盗や窃盗などの犯罪行為，暴力行為，車の危険走行，急病，火災，洪水やがけ崩れなどの自然災害，水難事故などの様々な危険事象に対して応用可能なプラットフォームとしての本システムが，ユーザにとって最も安全と思われる危険回避などの支援を自律的に行うことが可能となる．

第３章
静止画シーン認識に基づく危険事象検出

3.1 まえがき

高安全知能コンピューティングシステムを構築するにあたっては，まず，危険事象が発生していることをいかにして高精度に検出できるかが重要な課題である．これまでの危険事象検出の手法については，暴力行為，人物の転倒，交通事故や異常走行，工場での火災，プールでの溺者など様々な手法が報告されている[2，3，4，5，6]．しかしながら，これらの手法はある特定の事象のみを取り上げ，その危険事象が発生していない通常の状態に固有の特徴量を抽出・学習し，通常の状態から大きく逸脱しているものを危険であると検出するという手法が取られている．したがって，これらの手法のほとんどは，設定した危険事象の検出に限定されるため，同一の手法により他の危険事象の検出が難しいため，汎用的に本システムのプラットフォームとして活用することが困難である．

本章では，監視カメラなどの画像（静止画）から得られる危険事象等が発生している場面（シーン）であるか否かを認識することにより，様々な危険事象を検出するためのプラットフォームとなるシーン認識の手法について提案する．提案する 本システムで用いるシーン認識は画像に映し出されているある事象のシーンが危険か否かを認識するためのものである．

3.2 高安全知能コンピューティングシステムにおけるシーン認識の定義

画像を用いて私たちの生活のあるシーンを認識するためには，画像全体の領域の空間的構成やその関係に着目する必要があるとの考えから，これまでさまざまな手法が考案されてきた．それらの中でも大域的特徴量としてGIST特徴を用いたシーン認識は最も有名なものの 1 つである[11]．

現在ではこのようなシーン認識にもディープラーニングの1つである畳み込みニューラルネットワーク（CNN：Convolutional neural network)が利用されており[18, 20]，画素情報をそのまま CNN に入力し，特徴を自動的に学習してSuperpixel(スーパー画素)によるセグメンテーションも併用しながら，各ピクセルにシーンラベルを付与するシーン認識が行われている．Superpixel とは，輝度や色などの性質が似ている小領域で，処理の単位要素をピクセルから Superpixel に変更することで処理量を大幅に減少させることが可能なため広く利用されている[12]．また，膨大なラベル付きのシーン画像データベースを CNN で学習する手法も研究されている[13]．

しかしながら，そのシーン認識の精度は，約 50%程度(人間が約 80%といわれている)である[14]．現状では高安全知能コンピューティングシステムへの応用が可能なレベルに達しているとは言い難い．また，ここで述べたシーン認識については，静止画を用い，屋内・屋外，寝室・キッチン，駐車場・道路といった「場所」を特定するタスクが含まれているが，高安全知能コンピューティングシステムの危険事象検出が必要としている「画像に写っているシーンが，危険事象等が発生しているシーンであるか否か．」を識別するためのシーン認識は実現されていない．

すなわち，本研究でいうシーン認識とは，場所を特定するためのものではなく，危険事象等が発生しているシーンを認識するためのものである．

次節ではこれを実現するための手法を述べる.

3.3 静止画を用いた単一シーンの認識

コンビニ強盗が発生しているシーンの認識を例に,高安全知能コンピューティングシステムの危険事象検出で用いる静止画を用いた1種類のシーン(以下，単一シーンと呼ぶ）の認識手法について提案する.

まず,提案手法と一般に考えられているコンビニ強盗シーンの認識手法について比較する．一般には.

(1) コンビニ強盗が発生したと判断するための条件(凶器，帽子，マスクなど特定のオブジェクトの有無)をあらかじめ設定（判断条件）

(2) 画像内にある物体を検出（物体検出）

(3) 判断条件と検出物体との照合

(4) コンビニ強盗発生の有無の認識

といったプロセスを組合せて,コンビニ強盗が発生していることを認識することが可能であると考えられている.このため,物体カテゴリ認識や物体検出の精度向上に重きが置かれていたと考えられる.

ところが，コンビニ強盗が発生していると判断する条件の設定は，凶器の種類や犯人の服装など多様であり,個々のオブジェクトの組合せに基づく論理推論による条件設定には限界がある.また,画像には多くのオブジェクトが存在し，照明の具合，物体の位置・スケールのなども様々で，これらのオブジェクトを正確に検出するのは容易ではない．たとえば,「凶器が死角に入って検出できない.」,「花粉症，あるいは新型コロナウイルス感染予防のため，めがねとマスクをしている.」,「刃物を購入しようとしてレジへ持っていった.」といった場面では，コンビニ強盗が発生しているか否かの正確な状況の認識は不可能であり,コンビニ強盗発生の見落としなどの重大な誤

認識が発生してしまう．このような発想により，コンビニ強盗シーンを認識するは難しいと言える．そこで，凶器の有無や服装などをひとつひとつ正確に確認せずとも，画像全体から得られる何らかの特徴に基づき，直感的にコンビニ強盗シーンであると認識可能な手法を検討した．

本研究においては，画像，音声，言語などの分野で従来手法を圧倒する高い認識能力を示しているディープラーニング，その中でも特に画像認識の分野で目覚しい成果を挙げている CNN に注目した．この CNN の物体カテゴリ認識能力は，ILSVRC2016(ImageNet Large Scale Visual Recognition Challenge　2016)の結果では誤認識率約 3%となっており，人間の誤認識率と言われている 5.1%を上回るレベルに達している．

CNN は多階層ネットワーク構造となっており，膨大な数のパラメータ（重み）の最適化が必要となる．したがって，CNN の認識精度は学習に使用されるデータ量とその多様性によって大きく左右され，精度の高い学習のためには，数千から数百万もの教師ラベル付の学習用の静止画（トレーニングデータセット）と膨大な学習時間を要するため，たとえば，前述した場所の特定に関するシーン認識の分野では，大規模なデータベース[15]が構築され活用されている．しかしながら，本研究のテーマである危険事象発生シーンに関連する大規模なデータベースは，そのデータ収集の困難性から構築されておらず，また，本研究において大規模な危険事象発生シーンのトレーニングデータセットを構築することも容易ではない．したがって，このような状況においては，高精度なシーン認識結果を得るための CNN をゼロから生成することは難しいと言える．

そこで，本研究ではディープラーニングのフレームワークである“Caffe”により公開されている静止画を用いた物体カテゴリ認識のために学習済みの CNN モデル “bvlc_googlenet.caffemodel”(validation set における top-5 正解率は 88.9% である．)を利用する[15, 16]．このモデルは，Berkeley Vision

Learning Center（BVLC）が ImageNet[17]の ILSVRC2012 のデータセットにより学習させた GoogLeNet[18]の複製で，1000 種類の物体カテゴリ認識が可能である．

しかしながら，この学習済み CNN は，前述のとおり物体カテゴリ認識用に生成されたネットワークであり，当然，シーン認識の結果を得ることはできない．そこで，この学習済み CNN から抽出した画像の特徴量が他のタスクでも有効に活用が可能であるということに着目し[19], この学習済み CNN を単に画像の特徴抽出器として利用した．この特徴量を SVM により学習することによりシーンの認識を行う．なお，このシーン認識に SVM が最も適していることは，後述する評価実験により示す．

また，トレーニングデータセットについては，認識対象を 1 種類（以下，単一シーンと呼ぶ）とする場合は，認識対象とするシーン（コンビニ強盗シーン）を正例，正例シーンと同一環境（コンビニ）における正例シーン以外のシーン（通常の営業を行っているシーン）を負例とし，それぞれのシーンの画像を複数枚含んでいる．このトレーニングデータセットを機械学習することにより，正例／負例を識別する識別器を生成する．この識別器はトレーニングデータ以外の未知画像が入力されるとその識別結果，すなわちシーン認識結果を出力する．

なお，本研究では，2 値分類，2 クラス分類などと呼ばれているタスクを「識別」と呼ぶこととする．

図 3.1 に，静止画を用いた単一シーン認識の手法を示す．

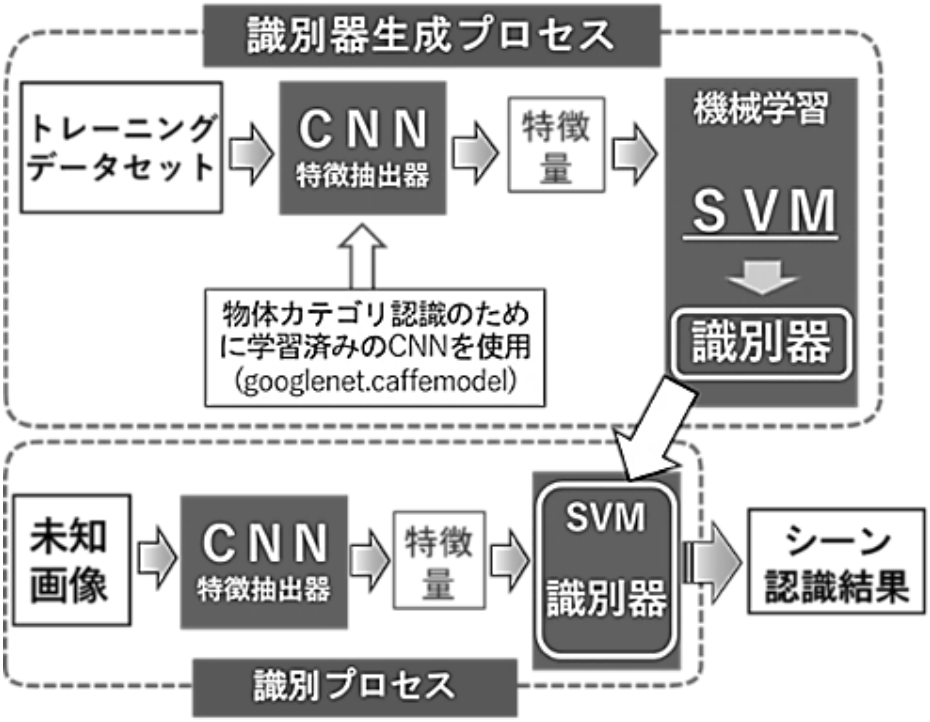


図 3.1　静止画を用いた単一シーン認識の手法

（1）識別器生成プロセス

トレーニングデータセットの画像を学習済みＣＮＮ（GoogLeNet）に入力し，画像の特徴量を抽出する．それらの特徴量を SVM で機械学習し，識別器を生成する． なお, CNN 特徴ベクトル抽出及び学習に使用したプログラムを付録に示す.

（2）識別プロセス

未知のシーン画像を学習済み GoogLeNet に入力し，画像の特徴を抽出する．識別器生成プロセスにおいて生成した SVM 識別器にその特徴量を入力し，シーン認識結果を得る．なお，この識別に使用したプログラムを付録に示す.

3.3.1　特徴抽出器として利用する学習済みＣＮＮ

主に物体カテゴリ認識に使われるディープラーニングとして発展してきた CNN は図 3.2 に示すような多階層構造をしており，霊長類の脳の高次視覚野と似た振る舞いを示すことが知られている．

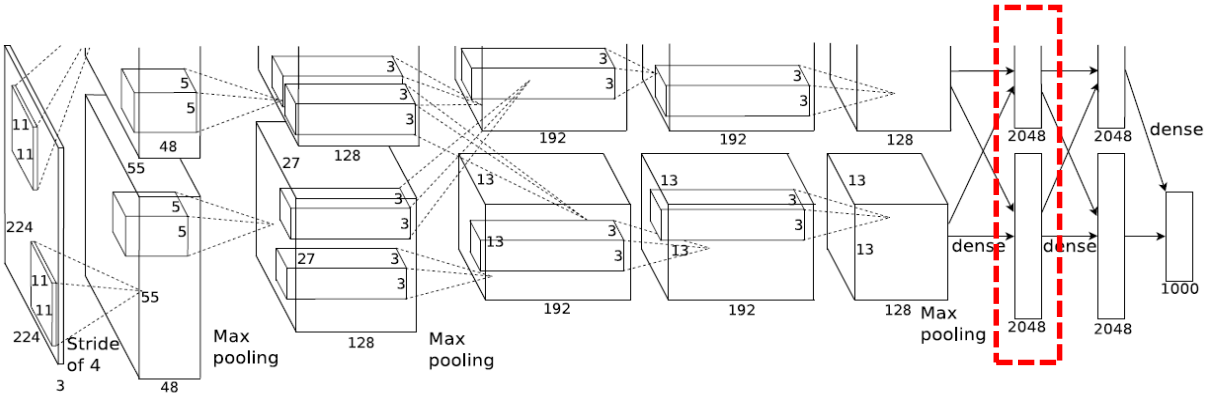


図 3.2　CNN の例：(AlexNet [20]より)

図 3.2 は ILSVRC 2012 で優勝した AlexNet である[20]．AlexNet は，畳み込み(convolution)層・プーリング(pooling)層それぞれ 5 層，全結合(fully-connected)層 3 層の構造で，サイズが比較的小さく，基準手法としてよく使われていた．CNN の各層の役割は，畳み込み層では，入力画像（前層）とフィルタで積和演算を行う．フィルタの係数の値は学習によって得られる．プーリング層では，畳み込みの結果出力される特徴の位置感度を若干低下させ，画像内での位置が変化してもプーリング層の出力が不変になるようにする．フル結合層は，畳み込み層・プーリング層で抽出された値をカテゴリに分類するための層である．すなわち，最終のプーリング層の出力（4096 次元ベクトル：図 3.2 の点線部の入力値）は画像の特徴量が出力されていると考えることができる．このような特徴は到底人間が設計できものではなく，人間には理解できない数値であるが，その有効性は認識結果（AlexNet の識別精度：正解率 83.6%．CNN 登場前は 74.2%）が物語っている．

最終的な出力結果は 1000 種類の物体カテゴリに分類（多クラス分類）になるので，ソフトマックス関数により 1000 クラスの確率として算出して出力される．上下 2 段となっているのはそれぞれ GPU により処理を高速化するためである．

以上，AlexNet を例に CNN の基本的な構造を概説した．

本研究では，先に述べたとおり，この AlexNet をさらに発展させた GoogLeNet を用いる．GoogLeNet は図 3.3(a)に示すとおり，AlexNet とは構造が大きく異なり，各層は convolution 層，pooling 層，Inception モジュールの組合せで構成されたニューラルネットワークである．特徴的なのは，図 3.3(b)の点線で囲まれた部分の Inception モジュールである．モジュール内で畳み込み演算が連結されており，これにより次元削減と投射の減少による疎性結合を実現している．このように，モジュール内で構造を作り込んだ方が総結合数は減少する．すなわち，1×1，3×3 などの異なる大きさの複数の畳み込みフィルタを並列に用いることで識別性能を向上させるとともに，ネットワーク全体のパラメータ数を減少させている．

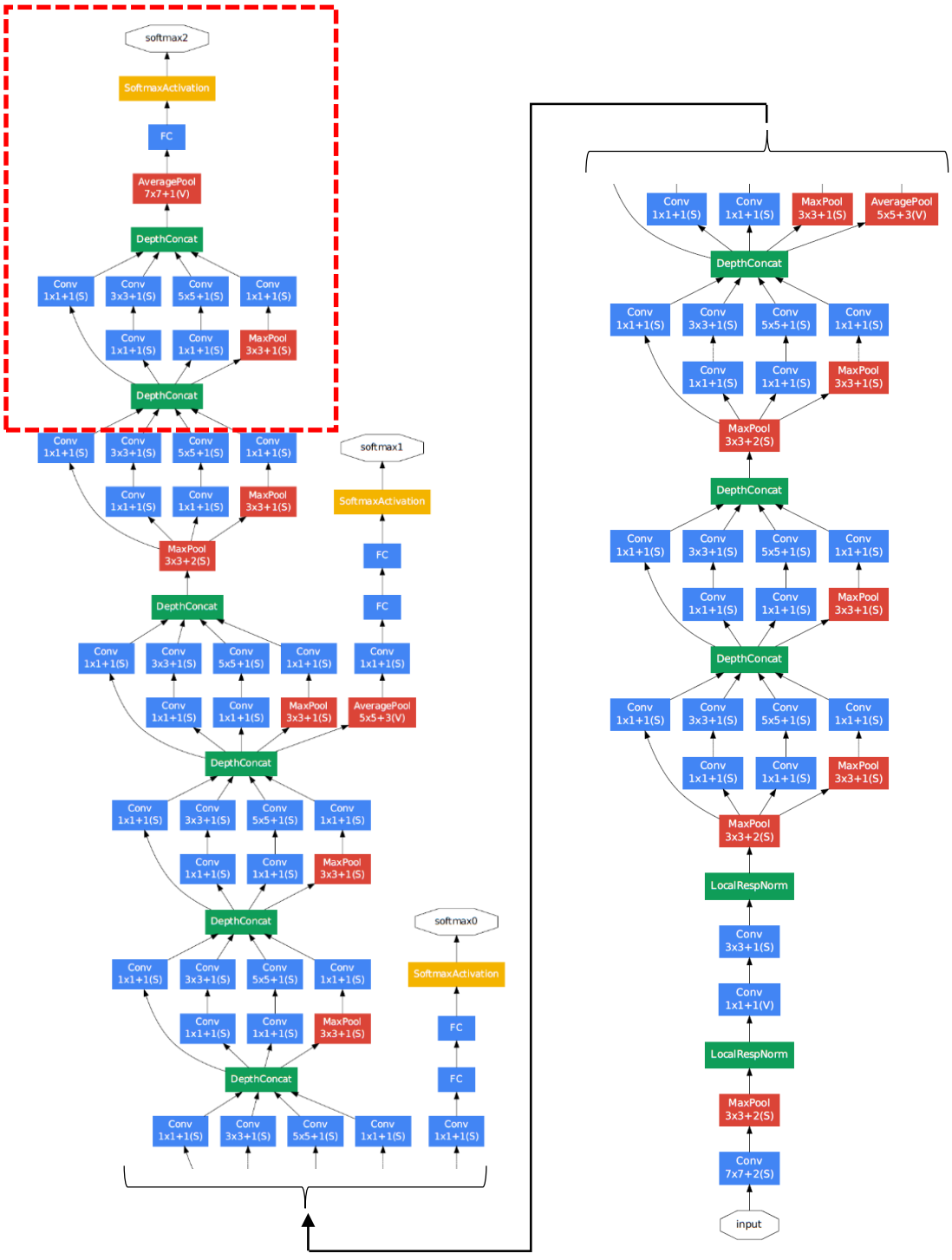


図 3.3(a)　GoogLeNet のネットワーク構造[18]より

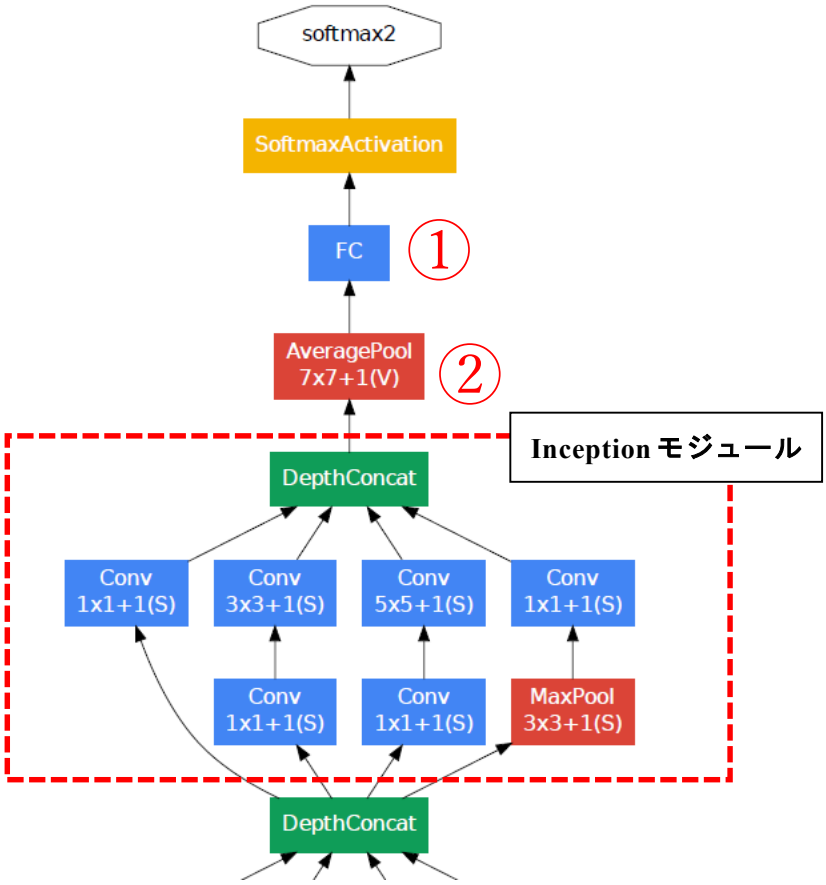


図 3.3(b)　GoogLeNet の最終出力付近の階層
（図 3.3(a)の点線部の拡大）［18］より改変

3.3.2　ＣＮＮから抽出する特徴量

ＣＮＮ（GoogLeNet）から抽出する特徴量について述べる．

図 3.2 の AlexNet のフル結合層にあたるのは GoogLeNet では図 3.3(b)の①の“FC”の階層である．したがって，識別に十分有効な入力画像の特徴量が出力されているのは図 3.3(b)の②の最終プーリング層“AveragePool 7x7+1(V)”の出力であることが分かる．この値を GoogLeNet から抽出する画像の特徴とし，以降，CNN 特徴ベクトルと呼ぶこととする．

CNN 特徴ベクトルは，1024 次元ベクトルで非負の値を取る．これは，ネットワークのユニット（ノード）が持つ活性化関数が正規化線形関数である

ユニット（ReLU：Rectified Linear Unit）であるためである[21]．これをグラフで可視化すると図 3.4 のようになる．

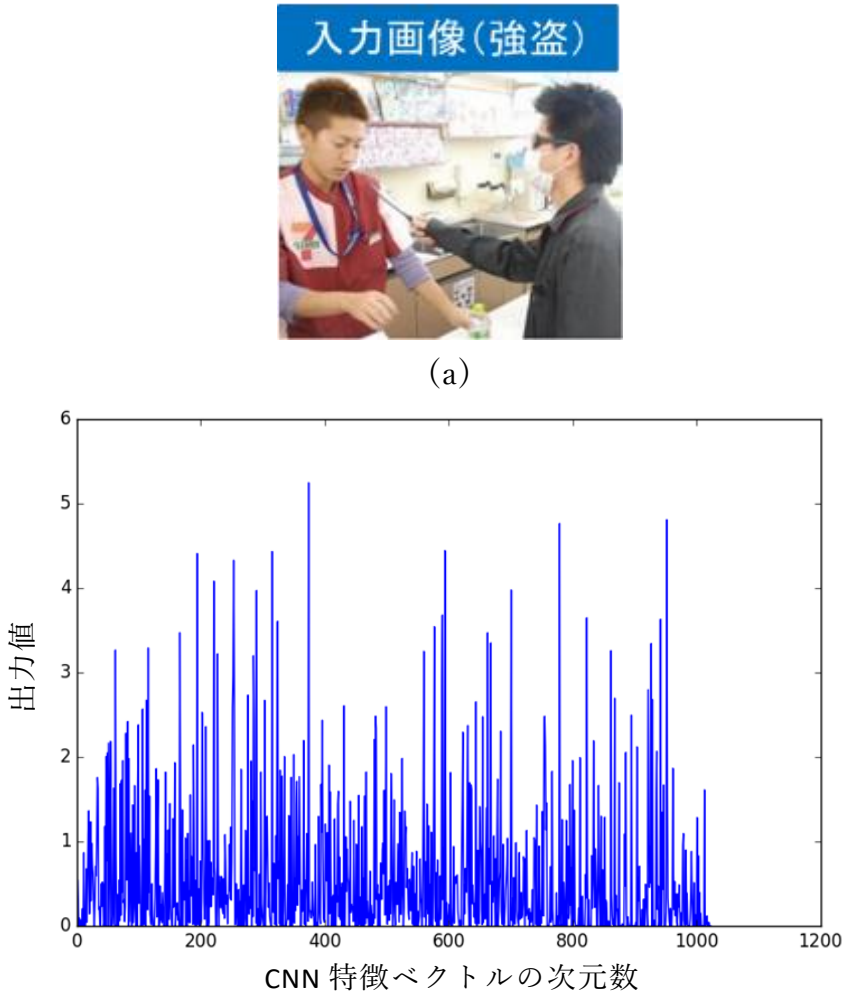


図 3.4　(a)の入力画像の CNN 特徴ベクトルの可視化

3.3.3　静止画を用いた単一シーン認識の評価実験

以下の 2 種類のシーンそれぞれの認識評価実験の結果を示す．なお，各シーンは正例／負例によって構成されているが，シーンの名称はそれらの正例を表す名称としている．また，実験に用いた以下のデータセットは，高安全知能コンピューティングシステムの実現に向け，その応用イメージに適合するシーンの画像である．

（1）データセット

実験 1：　強盗シーン認識

データセットの内容は，コンビニや銀行などの室内という同一環境中で発生した強盗の画像を正例とし，また，同様の環境で強盗が発生していない画像を負例とする．

トレーニングデータ数は，正例が 175 枚，負例が 175 枚である．

実験 2：　急病・転倒シーン認識

データセットの内容は，室内や路上などの環境中で急病や事件・事故等で人が倒れている，あるいは苦しんでいるような画像を正例とし，また，同様の環境中で居眠りやくつろいでいるような，けが・急病等が発生していない画像を負例とする．

トレーニングデータ数は，正例が 200 枚，負例が 200 枚である．

また，これらのデータセットは，著者がインターネットから収集した不特定多数の独立した静止画像である．静止画シーン認識では，様々な店舗等で発生する強盗シーンあるいは様々な場所で発生する急病シーンを認識できなければならないことから，それらの画像は，背景の類似性などにより認識がなされないよう，同一のビデオ映像から一定の比率で切り出した複数枚のフレーム画像は含まれていない．また，特徴抽出器である CNN の仕様に合わせ，画像はカラー画像であり，そのサイズが 224×224 ピクセルになるように縮小・拡大を行った．また，正例/負例の教師信号の付与については著者が行った．図 3.5 に各シーンのトレーニングデータ，テストデータのサンプル画像を示す．

(a)　(b)　(c)　(d)

図 3.5 サンプル画像：(a)強盗シーンの正例, (b)強盗シーンの負例,
(c) 急病・転倒シーンの正例, (d) 急病・転倒シーンの負例

（２）識別手法

一般に画像認識等で使われている以下の識別手法を選定し, 比較実験を行い，SVM の認識精度が最も高くなることを示す.

① ベイジアンネットワーク
② アダブースト
③ サポートベクトルマシン
④ ニューラルネットワーク

なお，本実験で機械学習に使用したツールは，Weka3.7 である[22]. Weka は, ニュージーランドの Waikato 大学でオープンソースソフトウェアとして開発され，公開されているものである.

以下に各識別手法を概説する.

① ベイジアンネットワーク（BayesNet）

BayesNet は，確率変数をノードとし，それらの間の依存関係をアークでグラフィカルに表現した確率モデルである[23]. 依存関係は，アークに付随する条件付確率表で定量的に表現される．BayesNet では，ノードの

値は, その親となるノードの値以外の影響を受けないという条件付独立の仮定を設定する.

この仮定を数式で表すとノードの値 $\{x_1,\cdots,x_n\}$ の同時確率は,

$$P(x_1,\ldots,x_n)=\prod_{i=1}^{n}P(x_i|\mathrm{Parents}(X_i)) \tag{3.1}$$

$\mathrm{Parents}(X_i)$：値 x_i をとるノードの親ノードの値

で表すことができる. その例を図 3.6 に示す.

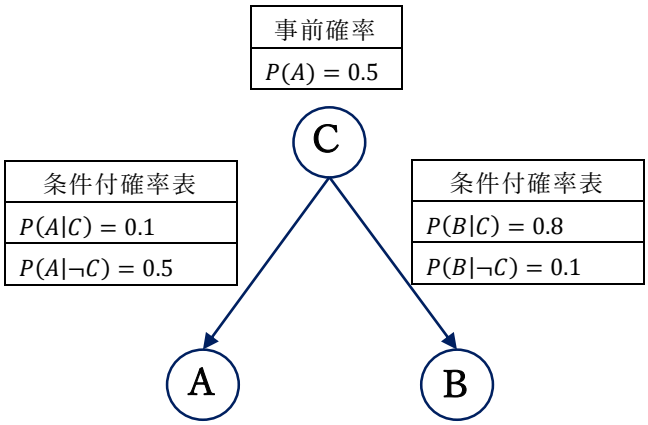


図 3.6　BayesNet の例

なお, アーク探索手法は, アークを新たに作成することにより対数尤度が増加するよう探索的にアークを追加していくための基本的方法である K2 アルゴリズムを用いている[24].

② アダブースト（AdaBoost）

AdaBoost アルゴリズムは, ブースティングの代表的なものである[25].

ブースティングは, 次の手順でデータを識別する.

・学習用の各データに重みを付け, 弱識別器を生成する.

・生成した弱識別器で誤識別されたデータに対してその重みを増やす.

・重みを増やしたデータで異なる弱識別器を n 個, 逐次的に生成する.

・後から作られる弱識別器は，前段の弱識別器が誤ったデータを優先的に識別するようになるので，各弱識別器は相補的な働きをする．

・識別結果は、各弱識別器の性能に応じた重み付き投票により決定する．

図 3.7 に，学習後の AdaBoost による識別の流れを示す．図中の $\boldsymbol{h_i(x)}$ は，学習により生成された弱識別器で，$\boldsymbol{\alpha_i}$ は，各弱識別器の性能に応じた重みである．

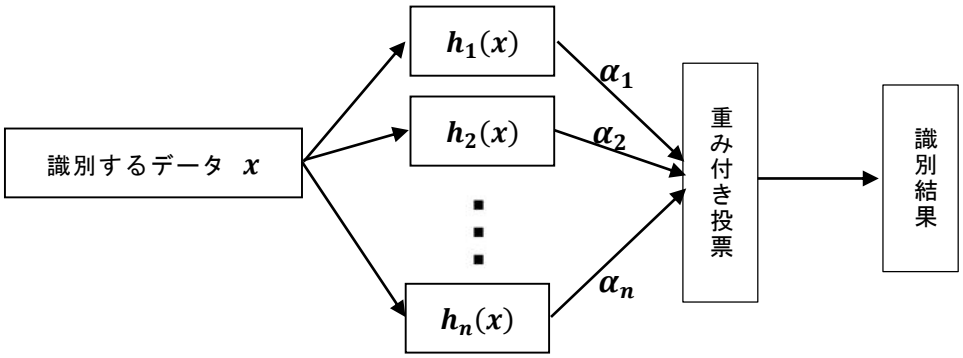


図 3.7　AdaBoost による識別の例

③ サポートベクトルマシン（SVM）

図 3.8 に示すような特徴空間上に，学習データの正例(図中の○)・負例(図中の▲)がそれぞれ分布していると仮定したとき，正例・負例を最も汎用性が高く識別できる境界線（超平面）を求める手法が SVM である[26]．最も汎用性の高い識別境界線とは，境界線に最も近いデータ（サポートベクトル）との距離（マージン），図 3.8 の実線と点線の距離が最大となるような境界線である．すなわち，マージンを最大化する問題である．

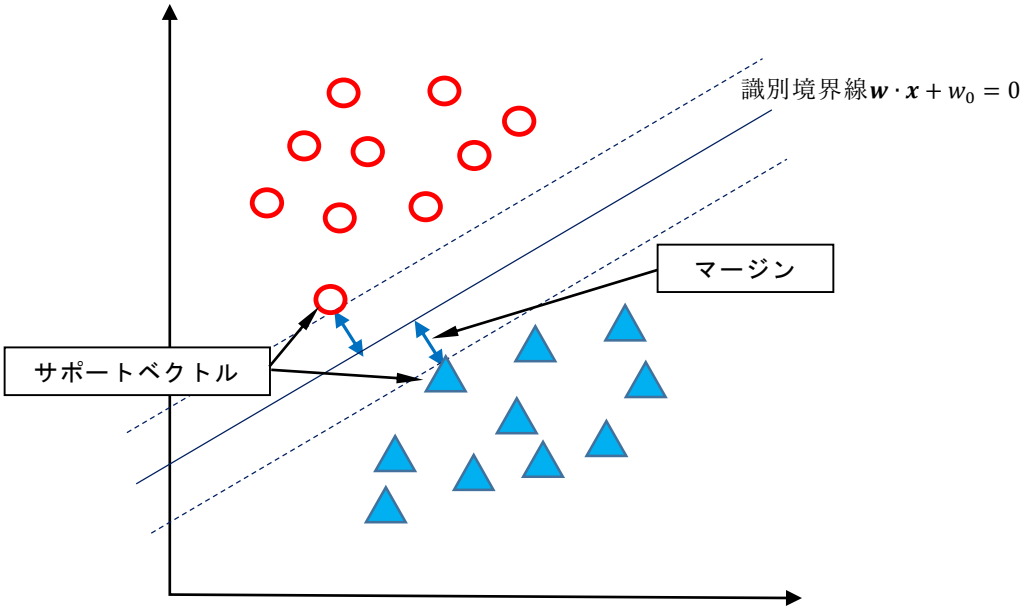


図 3.8　SVM による識別

しかし，このままでは，線形識別面しか求めることができないので，「カーネルトリック」と呼ばれる手法を用い SVM を非線形モデルに対応させる．

一般に非線形関数 ψ を使って，データを元の次元より高次元の特徴空間へ写像するとデータが線形分離できる可能性が高まる．したがって，ψの内積$\langle\psi(x_i),\psi(x_j)\rangle$ を計算できれば，非線形の SVM の学習が可能となるので，これをカーネル関数 $K(\boldsymbol{x},\boldsymbol{x}')$ によって直接定義する手法がカーネルトリックである．

カーネル関数には，多項式カーネル

$$K(\boldsymbol{x},\boldsymbol{x}')=(\boldsymbol{x}^T\boldsymbol{x}'+c)^p \tag{3.2}$$

p：自然数，　$c\geq 0$

やガウシアンカーネル（ＲＢＦカーネル）

$$K(\boldsymbol{x},\boldsymbol{x}')=\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2\sigma^2}\right) \tag{3.3}$$

σ：正の実数

などが，よく用いられる．

なお，SVM には，ソフトマージンにより誤識別をどの程度許容するかを決めるパラメータであるコストパラメータ C とガウシアンカーネルのパラメータ γ（式(2.3) の $1/2\sigma^2$）がある．

Cは，小さいほど誤識別を許容するように，大きいほど誤分類を許容しないように識別境界面を決定し，γ は，値が小さいほど単純な識別境界面となり，大きいほど複雑な識別境界面となる．これらのパラメータは，パラメータの探索空間を格子状（グリッド）に区切り，交点となるパラメータの組み合わせをすべて調べるというグリットサーチと呼ばれる方法で過学習とならないよう最適化する．

④ ニューラルネットワーク（NN）

NN は，生物の神経回路の仕組みを模倣した図 3.9 のような多層構造のネットワークモデルである．

信号が入力層から出力層に向かって流れ，入力層のノード（図 3.9 では，$\{x_1, x_2, x_3\}$）を特徴ベクトルの次元数用意し，それぞれの次元を入力値とする．出力層は，識別対象のクラス数だけノードを用意する．隠れ層は，入力値・出力層の数に応じ適当な数のノードを用意する．このように，ノードを階層状に組むことで，複雑な非線形識別面を実現することができる．

学習は，誤差伝播法によって行われる．誤差伝播法は，出力層の出力とクラスの値（正解情報）との誤差を求め，その誤差を中間層に伝播させて学習を行う手法である．ここでは出力と正解情報の 2 乗誤差が最小になるよう，一般には確率的最急勾配法が用いられ，学習が行われる．

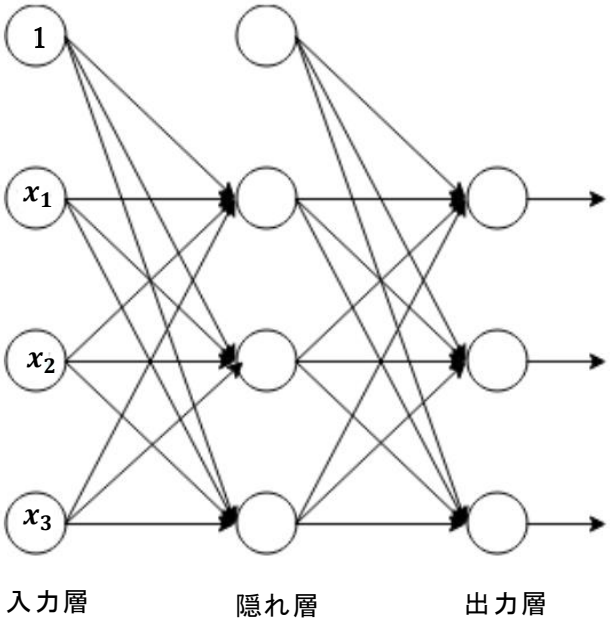


図 3.9　3 層のＮＮの例

なお, 学習の際のパラメータとして, ミニバッチのサイズ, 学習係数, モメンタムなどがあり, また, ネットワークのパラメータ(重み)の更新回数であるエポック数も決める必要がある.

（３）評価方法

評価方法として交差検証法（k－分割交差検証法）を用いる[27]. 以下にその方法について概説する.

交差検証法は, シーン認識モデルが未知の画像データに対して, どれだけの識別能力を有しているかという「汎化能力」を推測するための手法であり, 比較的小規模なトレーニングデータセットでも, 実データに対する識別能力に近い結果を示す.

交差検証法は, 図 3.10 のとおり, トレーニングデータセットを非復元抽出によりランダムに k 個に分割する. そのうちの $k-1$ 個を識別器の学習に使用し, 1 個をテストに使用する. この手順を k 回繰り返すことで,

k個の識別器と識別能力評価を取得する.この識別能力評価の平均値をとり，モデルの推定汎化能力とするものである.

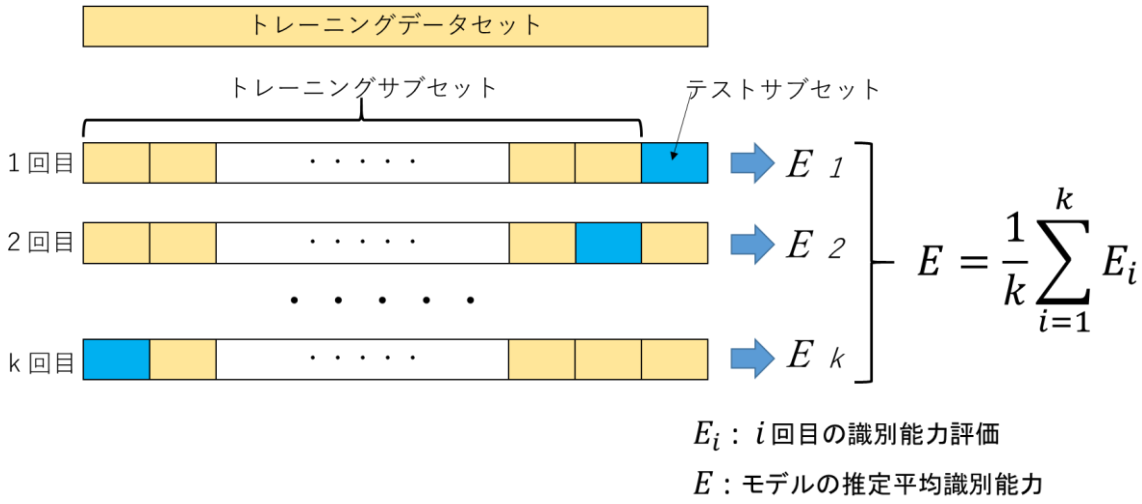


図 3.10　k－分割交差検証法

本実験では，トレーニングデータセットが比較的小規模であることから，10 分割で交差検証を行った.

（4）識別能力評価の指標

識別能力評価の指標は，データセットの正例のデータ数と負例のデータ数が同一であることから正解率(Accuracy)を用いることが最も妥当である．正解率は，全データ数に対する，正例/負例が正しく識別されたデータ数の割合であり，表 2.1 に示す混同行列(confusion matrix)から，式(3.4)により求めることができる.

本実験での識別能力評価結果は，この正解率により示すこととする.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{FP}} \tag{3.4}$$

表 3.1　混同行列

	識別結果が正例	識別結果が負例
正例	正解数(TP)	正例を負例と誤識別した数(FN)
負例	負例を正例と誤識別した数(FP)	正解数(TN)

（5）評価実験結果

評価実験結果は，図 3.11 及び図 3.12 のとおりである．

実験 1：強盗シーン

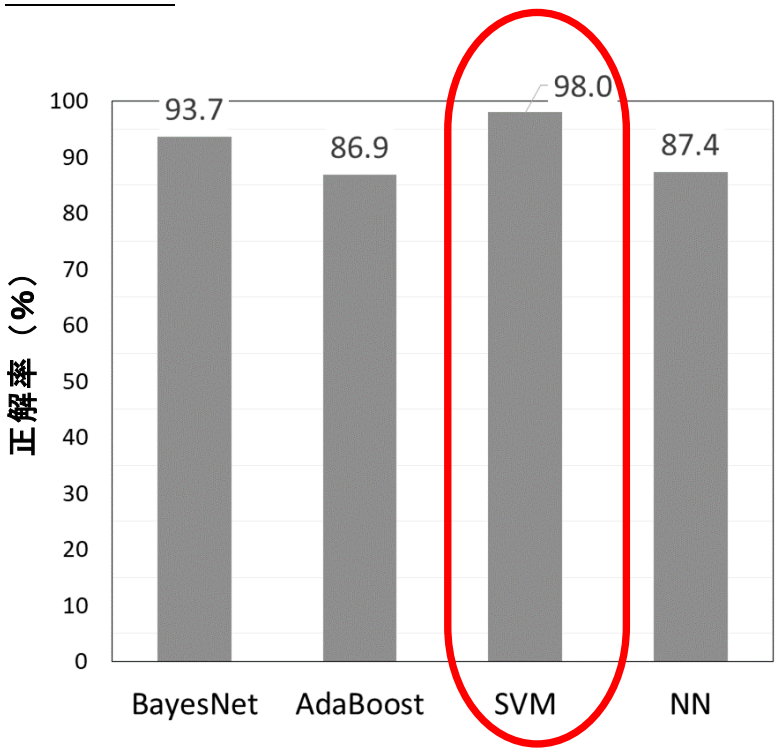


図 3.11　強盗シーン認識の評価実験結果

実験 2：急病・転倒シーン

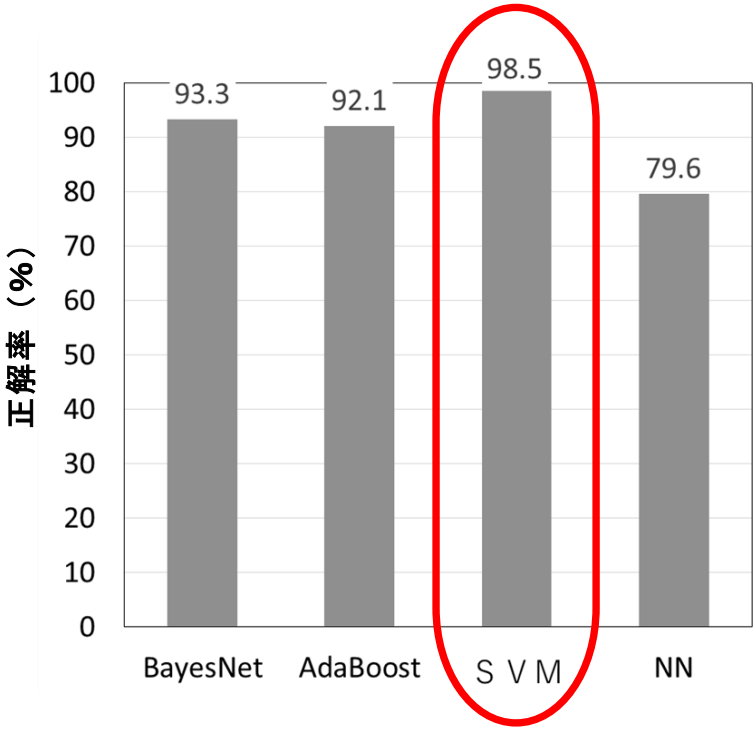


図 3.12　急病・転倒シーン認識の評価実験結果

いずれのシーンにおいても，ＳＶＭが 98%以上の高い正解率を示した．

このように，物体カテゴリ認識のために学習済の GoogLeNet を用い抽出した画像の CNN 特徴ベクトルを SVM で学習し生成した識別器によって，静止画を用いた単一シーン認識を高精度に実現できることが明らかとなった．

3.4　静止画を用いた複数種類のシーンの認識

前節では CNN とＳＶＭの組合せにより，特定の環境・状態下における単一シーンを静止画から高精度に認識が可能であることを明らかにした．しかしながら，この手法では，複数種類の危険事象が発生する可能性のある空間

における複数種類のシーンの認識（以下，多クラス分類と呼ぶ）を実現できない．

ある 1 種類の危険事象の発生を検出することはもとより，複数種類の危険事象の発生を同一の高安全知能コンピューティングシステムにより認識できれば，汎用性やコストの観点から有用である．このようなシステムが導入されたコンビニエンスストアや銀行においては，強盗や暴力行為などの発生が認識された場合，警察などへの通報や客の避難誘導などを支援し，負傷者や病人が発生した場合には，消防への通報など，適切な対応が可能となる．また，本システムの装置を携帯し移動する場合，様々な環境で発生する複数の危険事象（強盗，通り魔など）を検出し，ユーザへの注意喚起，関係機関への通報，危険回避行動等により支援することができる．

3.4.1 感情価に基づき定義されたデータセットの生成

多クラス分類を実現するための手法を検討するにあたり，改めてシーン認識の対象とするシーンについて定義し，新しいデータセット（以下，その他シーンと呼ぶ）を生成する．

後段で詳述するが，シーン認識の対象シーンの定義にあたっては，危険状態に対する人間の直感的な認知プロセスには，恐怖や不安等の感情に関わる脳内の偏桃体などが関与しているという心理学，脳科学などの知見が示されており，危険な場面に対して人間は恐怖心等の感情を生起させる，ということに着目した．

（1）感情価を考慮した危険シーンの定義

心理学における感情の分類には，基本感情に基づくカテゴリ分類，その中でも「喜び」，「驚き」，「恐れ」，「怒り」，「嫌悪」，「悲しみ」の 6 大感情と呼ばれるものに感情を分類しようとする考え方がある．しかしながら，このよ

うな基本感情により感情の全てをカテゴリに分類し，感情がこれらのうちいずれかであると決定することは難しい．

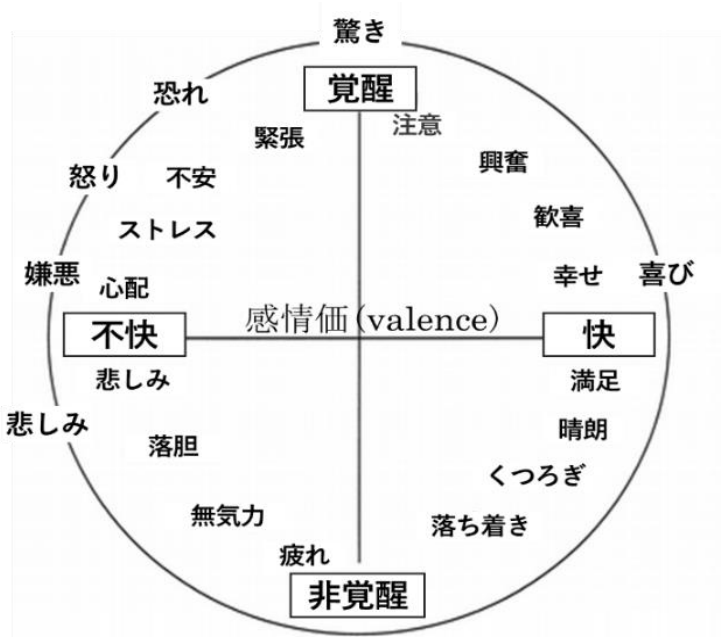


図 3.13　Russell の感情円環モデル[28]

それに対し，図 3.13 に示す Russell の提唱した円環モデルにより感情を理解しようという考え方がある[28]．このモデルでは，快－不快／覚醒－非覚醒の 2 次元上に感情が定義できるとされており，快－不快の軸には感情価(valence)，覚醒－非覚醒の軸には覚醒度(arousal)をとっている．この考え方は広く受け入れられており，本論文でも，このモデルに従って検討を進める．

ここで，快－不快の軸の「感情価」に着目すると，感情を単純化（2 値化）できる．例えば，恐れ，不安，嫌悪，心配などに類する感情は，感情価の軸に射影すれば「不快」となる．同様に幸せやくつろぎは「快」となり，感情を二元論的に扱うことができ 2 値分類器である SVM との親和性が高いと言える．

以上に基づき，これまで「危険事象が発生しているシーン」と述べてきたシーン認識の対象を拡張し，改めて，感情価を考慮した「危険シーン」として次のように定義する．

危険シーンとは，
危険因子が存在し かつ 不快を生起するシーン

ここで危険因子とは，画像により表現される一般に危険性を有する物，動作及び状態などを意味し．例えば，刃物，銃，暴力，人が倒れる，苦しんでいる，などである．
これにより，危険でないシーンすなわち非危険シーンは，
危険因子が存在しない または 不快を生起しないシーン

と定義できる．

このように危険シーンを定義することにより，危険因子が存在する画像に対して抱く感情価に基づいて危険シーンを認識できるようになる．例えば，感情価を考慮せず，危険因子の有無のみにより危険シーンを認識しようとした場合には，包丁自体は危険性を有する物（危険因子）であり画像にそれが写っていれば危険シーンと認識しなければならない．しかしながら，包丁が登場する料理の場面では，人間はそれを危険因子としては認知しないので，非危険シーンであると認識できなければならない．このような包丁の危険性に対する人間の認知の変化は，物自体の危険性は変化していないので，刃物に対する恐怖心（不快）が消失しているためであると考えることができる．また，感情価のみによって危険シーンを認識しようとした場合，授業中に学生があくびをしている危険性のないシーンに対し不快感を抱いたとすると，

危険シーンと認識しなければならなくなる．しかし，定義に従えば，危険因子が存在していないので，非危険シーンであると認識できる．

このように，危険シーンの認識は，危険因子の有無のみではなく感情価に依存しているおり，危険シーンの定義に基づき構築したトレーニングデータセットにより生成される識別器は，危険因子の有無のみに影響されるのではなく感情価に基づき危険／非危険シーンの認識が可能となる．

また，この危険シーンの定義にしたがえば，これまで用いてきた強盗シーンは，危険因子として凶器あるいは店員を脅している様子などが存在し，恐怖などの不快感を抱くシーンである．急病シーンについても，同様に危険因子として人が倒れているあるいは苦しんでいる様子などが存在し，不安などの不快感を抱くシーンであるということができる．このように強盗シーン等についても，ある特定の危険／非危険シーンとして，識別が可能となる．

（2）その他シーンデータセット

上述の危険シーンの定義に従いデータセットを生成した．なお，このデータセットには，強盗，急病・転倒シーンを含まない．トレーニングデータ数は，正例が 300 枚，負例が 300 枚である．

その他シーンの正例を詳述すると，火災，自然災害（家屋倒壊，津波被害など），交通事故，暴力行為，テロ，戦闘，高所，廃墟，危険生物などの様々な危険シーンである．これらの正例画像は，種類が異なるシーンであっても，共通に恐怖，不安，嫌悪といった感情を抱かせる画像である．また，負例については，家族団らん，学校・家屋などの部屋の様子，雑踏，観光地，スポーツなど，まさに恐怖や不安など感じさせない様々な日常的なシーンの画像である．

このように危険シーンの定義に従い様々なシーンにより構築されたその他シーンのデータセットである．このデータセットにより生成される識別器は，

汎用的に人間がその画像に抱く感情価を考慮してシーンを識別する特徴を有していると考えられる.

図 3.14 にその他シーンの画像の例を示す.

(a) (b) (c)

(d) (e) (f)

図 3.14　その他シーン：(a)(b)(c)正例画像の例，(d)(e)(f)負例画像の例

（3）データセットの主観評価実験

以上の危険/非危険シーンの定義に従い， 3.3 節で述べたとおり，著者が各シーンのデータに教師信号を付与した．しかしながら，この場合，著者が不快を感じることにより危険シーンと判断した画像とそうでない非危険シーンの画像に対して一般的な人物が,直感的に同様の感情価を生起するか否か,その感情価の一般性と整合性について検証する必要があると考えた.

そこで,第三者に各シーンの画像を提示し,不快を感じるか否かの単極尺度による主観評価実験を行った.

実験に用いた評価画像は，強盗，急病，その他の 3 シーンのデータセットの危険／非危険シーン画像をそれぞれからランダムに計 140 枚を抽出した．実験手順は，ディスプレイに提示された評価画像に対し，被験者が不快を感じた場合「はい」，そうでなければ「いいえ」と回答するよう求めた．被験者は，20 代から 60 代の 11 名（男 4 名，女 7 名）である．

実験結果を表 3.2 に示す．被験者の全回答数(1540 件(＝140 枚×11 人))のうち，危険シーンに対して「不快」，非危険シーンに対して「不快でない」と回答した割合は 97.5%であった．なお，表中の正解クラスは，前述のとおり，それぞれ著者が定義に基づき判断したものである．

以上のことから，危険／非危険シーン画像に対する感情価には一般性があることが示され，また，著者が各画像に付与した教師信号の整合性も確かめられた．

表 3.2　危険／非危険シーン画像に対する感情価の主観評価

		被験者の回答数	
正解クラス	画像 枚数	不　快	不快でない
危険シーン	70	750	20
非危険シーン	70	18	752

（4）その他シーン認識の評価実験

その他シーンを前節で述べた単一シーンとして扱い，これまでの実験結果と同様に高精度の静止画シーン認識が可能であることを示す．実験条件，実験結果は以下のとおりである．

○ SVM のパラメータ設定

①カーネルトリック：多項式カーネル関数を使用する．

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}')^2 \tag{3.5}$$

②コストパラメータ：C ＝1.0

○ 評価方法　トレーニングデータセットを用いた評価は，10 分割の交差検証法を用いる．

○ 実験結果　正解率 99.2%（トレーニングデータセット）

表 3.3 トレーニングデータセットを用いた実験結果の混同行列

	識別結果(正例)	識別結果(負例)
正例	299	4
負例	1	299

表 3.3 に示すとおり，これまでと同様の高い正解率となった．

3.4.2　識別器選択方式による多クラス分類

単一シーンだけの認識手法では，複数種類の危険シーンの認識である多クラス分類を実現できない．そこでこれを可能とする手法として，識別器選択方式を提案する．

提案手法は，入力画像とトレーニングデータセット画像の類似性を用い，あらかじめシーン毎に生成された識別器の中から入力画像の認識のために最適な識別器を選択することにより，　単一シーン認識と同等の精度で多クラス分類を実現する．

一般に SVM による多クラス分類の手法としては，1 対 1 方式（one-vs-one）と呼ばれる手法が用いられており，機械学習のさまざまなソフトウェアやライブラリの SVM の多クラス分類においてもこの手法が用いられている．

one-vs-oneは，n個のクラスからクラスc_i及びc_jに属するデータのみを取り出し，その2クラスを識別するSVM識別器を生成する．したがって，すべてのクラスの対を考えた場合，$(n-1)n/2$ 個の識別器が生成されることとなる．そして，あるデータ$\boldsymbol{x}$が入力されたとき，どのクラスに識別するかは，$(n-1)n/2$ 個の識別器の識別結果による投票（多数決）により決定するというものである．

このため，one-vs-oneにおいては，ターゲットとなるシーンとは無関係の環境のシーンによって生成される多数の識別器の影響により，大きく認識精度が低下するという問題が生じる．しかしながら，提案手法においては，ターゲットとなるシーンのトレーニングデータセットにより生成された識別器のみを選択することが可能であるため，単一シーンの認識と同等の認識精度を得ることができる．

図3.15に識別器選択方式による多クラス分類の手法を示す．

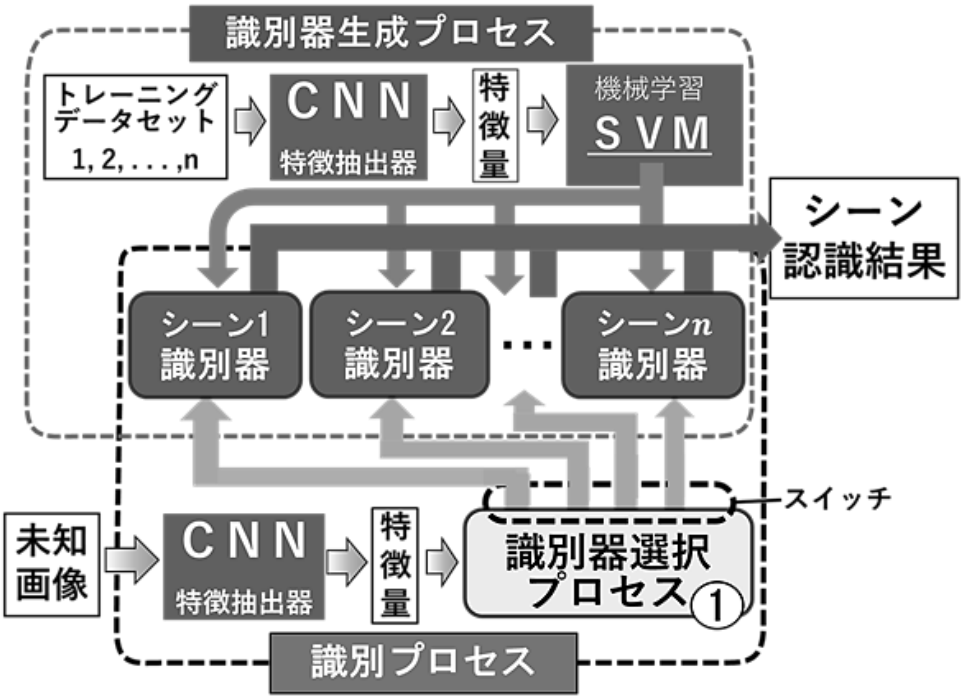


図3.15　識別器選択方式による多クラス分類

この識別器選択方式による多クラス分類と図 3 .15 に示した手法による単一シーン認識との違いは, 識別器生成プロセスにおけるシーン別の識別器の生成と識別プロセスにおける「識別器選択プロセス」(図 3.15 の①)である. 以下に各プロセスについて述べる.

（1）識別器生成プロセス

シーン 1, シーン 2, . . . , シーンn 毎に整備された各トレーニングデータセットを用いてシーン別に予めn個の SVM 識別器を生成しておく. これらの識別器は one-vs-one の場合とは異なりそれぞれカーネルとハイパーパラメータの設定を最適化した状態で生成でき, また, one-vs-one では未知の入力画像を誤認識するような識別器が多数生成されるが, これが不要となり認識精度が向上するとともに, 結果的に学習及び識別に必要な計算量も大幅に軽減されることとなる.

（2）識別プロセス

「識別器選択プロセス」により, 入力画像がどのシーンに属する画像であるかを推定し, 入力画像を識別するために最適な識別器を選択する. それにより選択された識別器による認識結果のみが入力画像のクラスと決定される. この方法により高精度の多クラス分類を実現する。

ここで重要となるのが「識別器選択プロセス」である. 着目したのは入力画像の CNN 特徴ベクトルと各シーンのトレーニングデータの CNN 特徴ベクトルの類似性である. 双方のデータの CNN 特徴ベクトルが類似していれば, 入力画像と類似度の高いトレーニングデータセットのシーンに帰属する画像であると推定でき, 当該トレーニングデータセットにより生成された識別器が最適な識別器であると選択することが可能となる.

したがって, CNN 特徴ベクトルの類似性をどのようにして決定するのかを検討しなければならない. 本研究では, 入力画像とトレーニングデータの

特徴ベクトルの類似性をコサイン類似度により求め, これにより最適な識別器を選択することとする.

コサイン類似度による識別器選択は次のように行う.

入力画像の CNN 特徴ベクトルを$\boldsymbol{x}$とし，すべてのトレーニングデータ数をN, トレーニングデータの CNN 特徴ベクトルを$\boldsymbol{y}_n$ $(n = 1, 2, \dots, N)$とする. Dは CNN 特徴ベクトルの次元数である.

入力画像と全てのトレーニングデータとの間のコサイン類似度を式(3.6)によりN個求める.

$$\cos(\boldsymbol{x}, \boldsymbol{y}_n) = \frac{\boldsymbol{x}}{|\boldsymbol{x}|} \cdot \frac{\boldsymbol{y}_n}{|\boldsymbol{y}_n|} = \frac{\sum_{d=1}^{D} x_d y_{n,d}}{\sqrt{\sum_{d=1}^{D} x_d^2} \sqrt{\sum_{d=1}^{D} y_{n,d}^2}} \tag{3.6}$$

なお，コサイン類似度の値は，CNN 特徴ベクトルの要素が非負であり，ベクトルの大きさで正規化されるので，$0 \leq \cos(\boldsymbol{x}, \boldsymbol{y}_n) \leq 1$の値をとる.

次に，全てのシーン数をMとし，式(3.1)により求めたN個の類似度の値が高い上位 k個（任意，20 個程度）を選択する．そこから，シーン$\boldsymbol{i}$の出現頻度h_iを求め，この h_i の集合を $\{h_i\}_{i=1}^{M}$ とすると，式(3.7)により，それが最大となるシーン番号 $S_{\max}$ を求めることができる.

$$S_{\max} = \underset{i}{\operatorname{argmax}} \ \ h_i \tag{3.7}$$

これにより, 入力画像が帰属するシーンを推測することで最適な識別器を選択することができる.

このプロセスを図 3.16 に例示する．図ではシーン 2 の識別器が選択される．なお，この cos 類似度の他に，クラスタリング（EM アルゴリズム），ユークリッド距離，重み付き Jaccard 係数なども類似性の指標となるが，実験による評価で cos 類似度が最も適している.

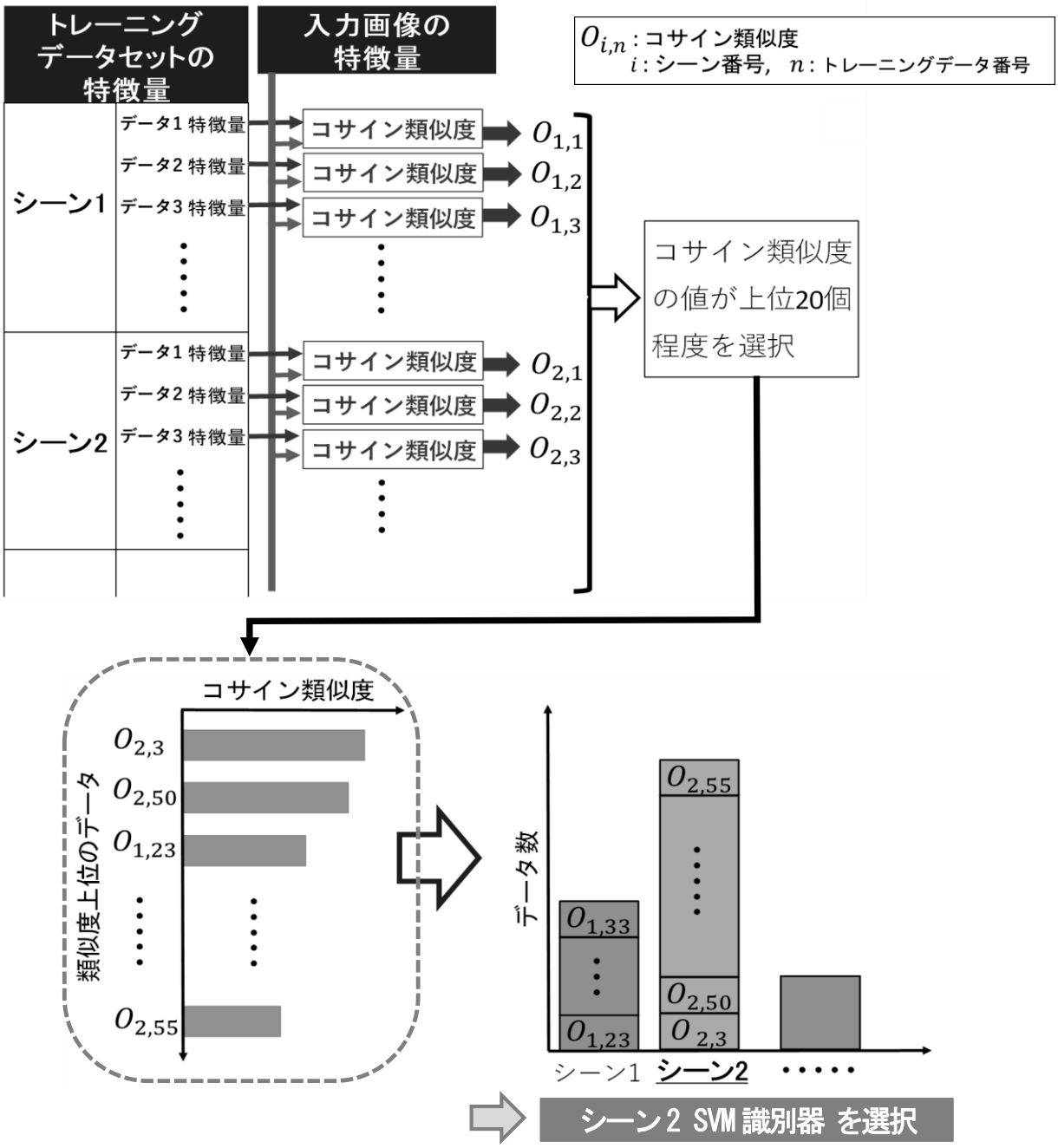


図 3.16　識別器選択プロセスでの処理の流れ

3.4.3　識別器選択方式による多クラス分類の評価実験

評価実験により，高精度の複数シーンの認識が可能となることを示す.

評価実験にあたっては,「強盗シーン」,「急病シーン」及び「その他シーン」を用いる．これにより，実験はそれぞれのシーンの正例／負例を識別するので 6 クラス分類となる.

（1）各シーンのテストデータの導入及びそれらの単一シーン認識実験

多クラス分類の評価実験を行うにあたり，強盗，急病・転倒及びその他シーンのテストデータセット，200 枚（正例 100，負例 100）を導入する．テストデータセットのデータ数の内訳は，次のとおりである.

① 強盗シーン　　50 枚（正例 25，負例 25）

② 急病シーン　　80 枚（正例 40，負例 40）

③ その他シーン　　70 枚（正例 35，負例 35）

また，各シーンそれぞれのトレーニングデータセットにより生成した識別器がどの程度の識別能力を有するのかを確認するため，各シーンのテストデータセットにより単一シーン認識実験を行った結果を以下に示す.

① 強盗シーンのテストデータセットの識別結果

正解率　96%（詳細は表 3.4 のとおり）

表 3.4 強盗シーンテストデータセット認識実験結果の混同行列

	識別結果(正例)	識別結果(負例)
正例	23	2
負例	0	25

② 急病シーンのテストデータセットの識別結果

正解率　95%（詳細は表 3.5 のとおり）

表 3.5 急病シーンテストデータセット認識実験結果の混同行列

	識別結果(正例)	識別結果(負例)
正例	38	2
負例	2	38

③ その他シーンのテストデータセット識別結果

・正解率 97.1%（詳細は表 3.6 のとおり）

表 3.6 その他シーンテストデータセット認識実験結果の混同行列

	識別結果(正例)	識別結果(負例)
正例	34	1
負例	1	34

いずれも，これまでの実験結果と同様の高い正解率である．

（2）テストデータセットを用いた多クラス分類の評価

テストデータセットを用いた識別器選択方式による多クラス分類の評価実験を行うにあたり，強盗，急病・転倒及びその他シーンのそれぞれの識別器はこれまでの実験で生成した識別器を使用する．表 3.7 に実験結果を示す．

表 3.7 識別器選択方式による多クラス分類の評価実験結果の混同行列

			認識結果						正解率
			強盗シーン		急病シーン		その他シーン		(シーン別)
			(正例)	(負例)	(正例)	(負例)	(正例)	(負例)	[%]
正解クラス	強盗シーン	正例	21	2	1	0	1	0	92.0
		負例	0	25	0	0	0	0	
	急病シーン	正例	0	0	35	3	2	0	92.5
		負例	0	0	0	39	1	0	
	その他シーン	正例	0	0	1	0	33	1	94.3
		負例	0	0	0	0	2	33	

表中の赤丸で示した箇所に若干の識別器選択の誤りによる誤認識が見られるが，識別器選択プロセスにおいて，概ねテストデータに対して最適な識別器が選択されており，これにより，単一シーン認識と同等の高い識別精度が維持されていることが分かる．

次に，一般に SVM による多クラス分類の手法である one-vs-one による評価実験結果を表 3.8 に示す．

表 3.8　SVM の one-vs-one による多クラス分類の評価実験結果の混同行列

			認識結果						正解率
			強盗シーン		急病シーン		その他シーン		(シーン別)
			(正例)	(負例)	(正例)	(負例)	(正例)	(負例)	[%]
正解クラス	強盗シーン	正例	14	1	1	0	8	1	78.0
		負例	0	25	0	0	0	0	
	急病シーン	正例	0	0	32	2	6	0	85.0
		負例	0	0	0	36	3	1	
	その他シーン	正例	0	0	2	0	32	1	92.9
		負例	0	0	0	0	2	33	

表中の赤丸で示した箇所に，多くの誤認識が生じており，全てのシーンにおいて正解率が低下していることが分かる．特に強盗シーン及び急病シーンではその低下が著しい．

これは，先に述べたone-vs-oneの特有の問題に原因があると考えられる．one-vs-oneでは，学習により全てのクラスの対の組合せを識別する識別器，すなわちクラス数がn個の場合，$(n-1)n/2$ 個の 2 クラス識別器が生成される．したがって本実験においては 6 クラス分類であるため 15 個の識別器が生成されることとなる．テストデータはこの 15 個の識別器による認識結果（クラス）の多数決によってテストデータのクラスが決定されるが，このとき，本来テストデータが帰属しているシーンとは全く無関係のシーンにより生成された識別器の認識結果に影響を受け，テストデータに最適な識別器の認識結果が必ずしも尊重されないという問題が発生するためである．

たとえば，表 3.5 で強盗シーン正例のデータがその他シーン正例のデータに誤認識された 8 個のデータのケースでは，本来当該データが帰属している強盗シーンに対応した強盗シーン識別器で正例と正しく認識されたとしても，他のその他シーン正例を識別するための 5 個の識別器（①その他シーン，②その他正例/強盗正例，③その他正例/強盗負例，④その他正例/急病正例，⑤その他正例/急病負例の各識別器）の多数でその他シーン正例と認識されたため，誤認識が生じた．

以上の結果をまとめると図 3.17 のとおりとなる．

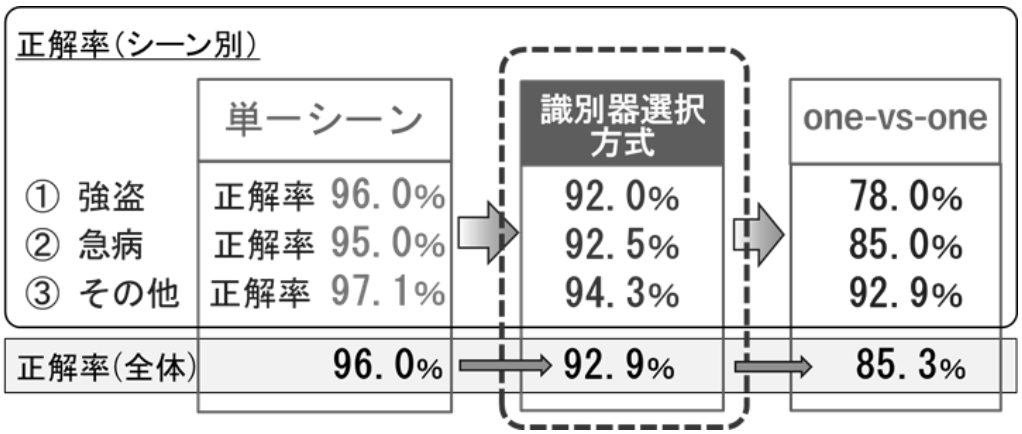


図 3.17　識別器選択方式による多クラス分類と
他の分類（識別）結果との比較

提案手法である識別器選択方式による多クラス分類においては, 単一シーンの認識精度と同等の 93%の正解率となり高精度の多クラス分類が実現された．これに対し one-vs-one では，大きく認識精度が低下することが分かる．

以上の評価実験結果から分かる通り, コサイン類似度を用いた識別器選択は有効に機能したが, コサイン類似度の他にも入力画像の帰属するシーンを推定する手法は存在する．1 つは CNN 特徴ベクトルを座標とみなしデータ間の距離が近いほど類似度が高いとするもの. また, クラスタリングにより, 帰属するシーンを決定する手法もある．しかしながら，コサイン類似度がいずれの手法よりも的確に識別器を選択することができる．これは，各シーンのトレーニングデータの CNN 特徴ベクトルが非常に類似しているケースが多数あるためであると考えられる．図 3.18 に CNN 特徴ベクトルを座標とし，主成分分析（PCA）により 2 次元に次元削減して可視化したものである．

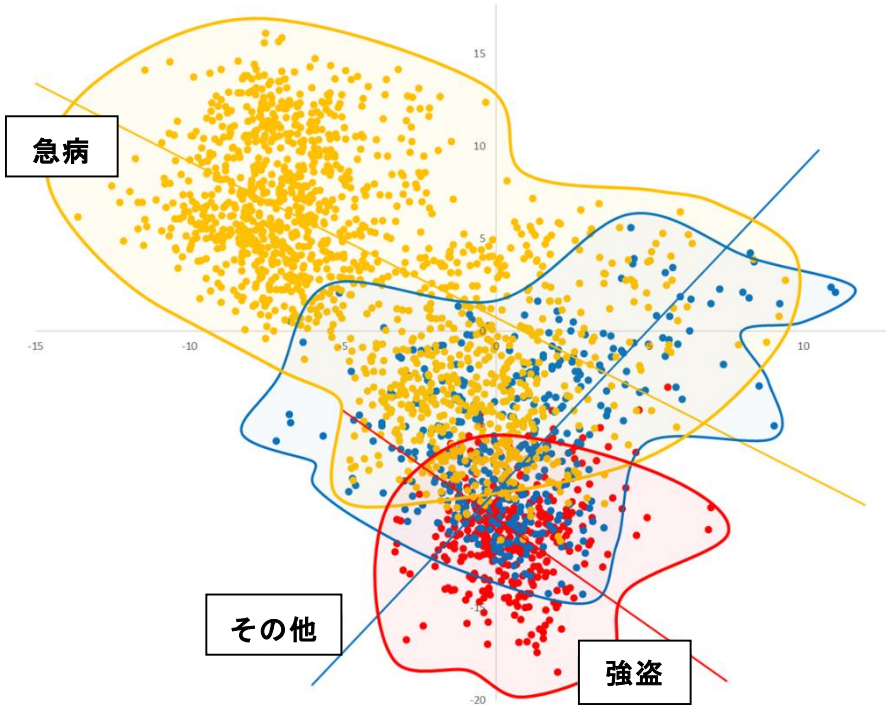


図 3.18　各シーンのトレーニングデータを 2 次元で表現した
場合のデータ分布
(赤：強盗シーン，黄：急病・転倒シーン，青：その他シーン)

図 3.18 から分かる通り，各シーンのデータが重なり合っている部分が見られ，データ間の距離により帰属シーンを類推するのは容易ではないことが明らかである．なお，以下に，識別器選択プロセスにおいてコサイン類似度以外を用いた場合の多クラス分類の結果を示す．

（1）クラスタリングを用いた識別器選択方式による多クラス分類結果

① 強盗シーン　　正解率 78.0%

② 急病シーン　　正解率 88.8%

③ その他シーン　　正解率 50.0%

（２）データ間の距離を用いた識別器選択方式による多クラス分類結果

① 強盗シーン　　正解率 92.0%

② 急病シーン　　正解率 91.3%

③ その他シーン　　正解率 90.0%

データ間の距離を用いた場合の強盗シーンの認識結果を除き, いずれの結果も，コサイン類似度を用いた場合の正解率には及ばない（図 3.17 参照）. これはコサイン類似度が「距離」ではなく，またベクトルの大きさを考慮しない「方向」によるデータ間の類似性を表わしているためではないかと考えられる.

また，コサイン類似度は，式(3.6)に示したとおり，ベクトルの大きさで正規化されており，$0 \leq \cos(\boldsymbol{x}, \boldsymbol{y}_n) \leq 1$の値をとることから，類似度にしきい値設定が可能であるというメリットがある. これにより閾値以下のデータは認識結果の信頼性が低下することから，SVM 識別器では本来不可能な「識別不能」という認識結果を出力することも可能となる. なお, データ間の距離, すなわち「近さ」は，データセットにより相対的に決定されるため，しきい値を設定することは難しいと考えられる.

3.4.4　感情価に基づく多クラス分類

識別器選択方式による多クラス分類手法では, あらかじめ各シーン別に識別器を生成し, 識別器選択プロセスにおいて対応する識別器を選択することにより，単一シーン認識に匹敵する認識精度を実現していた.

しかしながら，この手法では，3 シーン程度の多クラス分類であれば簡単にシーン別の識別器を生成することが可能であるが, 多数のシーンの多クラス分類を行うためには，当然，それと同数の識別器を生成する必要があり, その学習にかかる時間は膨大なものとなる. また，汎化能力の高い識別器を

生成するためには，相応のトレーニングデータ数を整備する必要があり，これも簡単なことではない．そこで，これらの課題を解決するための新たな多クラス分類の手法を提案する．そのためには，まず，3.4.1 で述べた「その他シーン」のトレーニングデータにより生成される識別器が持つ特徴的な識別能力を明らかにする必要がある．

（1）その他シーンの識別器の特徴

前述したとおり，その他シーンは，強盗シーン，急病シーン以外の危険シーンと非危険シーンとによって構成されている．あらためてその他シーンの危険シーン（正例）を詳述すると，交通事故，火災，自然災害（家屋倒壊，津波被害など），テロ，戦争，高所，廃墟，猛獣（に襲われている），危険生物，暴力など，人間に不快感を与える画像である．非危険シーン（負例）は，公園，道路・街路，建造物，雑踏，ペット，スポーツ，室内（ホテル，一般家屋など），家族団らんなど日常的な場面であり，不快感を与えるような画像ではない．したがって，その他シーンは「快/不快」のデータセットでもあり，その他シーン識別器は「快/不快」の感情価の識別器であるとも言える．

そこで，識別器選択方式による多クラス分類の評価実験で生成した，3 シーン（強盗，急病，その他）に対応した識別器を用い，それぞれの識別器が学習していないシーンのテストデータセットを識別させる評価実験を行う．これにより，その他シーンによって生成された識別器が，たとえ未学習の強盗，急病シーンであってもそれらのシーンが危険因子により人に対し不快感を与える（与えない）危険シーン（非危険シーン）の画像であるということを考慮すれば，強盗，急病シーン識別器に比して高い精度でそれらのシーンの識別が可能であることを示す．また，特定の危険シーンの認識のために生成された識別器である強盗，あるいは急病シーンの識別器では，未学習シー

ンの認識は困難であることを示し, その他シーンの識別器の特徴を明らかにする.

実験の種類は，以下のとおりである.

実験 A： 強盗シーンの識別器を用い，急病シーン，その他シーンのテストデータセットの危険シーン／非危険シーンを識別する.

実験 B： 急病シーンの識別器を用い，強盗シーン，その他シーンのテストデータセットの危険シーン／非危険シーンを識別する.

実験 C： その他シーンの識別器を用い，強盗シーン，急病シーンのテストデータセットの危険シーン／非危険シーンを識別する.

実験 A, B については, ある特定の危険事象の発生シーンに対して抱く感情価に基づき生成されたテータセットによる識別器であるため, 学習していないシーンのテストデータの識別は困難であり, 認識結果はコイントスの様な確率的なものとなり正解率は 50%程度になると考えられる．また，実験 C については, 学習していないシーンのテストデータでの認識であるが, その他シーンの識別器は, シーンの種類を問わず感情価に基づき様々なシーンを識別する能力を有していることから，実験 A, B と比較して，高い正解率になると推測される．表 3.9 に実験結果を示す．実験 A,B では，正解率が 50.0%前後であるのに対し，実験 C では 75.0%以上の正解率となった.

表 3.9　未学習シーンのテストデータセットによる各識別の正解率

実験名	識別器名	テストデータセット名	正解率(%)
A	強盗シーン	急病シーン	53.8
		その他シーン	47.1
B	急病シーン	強盗シーン	44.0
		その他シーン	58.6
C	**その他シーン**	**強盗シーン**	**82.0**
		急病シーン	**75.0**

特定の危険事象が発生している危険シーンのみではなく,様々な危険シーンを学習したその他シーン識別器は,学習済みの危険シーンはもとより未学習の危険シーンに対しても 75.0%以上の認識精度を示した.

以上のことから,感情価に基づき構築されたその他シーンの識別器が,汎用性を持った危険シーンの識別器の性質を有していることが明らかとなった.また,このことは,3.3 で述べた論理的推論によらない直感的なシーン認識を可能とする要因となっていると考えられる.

（2）感情価に基づく多クラス分類の手法

その他シーン識別器の特徴を考慮すると,更に多くの危険シーンを学習させることにより,その能力を拡充させた識別器を 1 個だけ用い多クラス分類を実現する手法を図 3.19 に示す．これにより，前述した，識別器選択方式における，シーン毎に識別器を生成する必要があり，また，それらの識別器を生成するための膨大なトレーニングデータを準備しなければならないという，学習コストの課題を解決する.

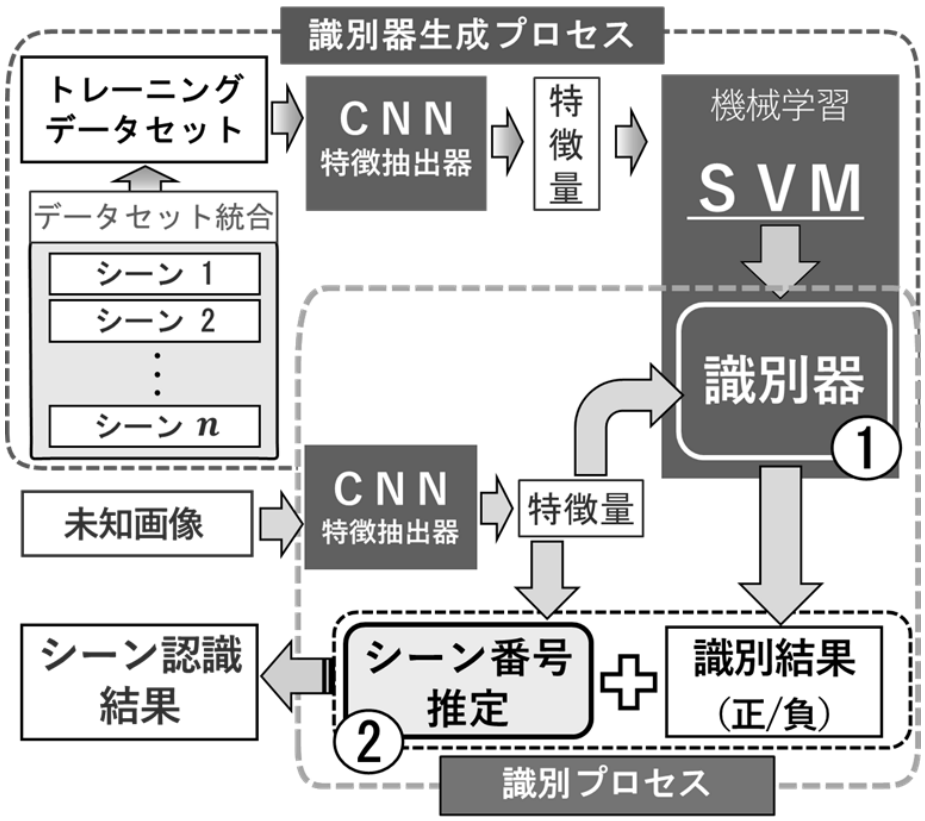


図 3.19 感情価に基づいた多クラス分類

○　識別器生成プロセス

多クラス分類する各危険シーン$(1,2,\ldots,n)$の画像をすべて統合した1つのトレーニングデータセットを作成する．次に，このデータセットをこれまで同様に単一シーン認識で用いた手法により学習し，危険シーンを識別する識別器（図 3.19 ①，以下，危険シーン識別器と呼ぶ）を生成する．

○　識別プロセス

未知画像の CNN 特徴ベクトルを抽出し，識別器生成プロセスで生成した識別器により，危険（正例）／非危険（負例）の識別を行う．したがって，この段階では未知画像が正例／負例であるかのみが識別される．次に，未知画像がどのシーンに属するのかを推定し，最終的なシーン認識結果，すなわ

ち，未知画像がシーンi の危険（危険性のない）シーンであるという認識結果を出力する．なお,「シーン番号推定」のプロセス（図 3.19 ②）においては，未知画像のシーン番号を 3.4.2 で述べた識別器選択プロセスと同様の手法を用い求める．まず，式(3.6)により cos 類似度を求め，式(3.7)によりシーン番号を推定し，未知画像がどのシーンに属するかを決定する．（以下，cos 類似度シーン推定と呼ぶ.）

ここで，cos 類似度シーン推定がどの程度の精度を有しているかを確認するため,これまでの実験で用いた 3 シーンのテストデータセットにより評価を行う．結果を表 4 に示すとおり，正解率 97.5%であった.

表 3.10　cos 類似度シーン推定の混同行列

		シーン推定結果		
		強盗	急病	その他
正解クラス	強盗	48	1	1
	急病	0	78	2
	その他	0	1	69

3.4.5　感情価に基づく多クラス分類の評価実験

（1）トレーニングデータセット及びテストデータセットによる危険シーン識別器の評価実験

危険シーン識別器のみの識別能力を評価する.トレーニングデータセットのデータ数は，これまで使用してきた，強盗シーン（350 枚），急病シーン（400 枚），その他シーン（600 枚）を合わせた 1350 枚とする．内訳は，正例数 675 枚，負例数 675 枚である．テストデータセットのデータ数は，強盗シーン（50 枚），急病シーン（80 枚），その他シーン（70 枚）を合わせた

200 枚とする．内訳は，正例数 100 枚，負例数 100 枚である．評価については，各シーンのトレーニングデータを統合し，データ数が増加したことから 5 分割の交差検証法を用い，また，テストデータセットによる評価も行う．トレーニングデータセットによる 5 分割の交差検証の評価結果は，表 3.11 に示すとおり正解率 96.9%となった．

表 3.11　5 分割交差検証法による識別の混同行列

		識別結果	
		正例	負例
正解クラス	正例	653	22
	負例	20	655

また，テストデータセットによる評価の結果は，表 3.12 に示とおり，正解率 96.0%となった．この結果は，図 3.17 に示した 3 シーンそれぞれのテストデータセットによる単一シーン認識の正解率の平均値と同等である．

表 3.12　テストデータセットによる識別の混同行列

		識別結果	
		正例	負例
正解クラス	正例	94	6
	負例	2	98

（2）トレーニングデータ数を減少させた場合のテストデータセットによる危険シーン識別器の評価実験

トレーニングデータ数によって危険シーン識別器の学習の度合いがどのように変化するのかを検証するため，テストデータセットを用いた実験を行う．

その方法は,特定のシーンのトレーニングデータ数を半減させて学習するというもので，強盗シーンは 350 枚から 176 枚に，急病シーンは 400 枚から 200 枚にデータ数を減少させた．ただし，その他シーンについては，危険シーン識別器の生成にかかる学習の中核となるデータセットであるためデータ数は変更せず 600 枚とし，トレーニングデータの総数は 976 枚である．内訳は正例数 488 枚，負例数 488 枚である．結果は表 3.13 に示すとおり，正解率 94.0%となった．この結果からわかるとおり，強盗，急病シーンのトレーニングデータ数を大幅に減少させたとしても，表 3.12 の識別結果の正解率 96.0%と比較して，さほど識別の精度は低下していない．

表 3.13　強盗，急病シーンのトレーニングデータ数を半減させた場合の実験の混同行列

		識別結果	
		正例	負例
正解クラス	正例	93	7
	負例	5	95

また,このトレーニングデータ数を半減させて生成した危険シーン識別器により，強盗，急病それぞれのテストデータセットを用い単一シーン認識を行った結果を表 3.14 に示す．

表 3.14　危険シーン識別器による単一（強盗，急病）シーン認識の混同行列

			認識結果				正解率
			強盗シーン		急病シーン		
			（正例）	（負例）	（正例）	（負例）	[%]
正解クラス	強盗シーン	正例	24	1			96.0
		負例	1	24			
	急病シーン	正例			36	4	92.5
		負例			2	38	

本来であれば，強盗，急病シーンのトレーニングデータ数を半減させればそれらのシーンの認識精度は大きく低下するはずである．しかしながら，この識別器では図 3.17（シーン別識別器の評価）の結果と比較して，ほとんど認識精度の低下は見られない（強盗シーンの正解率は 96.0%で変化なし，急病シーンの正解率は 95.0%から 92.5%に低下）．

このことは，識別器を生成するトレーニングデータセットを構成する個々のシーンのデータ数が，それらのシーン毎の単一シーン認識の学習に必要なトレーニングデータ数よりはるかに少ない場合であっても，認識精度を維持することが可能であることを示している．したがって，単一シーン認識に比べトレーニングデータ数が少なくて済むため，その収集・整理が容易になるとともに，トレーニングデータのアノテーションや識別器生成にかかるパラメータ調整などに掛かる労力や時間といった学習のコストは，学習データ数に比例することから，そのコストを大幅に軽減可能となる．

なお，参考のため，図 3.20 に単一シーン認識におけるトレーニングデータ数と正解率の関係を示す[7]．図は強盗シーン認識の例である．

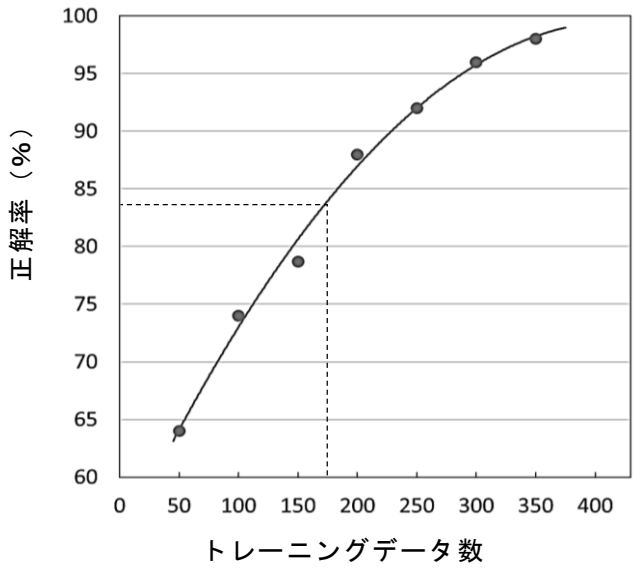


図 3.20　強盗シーンにおけるトレーニングデータ数の増加に伴う正解率の推移 [7].

横軸がトレーニングデータ数, 縦軸が正解率を表している. 図からわかるとおりトレーニングデータ数が 350 枚以上で正解率が 98.0%とほぼ収束しており，十分な学習ができていることを示している．また，図の破線は，データ数が 175 枚付近での正解率を示しており，およそ 84%である．このように単一シーン認識では, トレーニングデータ数が半減することにより, 14%ほど正解率が低下することがわかる. しかしながら, 危険シーン識別器では, 表 3.14 に示したとおり，強盗シーンにかかるトレーニングデータ数を 176 枚に減少させても，正解率が低下していない.

（3）テストデータセットを用いた感情価に基づく多クラス分類手法の評価実験

感情価に基づく多クラス分類の手法により，強盗シーン，急病シーン，その他シーンのそれぞれの正例と負例による 6 クラス分類の実験を行う. 図 3.19 に示すとおり，まず，3.4.5（1）の実験で生成した識別器（表 3.11 参照）

を用い危険シーンの識別を行った後，シーン番号推定を行い，未知画像に対する危険シーン認識の結果を出力する．表 3.15 にその結果を示す．また，比較のため，識別器選択方式による多クラス分類の結果を表 3.16 に示す．

表 3.15　テストデータセットを用いた感情価に基づく多クラス分類の混同行列

			認識結果						正解率（シーン別）[%]
			強盗シーン		急病シーン		その他シーン		
			（正例）	（負例）	（正例）	（負例）	（正例）	（負例）	
正解クラス	強盗シーン	正例	22	1	1	0	1	0	94.0
		負例	0	25	0	0	0	0	
	急病シーン	正例	0	0	34	4	2	0	92.5
		負例	0	0	0	40	0	0	
	その他シーン	正例	0	0	1	0	33	1	94.3
		負例	0	0	0	0	2	33	

表 3.16 テストデータセットを用いた識別器選択方式による多クラス分類の混同行列

			認識結果						正解率（シーン別）[%]
			強盗シーン		急病シーン		その他シーン		
			（正例）	（負例）	（正例）	（負例）	（正例）	（負例）	
正解クラス	強盗シーン	正例	21	2	1	0	1	0	92.0
		負例	0	25	0	0	0	0	
	急病シーン	正例	0	0	35	3	2	0	92.5
		負例	0	0	1	39	0	0	
	その他シーン	正例	0	0	1	0	33	1	94.3
		負例	0	0	0	0	2	33	

感情価に基づく多クラス分類のシーン別の正解率は,強盗シーンが94.0%.急病シーンが 92.5%．その他シーンが 94.3%であった．また，全シーンの正解率の平均が93.5%であった.識別器選択方式による多クラス分類結果と比較すると，強盗シーンで2.0%正解率が向上したが．急病シーン及びその他シーンの正解率に変化はなかった．また，識別器選択方式による多クラス分類の全シーンの正解率の平均では，93.0%であり，感情価に基づく多クラス分類の結果とほぼ同等であった.

以上のことから，多クラス分類においては，認識したいターゲットが多数存在する，あるいは，シーン別に十分なトレーニングデータを準備できない場合は，感情価に基づく多クラス分類手法が有効であり，また，その逆であれば，識別器選択方式も有効であると言える.

（4）感情価に基づく多クラス分類と人間の危険認知プロセスの関連性

このように,様々な危険シーンを学習した危険シーン識別器を用いることにより,論理的推論によらない人間のように直感的な危険シーンの認識というものを実現できることを示した．ここで，この人間の「直感」という情報処理とシーン認識との関連性について明確にしておく必要がある.このことは,人間の危険認知プロセスと感情価に基づく危険シーン認識との類似性を示すことにより明らかとすることができる.

心理学においては，人間のこころ（感情，理性）の働きに関して，迅速に無意識的，自動的な情報処理のプロセスと，遅く意識的，制御的な情報処理のプロセスとを対比する「二重過程理論」という考え方がある．前者の直感的，本能的な働きを「システム 1」，後者の論理的思考，自制心などの働きが「システム 2」と呼ばれている[29]．ここで注目したいのは，システム 1 の働きである.このシステム 1 が機能している状況は，外界の情報を素早く評価し，迅速な行動を取らなければならない場面，例えば，強盗，事故現場，火災などの差し迫った危険から回避行動を取らなければならない場面である.

このような迅速な行動はこころの働きに基づくものであり，当然，脳が関与している．しかし，システム 1 の機能の全体が脳のある特定の領域に属しているというわけではないが，限定的なシステム 1 のプロセスについてはその対応する脳領域として皮質下領域の関与が指摘されている [30]．また，脳科学の知見では，皮質下領域は本能的な欲求や意欲にかかわっており，そのうち側座核は「快」，扁桃体は「不快」の処理との関連が示されている．

このような人間のこころの働きとシーン認識の関連性について見てみると，危険シーン識別器は脳の皮質下領域で行われている本能的な「快－不快」の感情の処理を担い，また，シーン認識全体ではシステム 1 のような直感的な危険事象の認識を行っていると考えることができる．したがって，感情価を考慮した識別という発想を導入したシーン認識の処理と人間の「直感」という情報処理との関係は類比的であると言うことができる．

したがって，このことは，感情価に基づく危険シーン認識により，人間のように直感的に認識される危険事象のみに注意を払うことができ，それ以外の事象は無視できるということである．危険シーン認識を持ちなければ，常にすべての実世界の事象に目を光らせ，危険事象が発生しているか否かを監視するという膨大な情報処理が必要となる．しかし，世界はあまりにも広く多様であり，それを行うことは不可能であろう．

すなわち，危険シーン認識により，人間と同様に実世界の膨大な情報に翻弄されることなく，素早く高精度に危険事象への対応が可能となることを意味している．

3.5 むすび

本章では，CNN の構造等を概説し，様々な危険事象を検出するためのフラットホームとなるシーン認識の評価実験の結果を示した．これにより，物体カテゴリ認識用に学習された CNN である GoogLeNet から抽出した画

像の特徴量（CNN 特徴ベクトル）を SVM で学習し生成した識別器によって, 静止画による単一シーンの認識を高精度に実現できることが明らかとなった.

また，汎用的なシーン認識を可能とするため，複数シーンの認識（多クラス分類）を実現する「識別器選択方式による多クラス分類」の手法を提案した．この手法は，入力画像とトレーニング画像の特徴の類似性に着目し，予めシーン毎に生成した識別器の中から入力画像の識別に最適な識別器を選択することにより，多クラス分類を実現するものである．一般に SVM では．one-vs-one 方式により多クラス分類を行うが，シーン認識においては，その認識精度は大きく低下（正解率 86%）する．これに対し識別器選択方式においては，単一シーンの認識（正解率 96%）に匹敵する精度（正解率 93%）での多クラス分類を実現できることが評価実験により明らかとなり，その有効性が確かめられた．また，入力画像とトレーニング画像の特徴の類似性の指標としてコサイン類似度が最も適当であることが示された.

更に，シーン認識の汎用性を向上させるため，感情価に基づいた多クラス分類の手法を示した．この手法では，識別器の学習にかかる時間と汎化能力の高い識別器を生成するためのトレーニングデータの整備, という識別器選択方式での課題を解決することができる. 感情価に基づく多クラス分類では，様々なシーンに対して抱く快／不快の感情価を学習した識別器を用いることにより，危険／非危険シーンを識別することが可能となり，汎用性が向上することを明らかにした. この識別結果を識別器選択方式の識別器選択プロセスと同様にコサイン類似度を用いることにより, どんな危険シーンであるかを決定することができる. 評価実験における認識精度は，正解率 93.5%となり, 識別器選択方式の正解率 93.0%とほぼ同等の精度を得ることができた.

第４章
動画シーン認識に基づく危険事象検出

4.1 まえがき

前章では，静止画シーン認識を高精度に実現する手法について述べた．しかしながら，人間や車などの物体の動きによりはじめて不審な行動・異常な動作等であると判断可能となる危険シーンでは，静止画によるシーン認識のみでは検知できないという問題が発生する．例えば，「きょろきょろ」と辺りを見回したり，「うろうろ」と歩き回ったり，というような不審な行動，暴力行為，また，危険な走行をしている車両など，様々なシーンが想定される． そこで，静止画シーン認識手法を動画に対応したものへ拡張する必要が生じる．

本章では，動画のフレームと動き情報であるオプティカルフロー情報を重畳した静止画を用いることにより，静止画シーン認識の手法を動画シーン認識に応用可能であることに着目した認識手法を考案し，その有用性を明らかにする[10]．

4.2 動き情報を重畳した静止画の生成

動きのあるシーンを認識するためには動画を用いる．動画は，時系列に連続して変化する静止画を 1 秒間に数 10 枚の速さで切り替えることで視覚の錯覚として静止画が動いているように見えるという現象を利用してい

る．したがって，動画を用いたシーンの認識においても，静止画を用いたシーン認識の手法が応用可能であると考えられる．

しかしながら，それは単純には行かない．例えば，図 4.1 に示したのは自動車が左折している動画の 1 フレームである．

図 4.1　自動車が左折しているシーンの 1 フレームの静止画

この静止画だけからでは，自動車が斜めに「止まっている」のか，「左折」しているのかを認識することは不可能であり，このような場合，物体の移動/静止の動き情報が必要となる．そこで，オプティカルフローを導入する．

4.2.1　オプティカルフローの可視化（RGB 画像化）

オプティカルフローは動画の連続性から得られる隣接フレーム間の画素の変位ベクトルで表現でき，これにより，画素の移動の方向と大きさ（速さ）を知ることができる．オプティカルフローを求める方法には，勾配法の原理による Lucas-Kanade 法[31]や Farneback 法[32]などがあるが，画像中の各画素値を 2 次多項式で近似しその係数をフレーム間で比較することにより高精度なフローベクトル（密なオプティカルフロー）を推定する Farneback 法を用いる．また，これにより得られたオプティカルフローは，図 4.2 に示す

HSV 色空間で表現することができ，その方向（角度）を色相 H で，速さは HSV の明度 V で表現することができる．

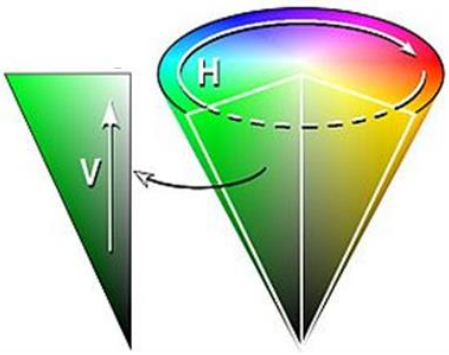
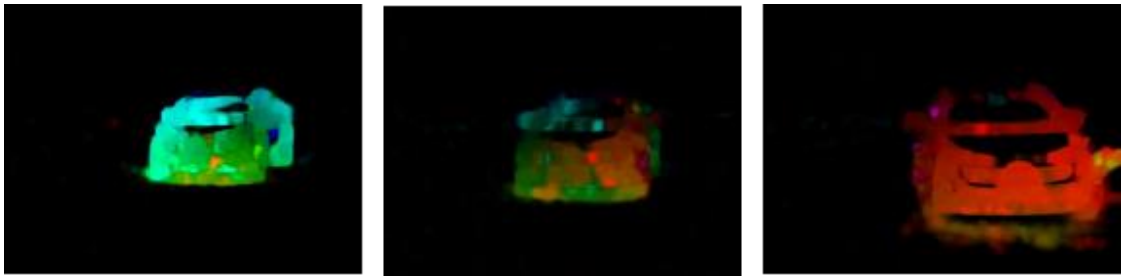


図 4.2　HSV 色空間の円柱

これにより，図 4.3(a)，(b)，(c)に例示すとおり，物体の動き情報は RGB 画像で表現可能となる．



(a)　(b)　(c)

図 4.3　左折してくる車両のオプティカルフローの可視化（図 4.2 を参照）(a)は車両がカメラの斜め前方から直進してくるフレーム，(b)は左折を開始したフレーム，左折のため減速し左方向へ若干移動していることが分かる，(c)は左折しているフレーム

4.2.2　重畳画像の生成

物体の動き情報すなわちオプティカルフローを RGB の静止画（以下，オプティカルフロー画像と呼ぶ．）として可視化可能であり，動画のフレーム

にオプティカルフロー画像を重ね合わせることにより，動画シーン認識においても，これまで述べてきた静止画シーン認識の手法が応用可能となる．

ただし，オプティカルフロー画像の画素値を単純に上書きあるいは加算した場合，フレームに写る物体等の形状などの情報が全く消失してしまう可能性があるため，オプティカルフロー画像を半透明（アルファブレンド）に変換して重ね合わせるものとする．

ここで重要となるのは，画像を重ね合わせる方法である．図 4.4 にその過程を示す．

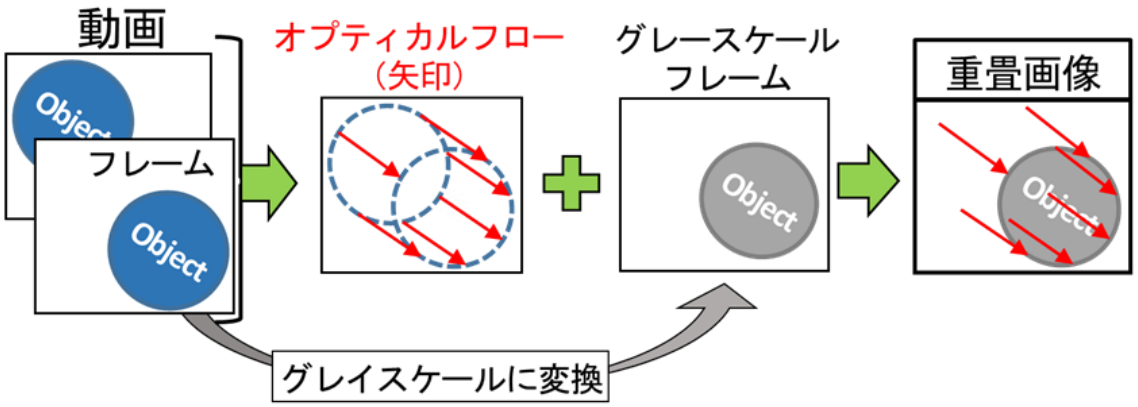


図 4.4　重畳画像の生成プロセス

フレームは RGB 画像であり，オプティカルフロー画像と重ね合わせた場合，オプティカルフローの色情報が影響を受け，動き情報が変化してしまう可能性がある．これを避けるため，元のフレームをグレースケール RGB 画像（以下，グレースケールフレームと呼ぶ）に変換し，そこにオプティカルフロー画像を重畳するという処理を行う必要がある．これにより，物体や背景の色情報による動き（方向）情報の変化は発生しなくなる．たとえば，赤い車両が左折している場合，フレームのグレースケールへの変換処理に伴い車両の「赤」という色情報は失われるが，車両が左折中である（オプティカ

ルフロー画像は「赤」）という情報は表現されているということであり，また，停止している赤い車両を左折中であると誤認識することもない．なお，物体の色情報が必要であれば，別途，画像処理により情報を取得すればよい．

図 4.5 にフレーム及びグレースケールフレームとオプティカルフロー画像を重畳した静止画（以下，重畳画像と呼ぶ）の例を示す．

(a)　　(b)

図 4.5　左折シーンの 1 フレーム画像の比較

(a)の静止画には車両の動き情報が全く表現されていないため，停車しているのか走行しているのかが分からない．(b)からは静止画であっても，車両が左折しているというシーンのある 1 フレームであることが表現されている．

ここで問題となるのが，重畳画像の生成におけるオプティカルフローの情報の変化である．図 4.7 に示すとおり，フレームに写る物体 A は左へ移動しているとしているので，オプティカルフロー画像は赤で表現される．しかしながら，オプティカルフロー画像の半透明化及びグレースケールフレームとの重ね合わせに伴い，オプティカルフロー画像の元の明るい赤から暗い赤に変化している．すなわち，オプティカルフローの速さの情報（HSV の明度 V）が変化してしまっているということである．しかしながら，重畳画像には，左に移動（赤）と物体の形状（四角形）の情報が表現されているので，

ＣＮＮ特徴ベクトルにもこれらの情報が反映（人間には理解不能である）されているものと考えられる．したがって，重畳画像による学習も十分に可能であると考えられることから，これを以下の実験により確認する．

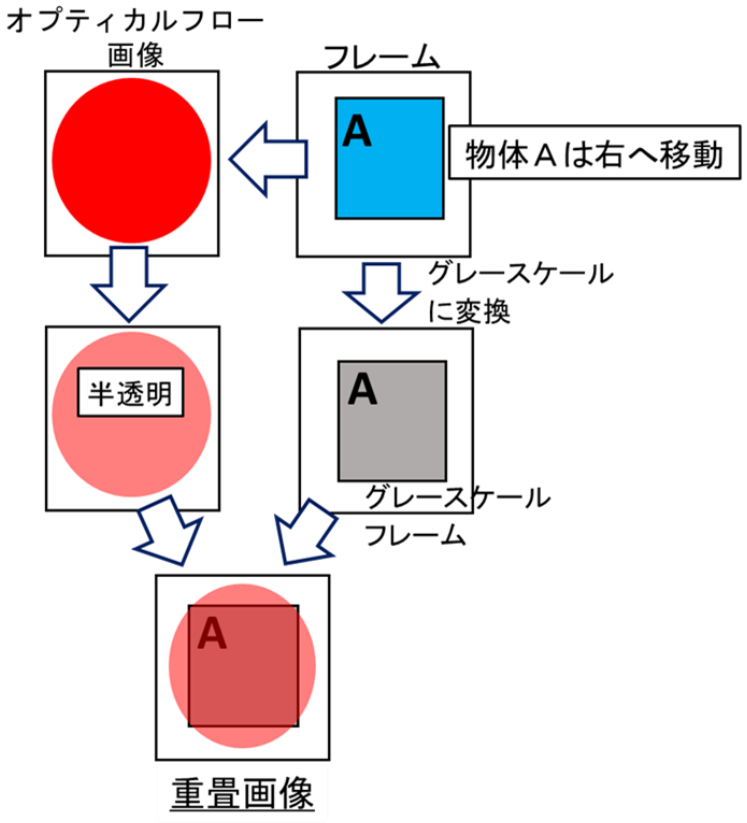


図 4.6　重畳画像の生成におけるオプティカルフロー情報の変化

実験では，車両が直進している（直進クラス）のか，左折している（左折クラス）のかを，重畳画像，オプティカルフロー画像及びフレームの，それぞれの画像を用いて学習した 3 種類の識別器のうち，どの識別器が最も識別能力が高いかを評価する．これにより，どの画像が最も学習に適しているかを明らかにする．

トレーニングデータセットは，重畳画像，オプティカルフロー画像及びフレーム画像の 3 種類で，データ数はそれぞれ，「直進クラス」は 1240 枚，「左

折クラス」は 1243 枚とした．また，図 4.7 にそれらのサンプル画像を示す．識別能力の評価は，トレーニングデータ数が比較的多いことから，5 分割の交差検証法による正解率により行った．表 4.1 に実験結果を示す．

	重畳画像	オプティカル フロー画像	フレーム
直進クラス			
左折クラス			

図 4.7　各トレーニングデータのサンプル画像

表 4.1　トレーニングデータの種類別の評価実験結果

	重畳画像	オプティカル フロー画像	フレーム
正解率（%）	99.96	99.52	99.88

評価実験の結果，重畳画像をトレーニングデータとして用いた場合，最も識別能力の高い識別器が生成可能であり，これにより，重畳画像ではオプティカルフローの情報に変化が生じてしまうものの，オプティカルフロー画像のみによる学習に比べ，高い識別精度が得られる学習が可能であることが示された．

4.3 多数決方式を用いた動画シーン認識

重畳画像を学習と識別に用いた動画シーン認識の手法を図 4.8 に示す.

識別器生成プロセスでは, 学習用動画のフレーム毎に生成された重畳画像から抽出した CNN 特徴ベクトルに正解情報が付与されたトレーニングデータを学習させることにより, SVM 識別器を生成する.

識別プロセスにおいては, 入力された未知動画のフレームから重畳画像を生成し, その CNN 特徴ベクトルを抽出する. これを識別器生成プロセスで生成した SVM 識別器に入力することにより, 重畳画像 1 枚ごとの識別結果が出力される. しかしながら, 暴力行為などのシーンでは, 一連の継続した動作・行動を複数フレーム（重畳画像）を用いることにより, 精度の高い認識が可能となることから, 複数の重畳画像を用いてシーン認識を行う必要がある.

図 4.8 に示した認識手法では, あらかじめ設定した, ある時間間隔(以下, 認識周期と呼ぶ）T 内に含まれる 1 から n 個の重畳画像の識別結果（正例／負例）の多数決を取り, その認識周期 T に対するシーン認識結果（危険／非危険シーン）を決定する. 多数決を取るのは, 認識周期内に暴力行為が発生（危険シーン）していれば, 正例の識別結果数が負例の識別結果数以上となると推測されるためである. これを「多数決方式」と呼ぶこととする.

なお, 認識周期はシーンのタイプに応じて適切な長さに設定する必要がある.

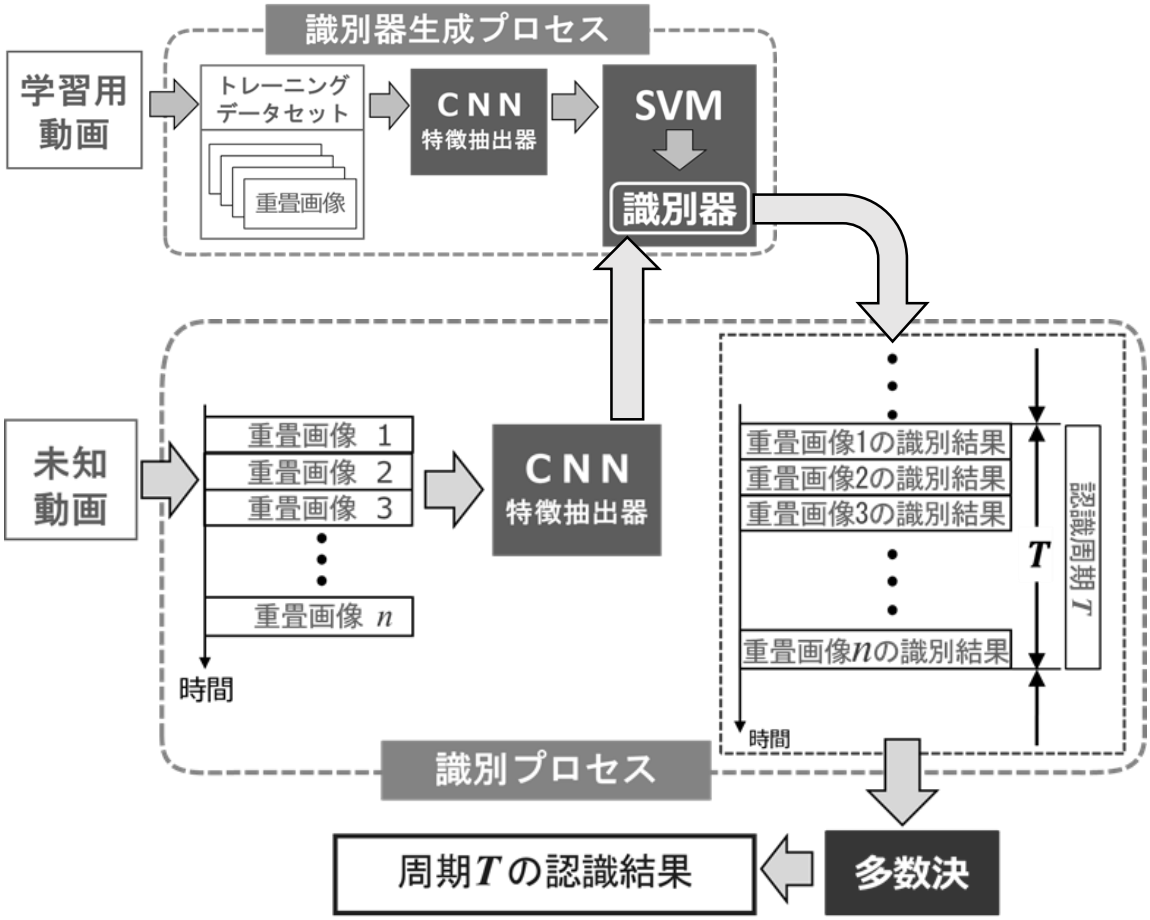


図 4.8　多数決方式を用いた動画シーン認識

この動画シーン認識手法における識別器生成プロセスにおいて留意すべき点は，動画の連続性を考慮したトレーニングデータセットの生成方法であることから．以下に，その方法について詳述する．また，多数決方式についても具体的な例を示し詳述する．

（1）トレーニングデータセットの生成方法

① 学習用動画は，認識対象とするシーンの部分がトリミングされているものとする．

② 学習用動画から，重畳画像を生成する．（30fps で 1 秒間に 30 枚）

③ 学習用動画の長さによっては重畳画像の枚数は膨大なものとなる場合がある，そこで SVM の学習時間を低減させるため，冗長な重畳画像

のいくつかを削除する(間引く)ことが可能である. これは図 4.9 に例示すとおり, SVM の学習ではサポートベクトルによってのみ識別境界線(面)が決定されるため, まとまって分布する冗長なデータは識別境界線の決定に大きな影響を与えることは少なく, したがって, データを間引く度合にもよるが, それによる認識精度への悪影響はほとんど起きないと考えられる.

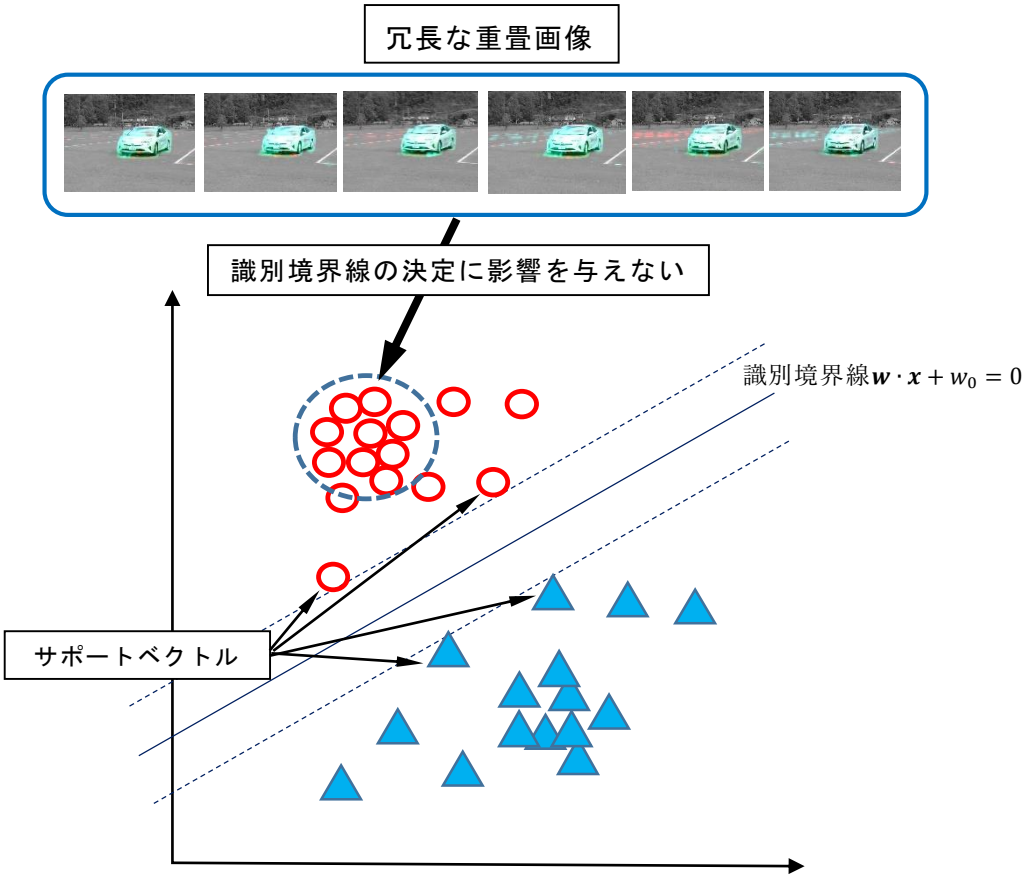


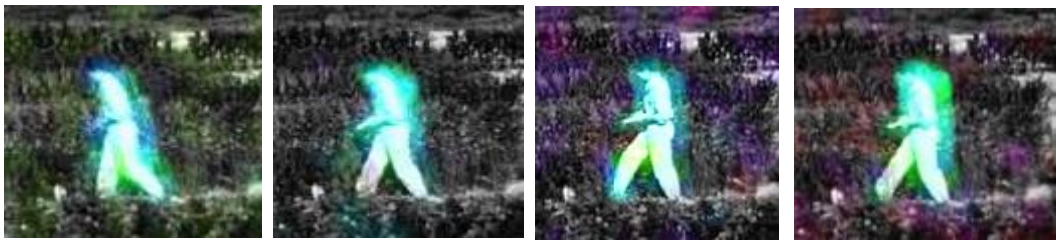
図 4.9　サポートベクトルによる識別境界線の決定

（２）多数決方式を用いたシーン認識結果の決定方法

認識対象である危険な行為・動作などのシーンは，複数フレーム(ある程度の継続的な時間)により構成されており，シーン中のある 1 フレーム（重畳画像）の識別結果だけを見て，それを直ちに，静止画シーン認識の場合と同様に．シーン認識結果とすることが適当ではないためである．たとえば，30fps（1 フレーム約 33ms）の防犯カメラ映像により暴力行為の発生の有無を監視する場合，暴力行為が発生していな状況であっても，間欠的に何フレームかは暴力シーンと誤認識されるケースが生じてしまう．しかし，暴力行為が 33ms 程度の間隔で発生したり消滅したりすることは考えらず，それを真に暴力行為が発生したとは判断できないということである．

したがって，動きのあるシーンを認識する場合，ある一定時間の継続した動き(複数フレーム)を認識周期により捉えシーンを認識する．

以下に多数決方式を用いたシーン認識結果の決定方法を，図 4.10 の「歩行シーン」を例に説明する．図は 1 から順番に，左足を一歩前に出す動作であり，前後に同様の左足を一歩前に出す動作，あるいは立ち止まっているなどのシーンが連続しているものとする．



4　3　2　1

8　7　6　5

図 4.10　歩行シーンの 1 フレーム毎の重畳画像

① 歩行シーン全体は8フレームの認識周期で足を一歩前に出す動作の繰り返しで構成されていると考えることができる．この認識周期でシーン認識結果を出力する．なお，認識周期は認識対象のシーンに応じて経験的に設定する必要があると考えられる．

② 認識周期をTとする．また，T内に含まれる重畳画像数をnとする．「歩行シーン」の例では，$T = 0.25\mathrm{s}$ ，$n = 8$ とする．また，重畳画像が歩行シーンである場合の識別結果を正例，それ以外の立ち止まっているなどの非歩行シーンを負例とする．

③ 認識周期Tの認識結果を歩行シーンであると決定するためには, SVM が 2 値分類器であることを考慮すると，多数決方式では重畳画像の識別結果が正例である回数を 4 回以上に設定すればよく，これにより非歩行シーンも決定される．すなわち，歩行シーン認識においては，重畳画像 8 枚中に 3 回しか正例が出現しなければ，その周期は歩行シーンではないと認識されることになり，その逆は歩行シーンであると認識される．

4.4 多数決方式による動画シーン認識の評価実験

多数決方式のよる動画シーン認識の手法を評価するため, 人物が危険な行為（暴力的な行為）を行っているシーンを認識する実験を行う．

人物の動作認識の分野では UCF-101[33, 34]と HMDB-51[35, 36]が標準的なベンチマーク用のデータベースとして公開されており, このデータベースに含まれる動画をこの実験に利用する．

UCF-101 は, 101 クラスの人物の動作(スポーツ中心)を認識するためのデータベースで 13,320 本，約 27 時間の動画で構成されており，YouTube のデータをもとに作成されている．HMDB-51 は，51 クラスの動作認識を行うためのデータベースで 1 クラス当たり約 100 本，1 本あたり約 2〜3 秒の

動画で構成されており，映画，YouTube，Web のデータをもとに作成されている．

これらのデータベースによる動作認識の精度は，表 4.2 のとおりである．（2017 年）

表 4.2　UCF-101, HMDB-51 における認識精度(正解率)

学習法	UCF-101	HMDB-51
Two-stream ResNet+IDT	94.6%	70.3%

※表中の正解率はテスト動画よる評価結果である．

ここで，表中の学習法について概説する．

Two-stream[37]は，図 4.11 に示すとおり RGB 画像とオプティカルフローの画像を別々の CNN で学習し，それぞれのクラススコアを統合させ認識結果とする手法であり，これが現在の動作認識手法の主流となっている．

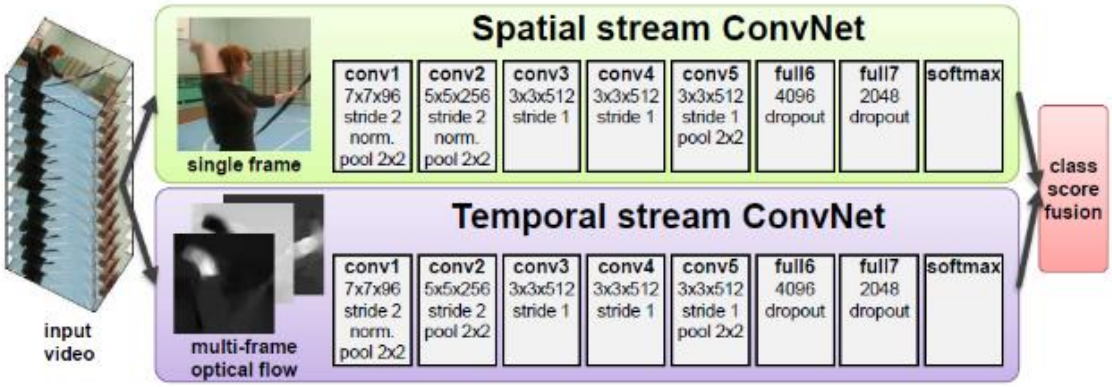


図 4.11　Two-stream アーキテクチャー による動作認識

また，このような動画を用いた CNN による動作認識においては，図 4.12 に示すような画像の特徴の融合も用いられる[38]. Single Frame は 1 フレームごとに学習される通常のネットワークである．Late Fusion は最初と最後のフレームを入力とし 2 つの Single Frame の特徴を最後の全結合層で融合

する．Early Fusion は複数の画像を深さ方向に連結し 1 枚の画像にして入力することで Single Frame と同様の学習を行う．Slow Fusion は Early Fusion と Late Fusion の中間的方式で入力と畳み込み層で特徴を融合する．

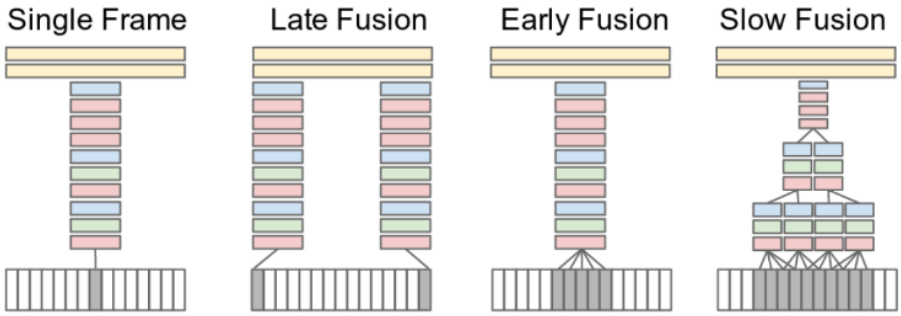


図 4.12 画像の特徴の融合方式[38]

また，ResNet [39]については，ILSVRC 2015 においてエラー率 3.57%で圧勝した CNN で，ネットワークの階層が 152 層（GoogLeNet でも 22 層）と非常に深い構造となっているのが特徴である．

IDT(Improved Dense Trajectories) [40]については，オプティカルフロー取得の改良を行い，従来発生していた背景の余分なフローを除去し，精度向させたものである．

以上を踏まえると，提案モデルは，“One-stream，Single Frame，GoogLeNet+Farneback”であると言える．

なお，近年は行動認識にも 3DCNN が導入され，フレーム 16 枚，最近は 32 枚を重ねて 3DCNN に入力し，学習と認識を行うケースも増えてきている．

この評価実験で使用するデータベースは HMDB-51 とした．このデータベースは，含まれる動画の画像サイズ，フレームレートがファイルによって異なり，また，動画自体が映画などのワンシーンをクッリプした映像である

ためクラス名とは全く異なるシーンが挿入されているケース（例：人が倒されるシーン中に顔のアップの映像が混入するなど）もあり，行動認識のためには非常に難易度の高いデータベースとなっている．したがって，学習/認識は容易ではなく，そのことは表 3.17 の認識精度(正解率 70.3%)にも表れている．なお，CNN が動作認識に導入された当初(2014)は正解率 59.4%であったが，2019 年時点では正解率は，82%程度となっている[35]．このような難易度の高いデータベースを用いた評価実験においても，良好な認識結果を得ることができ，そのことは多数決方式の有用性を示いている．

以下に，その評価実験について述べる．

（1）実験に使用する動画は，HMDB-51 の中から暴力的な行為のシーン(以下，危険シーンと呼ぶ)と暴力的な行為のない日常的に見られるシーン(以下，非危険シーンと呼ぶ）をそれぞれ 3 シーン（クラス）抽出する．

（2）抽出した各動画を学習用/テスト用とし，危険シーン(正例)/非危険シーン(負例）に 2 値分類する．なお，多数決方式のよる動画シーン認識の手法は，3.3.1 で述べたとおり，手の動き，足の動き，体の傾きなどの個々の動作（3.3.1 ではオブジェクト）の組合せに基づく論理推論により，危険シーンであるか否かを認識するものではないことから，個別の動作の認識は行わない．

データベースから抽出した動画は次のとおりである．

① 危険シーン

危険性のある暴力的な行為と考えられる次の 3 シーンである．

・人が倒れる，倒されるシーン(クラス名：“fall_floor”)

・凶器で物や人を叩くシーン(クラス名：hit)

・人や物を殴るシーン(クラス名：punch)

したがって，それぞれのクラスの動画のクラス名は「危険シーン」に変更する．

これらのシーンの例を図 4.13 に示す．

(a)　　(b)　　(c)

図 4.13 危険シーンの例（a) fall_floor クラス，(b) hit クラス，(c) punch クラス [35]より

②非危険シーン

正例に対応して日常的な行為と考えられる次の 3 シーンである．

・歩くシーン(クラス名：walk)

・抱擁するシーン(クラス名：hug)

・握手するシーン(クラス名：shake_hands)

したがって, それぞれのクラスの動画のクラス名を「非危険シーン」に変更する．

これらのシーンの例を図 4.14 に示す．

(a)　　(b)　　(c)

図 4.14 非危険シーンの例（a) walk クラス，(b) hug クラス，
(c) shake_hands クラス [35]より

（3）学習用動画数

それぞれのシーンから 70 本×6 シーンで計 420 本とした.

（4）トレーニングデータセット

各学習用動画から，重畳画像を生成した．ただし，すべての重畳画像は数十万枚となることから，前述した方法により，冗長な画像は適宜削除した.

正例は，7593 枚，負例は 9208 枚とした.

（5）テスト用動画数

それぞれのシーンから 30 本×6 シーンで計 180 本とした.

（6）評価方法

① トレーニングデータセットによる評価

データ数が多いので 5 分割交差検証法を用いた.

② テスト用動画による評価

テスト用動画は,動画中の危険シーンの部分のみに対して教師信号が付与されているわけではなく,1本の動画全体に対して1クラスの教師信号が付与されているので，認識周期は動画全体の長さとし，(正例と認識された重畳画像数)／(動画全体の重畳画像数)≥ 0.5 のときは危険シーンの動画, それ以外を非危険シーンの動画とし, それらの正解率により評価する.

（7）評価実験結果

評価実験結果は，表 4.3 及び 4.4 のとおりである.

評価方法①のトレーニングデータセットによる 5 分割交差検証結果

正解率 92.6%（詳細は表 4.3 のとおり）

表 4.3　危険シーン認識実験結果の混同行列

		シーン認識結果	
		(危険シーン)	(非危険シーン)
正解クラス	危険シーン	6873	720
	非危険シーン	531	8677

評価方法②のテスト用動画による評価結果

正解率 81.7%（詳細は表 4.4 のとおり）

表 4.4　テスト用動画による危険シーン認識実験結果の混同行列

		シーン認識結果	
		（危険シーン）	（非危険シーン）
正解クラス	危険シーン	74	16
	非危険シーン	17	73

多数決方式のよる動画シーン認識の評価結果は 2 値分類であり，表 4.2 の HMDB-51 による動作認識結果は 51 クラス分類であることから，それらの結果を単純に比較することはできないが，最新の HMDB-51 による動作認識の精度が 82%前後であることを考慮すると，多数決方式のよる動画シーン認識の精度は概ね良好であると言える．

しかしながら，この多数決方式を用いた動画シーン認識では，認識対象とする危険シーンに合わせてヒューリスティックに認識周期を決定しなければならないという課題がある．認識対象とする危険シーンのデータが大量に入手可能で，最適と考えられる認識周期を実験等により決定可能であれば，多数決方式は有効に機能すると考えられるが，適切に認識周期を設定できなかった場合，例えば認識周期が長すぎ場合は，危険シーンの見落としが発生する．

4.5　隠れマルコフモデルを用いた動画シーン認識

多数決方式を用いた動画シーン認識の課題を解決するため，図 4.15 に示す隠れマルコフモデルを用いた動画シーン認識の手法を提案する．

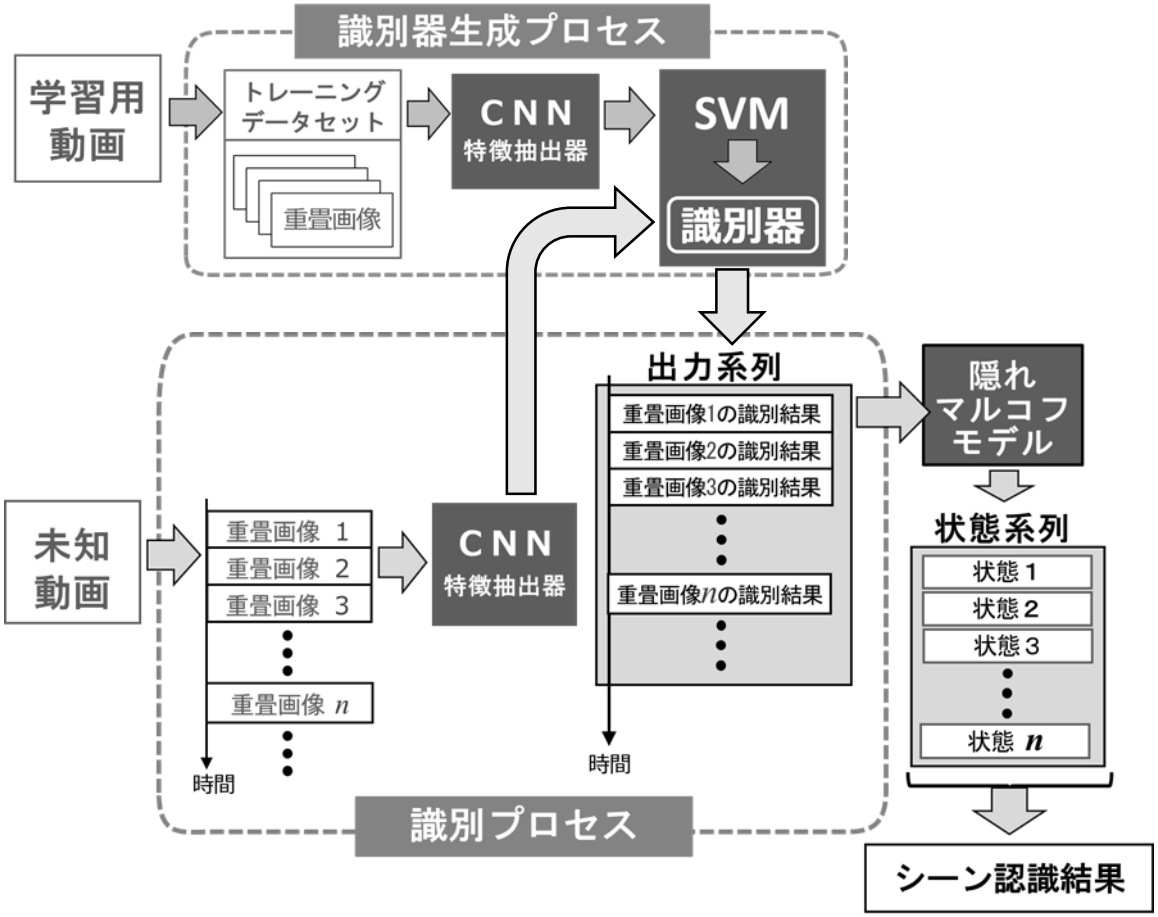


図 4.15 隠れマルコフモデルを用いた動画シーン認識

この手法では,

（１）識別器生成プロセス

多数決方式と識別器生成プロセスは全く同様で, 学習用動画から生成した重畳画像を用い，重畳画像の識別器を生成する.

（２）識別プロセス

識別プロセスでは，入力された未知動画のフレームから，4.2（２）で述べた方法により重畳画像を n 枚生成する．これらを CNN 特徴抽出器に入力し，CNN 得られた特徴ベクトルを，識別器生成プロセスで生成した重畳画像識別のための学習済み SVM に入力する．これにより，n 個の重畳画像識

別結果からなる系列（以下，出力系列と呼ぶ.）を得ることができる． ただし，出力系列の要素数 n については，認識対象としている危険シーンの平均的な継続時間を考慮し，適宜決定する.

（3）シーン認識結果の決定方法

図 4.15 に示すとおり，危険／非危険のシーン認識結果は，隠れマルコフモデルによって，要素数 n （以下，系列長と呼ぶ.）の出力系列から得られた系列長 n の状態系列が，ある一定の条件を満たしたときに，危険シーンであると決定し，それ以外の場合は，非危険シーンであると決定する.

このように，状態系列を用いシーン認識結果を決定することにより認識精度の向上が期待できる．その理由について以下に述べる.

① 出力系列

図 4.16 に危険シーンの動画から得られた出力系列の例を図示する．縦軸は識別結果（識別結果が正例の場合はO_1，負例の場合はO_0とする.），横軸は時間経過を重畳画像番号で表している. なお, 動画のフレームレートを 30fps としているので，この出力系列は約 1.1sec（系列長 $n = 36$）である.

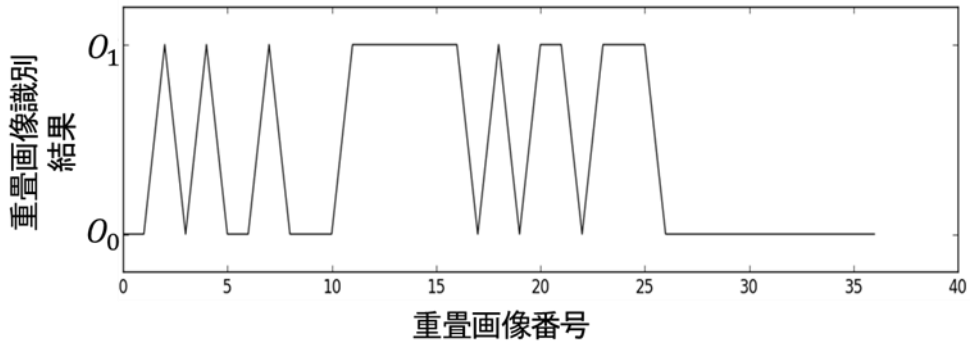
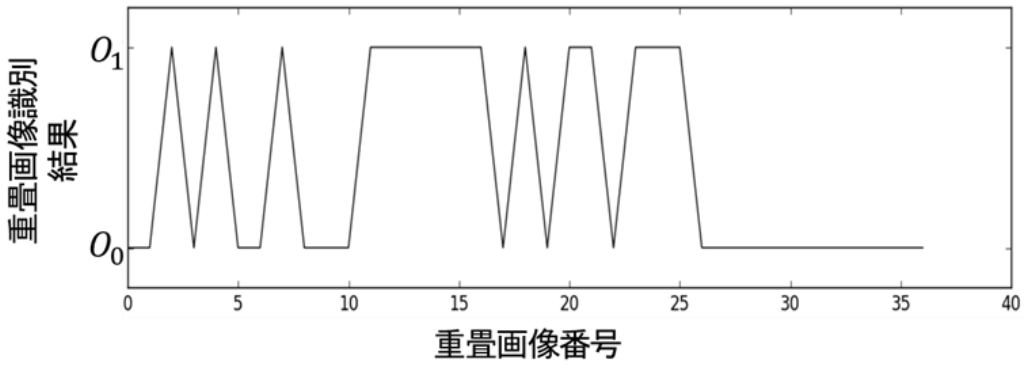


図 4.16 出力系列（系列長 $n = 36$）

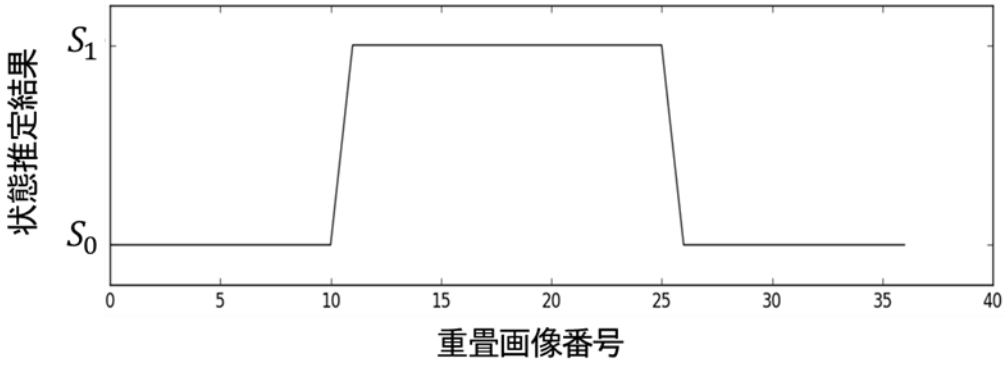
この動画において危険事象が発生しているシーンに対応する部分は，おおよそ重畳画像 10～25 番であると推測できる．しかしながら，出力系列では，重畳画像 2～7 番あるいは 17～23 番の区間において，0.03sec 毎あるいは 0.06sec 毎に識別結果が変化している．このような短い時間間隔で変化する出力系列と，動画における危険事象の発生・消滅に対応する危険シーンとは一致しない識別結果が発生している部分があることが分かる．このような不一致を誤識別と呼ぶこととする．したがって，図 4.16 では，重畳画像 2，4，7，17，19，22 番において誤識別が発生しているものと推測できる．このような出力系列を用いてシーン認識結果を決定した場合，高い認識精度を得ることは難しいと考えられる．

② 状態系列

そこで，このような誤識別の問題を改善するために隠れマルコフモデルを導入する．あらかじめ学習済みの隠れマルコフモデルを用い出力系列から危険事象発生の有無の状態系列を推定する．これにより，図 4.17(a)の出力系列とは異なり真値に近い，図 4.17(b)に示すような状態系列を推定（推定結果が正例の場合はS_1，負例の場合はS_0とする．）でき，この状態系列を用いることにより，誤識別の影響を軽減できるため，高精度のシーン認識結果を得ることができる．



(a)出力系列



(b)状態系列

図 4.17 隠れマルコフモデルを用いた出力系列からの状態系列の推定

この動画シーン認識手法で用いる隠れマルコフモデルを図4.18に例示す．S_1は危険事象が発生している状態，S_0は危険事象が発生していない非危険状態を表している．O_1は，状態S_1あるいは，S_0のときに出力された重畳画像識別結果（正例）であり，同様にO_0は重畳画像識別結果（負例）である．また，実線の矢印は状態遷移を，破線の矢印は状態からの出力を示している．

隠れマルコフモデルの学習では，動画シーン認識の対象とする危険シーンと非危険シーンの動画をそれぞれ数十本用意し，それらの動画から得られた重畳画像の出力系列を隠れマルコフモデルに入力し，Baum-Welch ア

ルゴリズム（EM アルゴリズム）を用いて状態遷移確率と出力確率を推定する[41，42]．この学習済みの隠れマルコフモデルから Viterbi アルゴリズムにより，出力系列から状態系列を推定することができる[43]．

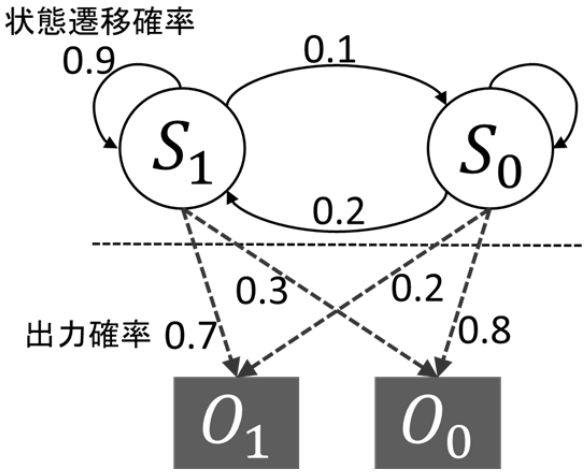


図 4.18 動画シーン認識のための隠れマルコフモデル

③　状態系列を用いたシーン認識結果の決定手法

図 4.17(b)の例を用いシーン認識結果を決定する手法について述べる．図では危険シーンの認識対象とする系列長を $n = 36$ と設定している．

認識結果は，状態系列において，S_1 の継続時間がある一定時間以上の条件を満たす場合は，危険シーンであると決定し，それ以外の場合は，非危険シーンであると決定する．すなわち，S_1 が連続する系列長をmとすると，mがある一定数以上であれば,危険シーンであると決定するということである．このmは，対象としている危険シーンを認識可能となる最小の継続時間である．これを図に対応させ，たとえば $m = 15$ （30fpt の場合，あるシーンが始まって約 0.5sec で危険シーンであると認識可能とである．）とすると，図

の状態系列は 11～25 番の区間においては，S_1 の継続時間が 15 となっているので，危険シーンであると決定する条件を満たしていることとなる．

以上は，ある長さにクリップされた動画において，危険シーンを認識するための例であったが，実用上は監視カメラなどから得られるリアルタイム映像に対し，連続してシーン認識結果を出力する必要がある．そのタイミングについて述べる．図 4.19 に示すように，$n \geq m$ となる系列長 n を設定し，隣り合う出力系列の先頭要素の重畳画像番号の間隔を s に定める．各出力系列から隠れマルコフモデルを用いて得られた状態系列において，S_1 の継続時間がある一定時間 m 以上の条件を満たすかどうかにより，危険シーンであるかどうかを決定する．図 4.19 では，認識結果が危険シーンであると決定されるのは，3 番目の出力系列から得られた状態系列 3 となる．それ以外の状態系列の認識結果は非危険シーンであると決定される．

間隔 s については，$s \leq n - m + 1$ に設定することにより，映像のどの時間帯に危険シーンの決定条件を満たす幅 m の区間が存在しても，その危険シーンを認識できる．図 3.37 の例では，$n = 20$，$m = 15$，$s = 5$ に設定されており，いずれかの状態系列において危険シーンの認識が可能となることがわかる．

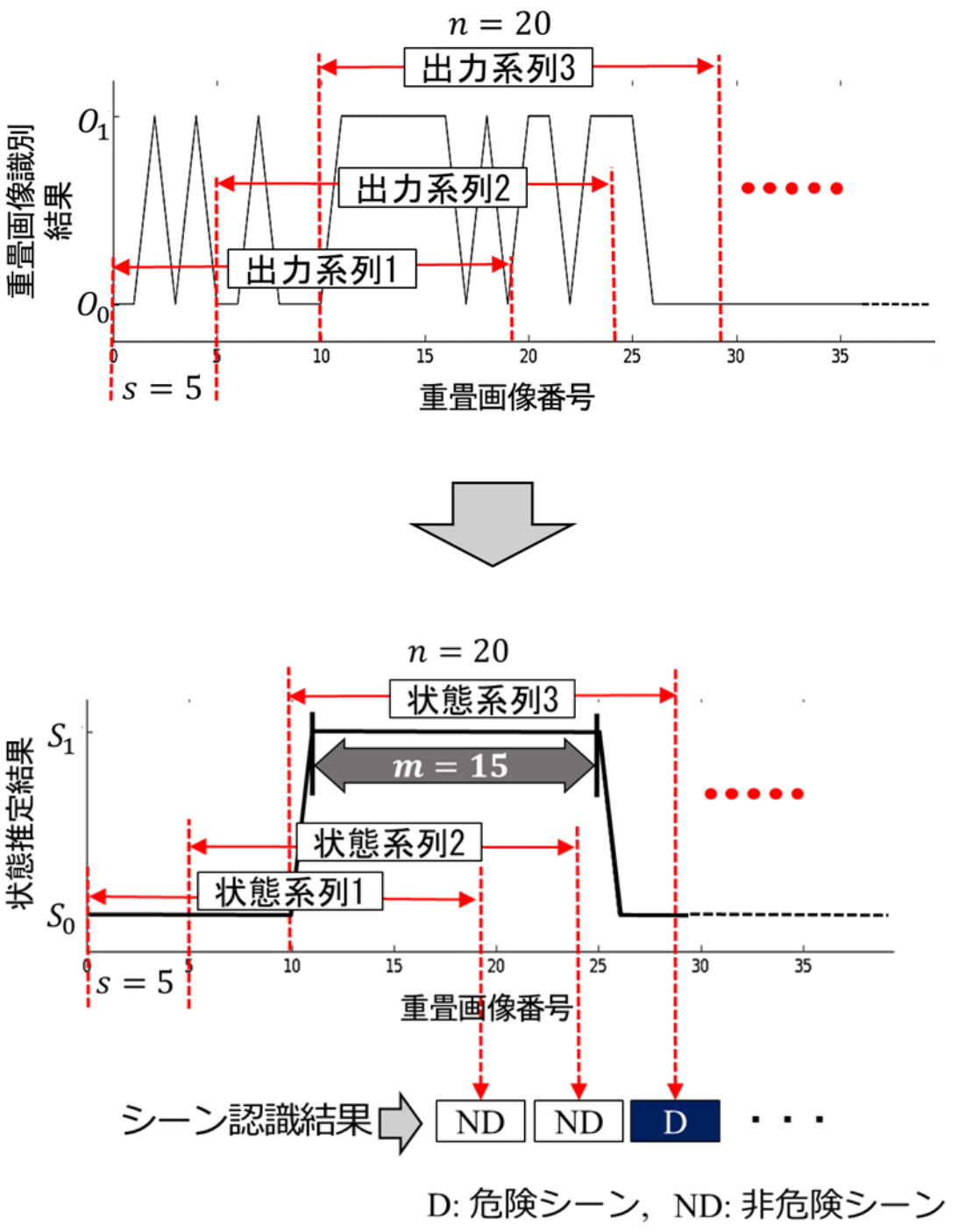


図 4.19 動画シーン認識のタイムチャート

4.6　隠れマルコフモデルを用いた動画シーン認識の評価実験

多数決方式を用いた動画シーン認識の評価実験に用いた HMDB-51 を利用し, 隠れマルコフモデルを用いた動画シーン認識手法の有効性を評価する.

HMDB-51の中から危険シーン及び非危険シーンとして選択するトレーニング用の動画は, 多数決方式と同様の 6 種類のクラスからそれぞれのシーンから 70 本を選択し 70×6＝420 本とした．また，テスト用の動画数についても，同様に，6 種類のクラスからそれぞれ 30 本を選択し 30×6＝180 本とした.

隠れマルコフモデルについては, 状態遷移確率と出力確率を推定するための学習用の出力系列が必要となる．この学習用出力系列は，危険／非危険シーンの動画からランダムにそれぞれ 30 本を抽出し，それらの動画から生成される重畳画像を多数決方式において学習済みの SVM 識別器に入力することにより得ることができる．図 4.20 に作成された学習用の出力系列を図示する.

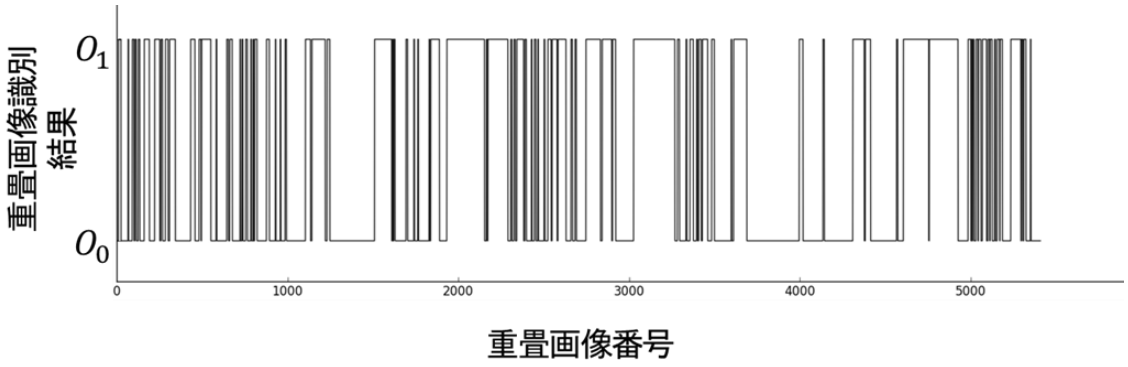


図 4.20 隠れマルコフモデルの学習用の出力系列

また, この評価実験における前述の危険／非危険のシーン認識結果を決定するための条件は, テスト用動画の出力系列を隠れマルコフモデルに入力して得られる状態系列の中に S_1 が 15 回（$m = 15$，約 0.5sec）以上継続して出現した場合, その動画のシーン認識結果を危険シーンと決定することとし

た．したがって，非危険シーンの動画であると認識するのは，S_1 が状態系列の中に 15 回数以上連続して出現しなかった場合である．なお，出力系列長 n は動画の長さとした．

評価実験結果を表 4.5 に示す．正解率は 86.1%であり，良好な認識精度が得られた．

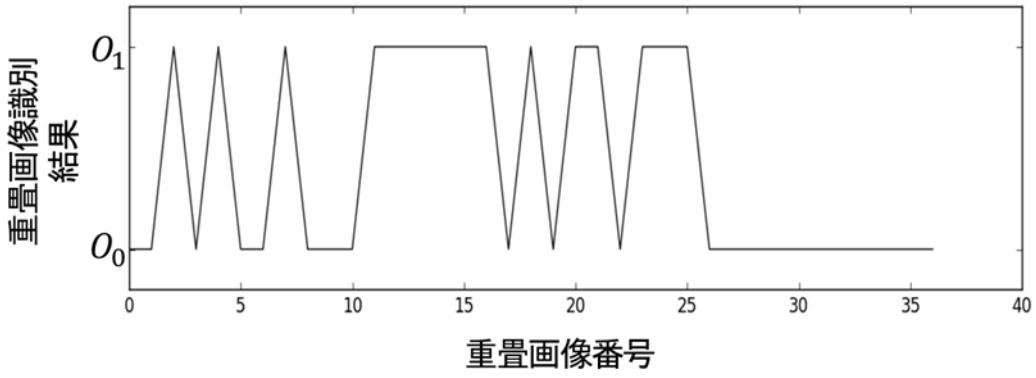
表 4.5 テスト用動画による危険シーン認識実験結果の混同行列

		シーン認識結果	
		（危険シーン）	（非危険シーン）
正解クラス	危険シーン	74	16
	非危険シーン	17	73

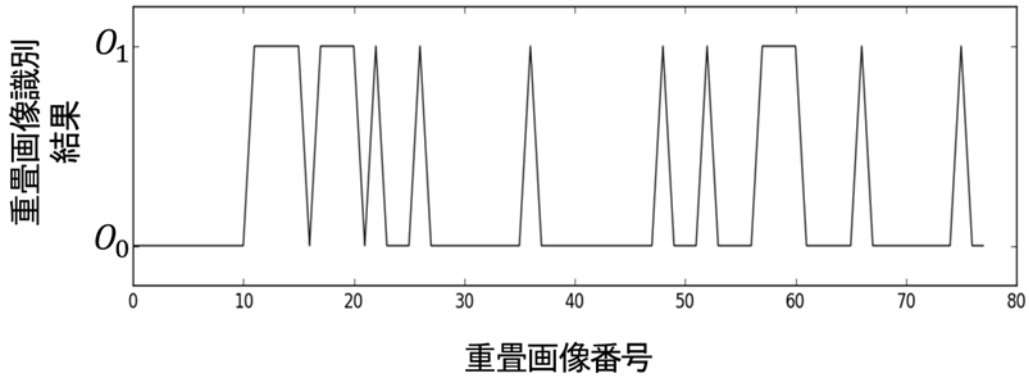
4.7 多数決方式を用いた動画シーン認識と隠れマルコフモデルを用いた動画シーン認識との比較

動画シーン認識の手法においては，危険／非危険シーンがある行動や動作であり，それはある一定の時間継続することを前提としている．多数決方式を用いた動画シーン認識では，動画をある時間周期（以下，認識周期と呼ぶ．）で分割し，この認識周期に対するシーン認識結果を決定する．そこでは，認識周期内の複数枚のフレームから重畳画像を作成し，重畳画像 1 枚毎に危険（正例）／非危険（負例）を識別し，次に，それらの重畳画像識別結果の多数決を取り，正例の枚数が負例の枚数以上であれば，その認識周期に対するシーン認識結果を危険シーンであると決定する．

テスト用動画を用いた多数決方式による評価実験結果は,正解率 81.7%であり,隠れマルコフモデルを用いた手法による正解率 86.1%と比較して 4.4%低下している．多数決方式の問題点として，認識周期をすべての動画に対して適切に設定することは困難であることがあげられる．たとえば，図 4.21 (a)にテスト用の危険シーン動画（以下，動画 D1 と呼ぶ.）と図 4.21(b)に非危険シーン動画（以下，動画 D2 と呼ぶ.）の重畳画像識別結果の例を示す.



（a）危険シーン動画（D1）



（b）非危険シーン動画（D2）

図 4.21　テスト用動画から得られた重畳画像識別結果の例

認識周期を動画の長さと同じ長さに設定した場合,動画 D1 では重畳画像識別結果が，O_1＝15 回，O_0＝21 回となり，これらの多数決をとるとシーン

認識結果は，非危険シーンとなってしまう．動画 D2 については同様に多数決をとると，正しく非危険シーンと認識される．そこで，動画 D1 の誤認識を解消するため認識周期を短くして 0.3sec(重畳画像 10 枚)に設定すると，第 2 周期目の多数決で危険シーンと正しく認識される．しかしながら，動画 D2 では，第 2 周期目の多数決で危険シーンであると誤認識されてしまう．また，2 つの認識周期にまたがって危険シーンが発生した場合，そのシーンの見落としが発生する可能性がある．

一方，隠れマルコフモデルを用いた手法では，出力系列から危険事象が発生しているか否かの真の状態に近い状態系列を推定することができるため，多数決方式のように出力系列から直接シーン認識結果を決定する場合に比べ，認識精度が向上すると考えられる．

4.8 むすび

本章では，動画を用いたシーン認識の手法を示した．動画シーン認識では，オプティカルフローにより得られる物体の移動/静止の動き情報を RGB 画像に変換しグレースケールフレームと重ね合わせた重畳画像を用いる．

多数決方式による動画シーン認識については，認識周期内の重畳画像の識別結果の多数決を取ることにより，シーン認識結果を決定する手法である．評価実験では，人物動作認識の評価に標準的に利用されているデータセットを用い認識精度を検証した．その結果，正解率 81.7%であった．単純には比較できないが，当該データセットによる最先端の認識手法による認識精度と同等の性能を有していることが明らかとなった．隠れマルコフモデルを用いた動画シーン認識については，危険事象が発生しているか否かの真の状態に近い状態系列を隠れマルコフモデルを用い出力系列から推定することにより，多数決方式で課題となった認識周期の設定問題を解決することができた．評

価実験では，多数決方式の評価と同じデータを用い，その結果，正解率は86.1%となり認識精度が大きく向上した.

以上の結果から，シーン認識は，危険シーンに合わせトレーニングデータを入れ替え,SVM識別器あるいは隠れマルコフモデルを学習させるだけで,容易に多くの種類の危険シーン認識が可能であることが明らかとなった.このことにより,シーン認識が高安全知能コンピューティングシステムにおける危険事象検出のためのプラットフォームとして有効に機能することが示された.

第５章
シーン認識に基づく危険事象の予測

5.1 まえがき

高安全知能コンピューティングシステムにおいては，危険事象が発生した場合に本システムのユーザが危険を回避し安全な状態を確保するための支援を行うが，更に，危険事象の発生が予測される場合についても，その被害を未然に防止するための支援を行うことを考慮している．この未然防止という考え方が重要であることは言うまでもない．

本章では，このような未然防止行動支援のための危険予測の手法について提案する．提案手法では，監視カメラなどの映像から得られる危険事象が発生する前兆である可能性が高いと考えられる前兆シーンの認識結果と過去の危険事象発生に関する統計データを用いて危険予測を行う．予測結果は危険事象が発生する危険性の度合いとして出力され，行動計画において，その出力値に応じた行動支援が行われる．

5.2 危険予測の対象とする危険事象

高安全知能コンピューティングシステムにおける危険予測の対象とする危険事象について述べる．

本システムは, ユーザが危険を回避し安全な状態を確保するための支援を目的としており, 危険予測に基づく未然防止行動についても, ユーザが対処可能な範囲で支援することとなる. したがって, ユーザがその発生を未然に防止することが不可能な危険事象である自然災害などについては, 危険予測の対象とはしない. たとえば, 津波, がけ崩れ, 洪水などの被害発生を未然防止する対策は, 堤防や砂防ダムなどの建設であり, 到底, 本システムのユーザがそれを行うことは不可能である.

また, 危険予測には前兆シーン認識を用いる. したがって, 対象とする危険事象に対応する前兆シーンが存在し, 定義可能であり, そのシーンを画像情報として取得可能でなければならない. たとえば, シーン認識では心筋梗塞が発症して苦しんでいるシーンは認識可能であるが, その前兆現象として考えられる, 胸の軽い痛みや違和感を覚え, 胸に手を当てるなどのシーンは日常的なシーンと識別が困難であり認識は不可能である. このような事象についても, 危険予測の対象としない. 加えて, 本システムにおける危険予測では, 前兆シーンの認識結果と危険事象発生の因果関係を明確に示す統計データが存在しない場合もあり, 前兆シーン認識のみにより危険予測を行うことは困難な場合もありうる. たとえば, うろうろしたりきょろきょろしたり不審な行動を取っている人物の何パーセントが犯罪行為を実行するか, といった統計データは存在しないことが多いということである. このため, 前兆シーン認識と過去の危険事象発生に関する統計データを用い危険予測を行う必要が生じるのである. したがって, 過去の危険事象発生に関するデータが存在しない, あるいは著しく少ない場合, たとえば, 現在の日本ではほとんど発生していないテロなどは, 危険予測の対象としない.

以上のことから, 危険予測の対象とする危険事象は, 人為的に発生し, 前兆シーンの認識が可能であり, ユーザがその発生を未然に防止できる可能性がある過去に何度も発生しており十分な統計データが存在する危険事象である.

たとえば,コンビニ強盗や万引きなどの犯罪行為や居眠り運転などが危険予測に対象となる.客の不審な行動を前兆シーンと定義し,そのシーン認識結果と過去の犯罪発生の統計データを用いて,犯罪が発生する危険性を予測するといった場合が危険予測の対象である.

5.3 危険予測の手法

危険予測の結果は危険事象が発生する危険性の度合い(以下,確信度と呼ぶ.)として出力する.危険予測の典型例として犯罪行為を取り上げ,前兆シーン認識と統計データ分析に基づいた高安全知能コンピューティングシステムにおける危険予測の方法について提案する.

5.3.1 前兆シーン認識

前兆シーンの認識は,前述のシーン認識と同一手法により行われ,前兆シーンが認識されたときのみ危険予測処理が実行される.また,危険事象発生シーンの認識は並列で実行されている.

危険予測は,危険事象が発生する危険性の度合いに応じて,その発生を未然に防止する対策を講じるためのものであることから,前兆シーン認識の結果についても,それがどの程度確からしいか,というクラス事後確率を推定する必要がある.しかしながら,静止画及び動画シーン認識では,SVM を用いていることから,一般に認識結果は 2 クラス(正例/負例)分類され,クラス事後確率を得ることができない.この課題を解決するため,SVM の識別処理の後,識別境界面からの符号付距離を分類スコアとし,シグモイド関数によりクラス事後確率を求める方法が提案されている[44].前兆シーン認識ではこの方法を用いクラス事後確率を求めることとする.

ここで求めたクラス事後確率は,前兆シーン認識の結果のみから得られる今後危険事象が発生する確率であり,クラス事後確率が高ければ前兆シ

ーンとしての確からしさが高い，すなわち危険事象の発生確率が高いことを示していると言える．

静止画シーン認識においては，1 枚の画像に対して 1 つのクラス事後確率を推定することができる．動画シーン認識では，図 5.1 に示した危険シーンと認識された状態系列 3 の $m = 15$ の区間に対応する，重畳番号 11～25 番までの，15 枚の各重畳画像のクラス事後確率の平均値を求め，その危険シーンのクラス事後確率とする．

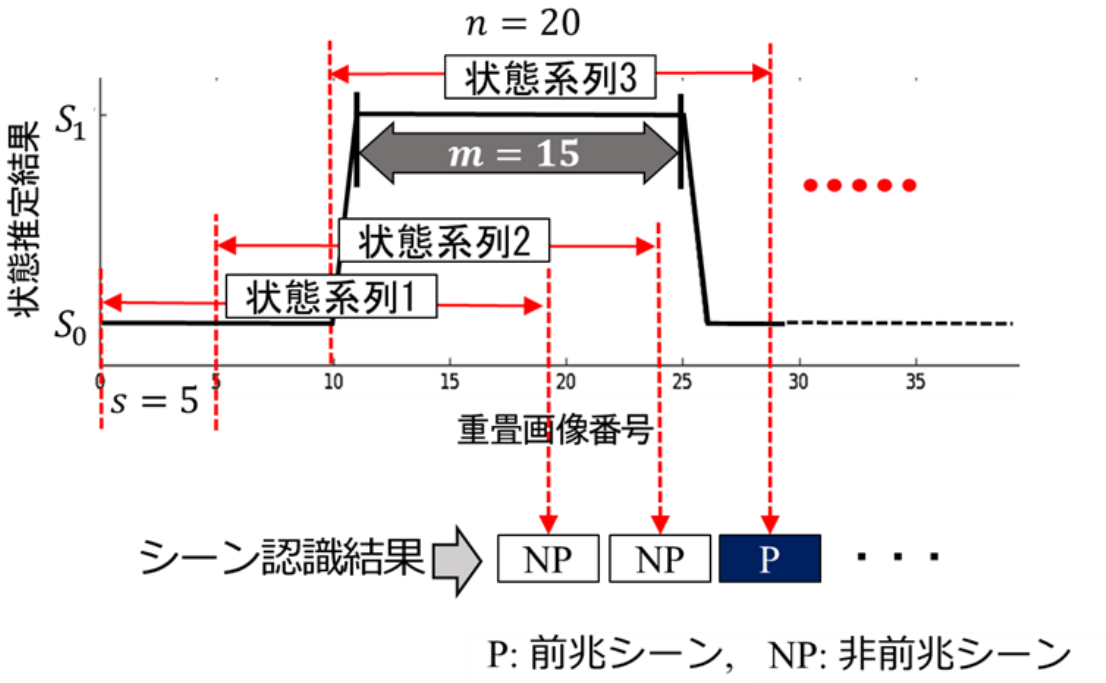


図 5.1 動画シーン認識における前兆シーンのクラス事後確率の推定

5.3.2 危険事象の発生に関する統計データの分析

危険予測に用いる前兆シーン認識では，認識対象とするシーンがどの程度前兆シーンとして確からしいかというクラス事後確率を推定することは可能である．しかしながら，その確率が高いからと言って必ずしも犯罪行為が発生するとは限らない．すなわち，人為的に発生する犯罪行為の前兆

シーンは物理現象とは異なり，その発生の必要十分条件とはなり得ないことが多いためである．したがって，前兆シーン認識の結果のみにより危険予測を行うことは困難な場合もありうる．

そこで，前兆シーン認識によって得られるクラス事後確率と，犯罪が発生しやすい時間帯や場所などの統計データの分析結果を用いて危険予測を行うこととし，これにより危険予測精度を向上させることが可能となる．

犯罪発生状況に関する統計データには，地域別の発生件数，月・曜日・時間帯別の発生件数，被疑者の変装方法，被害者数，発生場所の環境などが考えられる．これらのデータから，たとえば，平日の午前中，人通り多い環境では，ある犯罪行為は全く発生していない，あるいは，金曜の深夜帯，人通りの少ない場所で頻繁に犯罪が発生している，といった情報を得ることができる．これらの情報から犯罪の発生確率を求めることが可能である．

たとえば，年間ある一定数の犯罪が発生する地域において，ある月・曜日・時間帯に犯罪が発生する確率 p は，確率変数Mを発生月，Wを発生曜日，Tを発生時間帯とした場合，式（5.1）で求めることができる．

$$p = P(M, W, T) \tag{5.1}$$

また，対象地域のすべての場所で犯罪発生確率が等しければ，それを q とすることは可能である．しかしながら，その発生確率は，犯罪の発生場所 i $(i = 1, \cdots, n)$ の環境（たとえば，住宅街，繁華街，人通りが多い少ない，など）によって変化し，q_i となるのが現実的である．したがって，被説明変数を y_i とすると式（5.2）

$$y_i = \begin{cases} 1, \text{（犯罪発生）} \\ 0, \text{（発生せず）} \end{cases} \tag{5.2}$$

となり，これがベルヌーイ分布に従うと仮定すると式（5.3）

$$p(y_i|q_i) = q_i^{y_i}(1 - q_i)^{1-y_i} \ , \qquad 0 \le q_i \le 1 \tag{5.3}$$

となる．

また，この式のベルヌーイ分布の期待値は，$\mathrm{E}[y_i] = q_i$ であることから，ここで，回帰モデルを用いて，k個の説明変数 $\boldsymbol{x}_i = [x_{1i}; \cdots; x_{ki}]$ が与えられたもとでの y_i の条件付き期待値を式（5.4）

$$q_i = \mathrm{E}[y_i|\boldsymbol{x}_i] = \beta_1 x_{1i} + \cdots + \beta_k x_{ki} = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} \tag{5.4}$$

としてしまうと，$0 \leq q_i \leq 1$ とならない可能性が生じる．

そこで，線形回帰モデルを一般化した一般化線形モデル（GLM：Generaized Liner Model）を用いることとする[45]．

GLM においては，ベルヌーイ分布で使われるロジットをリンク関数としたロジット・モデルを用いることとする．この場合の尤度は，式（5.5）

$$p(D|\boldsymbol{\beta}) = \prod_{i=1}^{n} q_i^{y_i}(1-q_i)^{1-y_i}, \qquad D = (y_1, \cdots, y_n) \tag{5.5}$$

となり，発生場所別の犯罪発生確率q_iは，式（5.6）のロジット・モデルにより求めることができる．

$$q_i = \frac{1}{1 + \exp(-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})} \tag{5.6}$$

ただし，このモデルを用いるためには，あらかじめ係数$\boldsymbol{\beta}$を推定しておく必要がある．これには，マルコフ連鎖モンテカルロ法（MCMC：Markov chain Monte Carlo）法が用いられる．本研究においては，これを python のライブラリである PyMC を利用して行う．

MCMC 法では，まず，$\boldsymbol{\beta}$の事前分布を設定し，これと前述の尤度から事後分布を求める．この事後分布からマルコフ連鎖サンプリング法により生成された乱数からなるモンテカルロ標本を生成する．この標本からモンテカルロ法を用いて$\boldsymbol{\beta}$の平均値や HPD 区間などの事後統計量を求める．なお，PyMC では乱数の生成に NUTS[46]を使用している．

以上により，統計データから時間帯別と発生場所の環境別の犯罪発生の確率を求めることができる．

5.3.3　確信度の算出

危険予測の結果は, 危険事象が発生する危険性の度合いである確信度として出力する. この確信度に基づき, 危険事象の発生を未然に防止するための行動計画を決定する.

確信度は, 前述の前兆シーン認識結果と過去に発生した危険事象に関する統計データの分析結果を用い以下により求める.

ある時刻及び場所 j において，前兆シーンが認識されたとき，その確信度を Z_j とすると,

① 前兆シーンのクラス事後確率は，前兆シーン認識のみに基づいた犯罪が発生する確率を表しおり，それを r とする.

② 統計データ分析から得られたその時刻（時間帯）における犯罪発生確率を p（式（5.1）より）とする.

③ 場所 j における犯罪発生確率を q_j （式（5.6）より）とする.

④ p 及び q_i の最大値をそれぞれp_{max} ，q_{max} とする.

Z_j は，①～④より,

$$Z_j = r\frac{pq_j}{p_{max}q_{max}} \quad , \qquad 0 \le Z_i \le 1 \tag{5.7}$$

と表される.

なお，すべての事象は，独立であると仮定している．また， pq_jが非常に小さい値となるため$p_{max}q_{max}$により正規化している.

5.4　危険予測の適用例

具体的な統計データの数値例を用いた危険予測のための統計データの分析と確信度の算出例を示す.

参考とした統計データは,ある県でのコンビニエンスストアにおける犯罪（強盗）発生状況に関する資料である[47]. なお, コンビニの店舗数については,資料に記載がないので適宜設定した.また,強盗発生件数についても,資料では3年で79件となっているが,簡単のため年80件に変更し,また,ここでは, ロジット・モデル（式（5.6））を用いて確信度を算出することするので同一店舗が複数回被害に遭わなかったこととした.

【統計データ】

1．A県には，４００件のコンビニエンスストア（年中無休，24ｈ営業）があり，毎年約８０件の強盗が発生する.

2．強盗事案のデータ

（1）月，曜日，時間帯ごとの強盗発生件数

表5.1　月ごとの強盗発生件数

1月	2月	3月	4月	5月	6月
1	5	6	7	9	7
7月	8月	9月	10月	11月	12月
8	4	6	8	14	5

表5.2　曜日ごとの強盗発生件数

月	火	水	木	金	土	日
8	11	15	14	7	16	9

表 5.3　時間帯ごとの強盗発生件数

時間帯 1	時間帯 2	時間帯 3	時間帯 4
0 時～ 6 時	7 時～12 時	13 時～18 時	19 時～23 時
65	2	5	8

（2）被疑者の特徴

① 侵 入 口：正面出入口　（１００%）

② 変装方法: 帽子のみや帽子とマスクなどにより,顔を隠す変装(９４%)

③ 凶器の種類：包丁やナイフなどの刃物類　（８７%）

その他（１０%), 凶器なし（３%）

（3）強盗発生時の従業員数と客数

① 従業員数：　1～2人　（８７%）

② 客　　数：　0～3人　（９１%）

（4）防犯設備の利用状況

① 防　犯　ベ　ル　：　設置あり（８１%）➡　活用（１７%）

② 防犯カラーボール：　設置あり（９９%）➡　活用（３%）

（5）従業員の初動にかかる既遂（５５%)・未遂（４５%）

① 既遂：　初動あり　（0%）

② 未遂：　初動あり　（８３%）

※ 初動の内容：　防犯設備による対応, 取り押さえた, 犯人に抵抗, 大声で叫ぶなど

（６）店舗の立地条件よる強盗被害店舗（８０件），強盗被害を受けなかった店舗（非被害店舗）の特徴（３２０件）の特徴

① 立地環境：　店舗周辺の街灯・電灯の数，人通り，隣接道路幅など．（属性：良い／悪い）

「良い」の例：　街灯があり，見通しが良く，人通りも少なくない．

② 立地地域：　店舗が立地している用途地域　（属性：住宅系／商業系／工業系／指定なし）

表 5.4　立地環境（良い/悪い）別による店舗割合

属性	強盗被害店舗	非被害店舗
良い	50%	65%
悪い	50%	35%

表 5.5　立地地域（用途地域）別による店舗割合

属性	強盗被害店舗	非被害店舗
住宅系	32%	41%
商業系	39%	29%
工業系	7%	23%
指定なし	22%	7%

なお，参照した資料では，どの店舗が被害に遭ったのか，また，各店舗ごとの立地条件等が示されていないため，表 5.4，5.5 の割合でランダムに発生させた数値（店舗番号）に基づき，各店舗の属性データ等を生成した．

【統計データに基づく強盗発生確率の算出】

前述の統計データに基づき，コンビニ強盗が，どの時間帯にどんな環境に立地している店舗において発生しやすいのか，それを確率（以下，強盗発生確率と呼ぶ.）として求める.

１．月，曜日，時間帯別の各店舗に共通する強盗発生確率

確率変数　X_1: 発生月，X_2：発生曜日，X_3：発生時間帯,

$X_1 = 1$ 月,…,12 月　，　$X_2 =$ 月,…,金　，　$X_3 =$ 時間帯 1,…,時間帯 4

とすると,あるX_1, X_2, X_3 においてコンビニ強盗が発生する確率 p は,

$$p = P(X_1, X_2, X_3)$$

により求めることができる.

たとえば，p の最大値を p_{max} とすると，表 5.1，5.2，5.3 より，

$$p_{max} = \max \ P(X_1, X_2, X_3) = 0.028$$

となる.

このとき，$X_1 = 11$ 月，$X_2 =$ 土曜日，$X_3 =$ 時間帯 1，となり，最も強盗が発生しやすいのは，11 月の土曜日の午前 0 時から 6 時の間ということになる.

また，同様に p の最小値を p_{min} とすると，表 5.1，5.2，5.3 より，

$$p_{min} = \min \ P(X_1, X_2, X_3) = 2.734 \times 10^{-5}$$

となる.

このとき，$X_1 = 1$ 月，$X_2 =$ 金曜日，$X_3 =$ 時間帯 2，となり，最も強盗が発生しないのは，1 月の金曜日の午前 7 時から 12 時の間ということが分かる.

２.立地条件別の店舗ごとの強盗発生確率

店舗の立地条件により，強盗発生確率が異なると仮定し，5.3.2 で述べた GLM を用いて各店舗別の強盗発生確率 q_i を求める．

i は，店舗番号を表し，店舗数を４００店としているので，$i = 1,\cdots,400$ である．

q_i はベルヌーイ分布にしたがうことから，式（5.6）のロジット・モデルを用いる．

$$q_i = \frac{1}{1 + \exp(-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})} \tag{5.6}$$

まず，説明変数 $\boldsymbol{x}_i$ は，各店舗の立地条件であり，

$$\boldsymbol{x}_i = \begin{pmatrix} 1 \\ x_{1i} \\ x_{2i} \end{pmatrix},$$

とする．なお，説明変数 $\boldsymbol{x}_i$ の第 1 項目の 1 はモデルの式に定数項が含まれていることを考慮したものである．

x_{1i} は，店舗の立地環境とし，その値を

$$x_{1i} = \begin{cases} 1,（良い） \\ 2,（悪い） \end{cases}$$

とする．

また，x_{2i} は，店舗の立地地域（用途地域）とし，その値を

$$x_{2i} = \begin{cases} 1,（住\ 宅\ 系） \\ 2,（商\ 業\ 系） \\ 3,（工\ 業\ 系） \\ 4,（指定なし） \end{cases}$$

とする．

次に， MCMC 法を用いて係数 $\boldsymbol{\beta}$ を推定する．

係数 $\boldsymbol{\beta}$ を

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

とし，その事前分布には，式（5.8）の多変量正規分布を設定する．

$$\boldsymbol{\beta} \sim N_3(\boldsymbol{\beta}_0, \boldsymbol{A}_0^{-1}) \tag{5.8}$$

ただし，

$$\boldsymbol{\beta}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{A}_0 = \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}$$

とする．

また，被説明変数 y_i を

$$y_i = \begin{cases} 1, \text{（強盗発生）} \\ 2, \text{（発生せず）} \end{cases}$$

とすると，この場合の尤度は，次式のとおり，

$$p(D|\boldsymbol{\beta}) = \prod_{i=1}^{400} q_i^{y_i}(1-q_i)^{1-y_i}, \qquad D = (y_1, \cdots, y_{400})$$

となる．

以上の事前分布と尤度から，事後分布を求め，ＭＣＭＣ法により得られた事後統計量から，係数 $\boldsymbol{\beta}$ を点推定する．

推定結果を図 5.2 及び図 5.3 に示す．

	mean	sd	hpd_3%	hpd_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	r_hat
β[1]	-2.000	0.487	-2.903	-1.074	0.001	0.001	141997.0	140332.0	1.0
β[2]	0.140	0.267	-0.361	0.642	0.001	0.001	152366.0	136328.0	1.0
β[3]	0.198	0.125	-0.039	0.431	0.000	0.000	165673.0	160236.0	1.0

図 5.2 係数 $\boldsymbol{\beta}$ に対する事後統計量

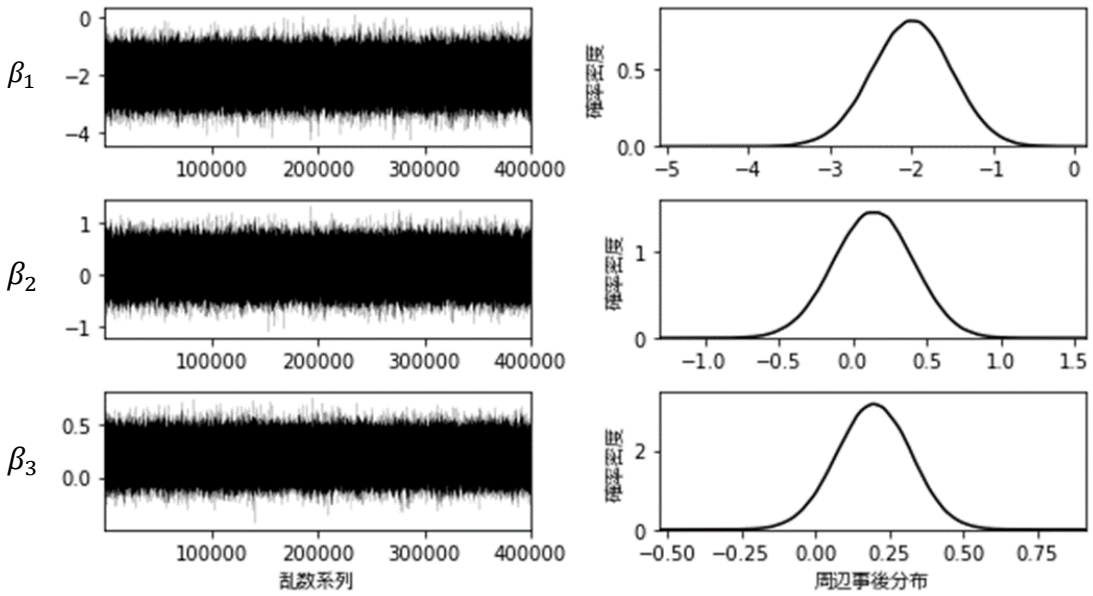


図 5.3 係数 $\boldsymbol{\beta}$ 乱数系列のプロットと周辺事後分布

図 5.2 に示した主な事後統計量について説明する. mean は係数 $\boldsymbol{\beta}$ の各要素の平均値, sd はそれらの標準偏差である. hpd_3%及び hpd_97%は 94%HPD 区間の下限，上限を示している．mean の値がこの HDP 区間に入っており, 概ね良好な結果であると言える. r_hat は, Gelman-Rubin 収束判定[48]の結果である．この判定基準では，サンプル系列間の分散が小さくなれば r_hat が 1 に近づくとされ，図 5.2 の r_hat はい

ずれも 1.0 なので，MCMC 法によるサンプリングはうまく収束したと判断できる．

以上のことから，係数 $\boldsymbol{\beta}$ の点推定結果は，その平均値とすることとし，

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 = & -2 \\ \beta_2 = 0.140 \\ \beta_3 = 0.198 \end{pmatrix}$$

という値が得られた．

これにより，各店舗の立地条件 $\boldsymbol{x}_i$ を前述の式(5.6)のロジット・モデルに代入することにより，店舗ごとの強盗発生確率 q_i を求めることができる．

ここで，q_i の集合を $Q = \{q_1, \cdots, q_{400}\}$ とし，q_i の最大値を q_{max} とすると，

$$q_{max} = \max \ Q = 0.283$$

となる．

このとき，$x_{1i} = 2$（悪い），$x_{2i} = 4$（指定なし），となり，最も強盗が発生しやすい店舗は，立地環境が悪く，立地地域（用途地域）の指定がない立地条件の店舗であることが分かる．

また，同様に q_i の最小値を q_{min} とすると，

$$q_{min} = \min \ Q = 0.160$$

となる．

このとき，$x_{1i} = 1$（良い），$x_{2i} = 1$（住宅系）となり，最も強盗が発生しない店舗は，立地環境が良く，立地地域（用途地域）が住宅系である店舗であることが分かる．

なお，各店舗の立地条件を考慮しない場合は，A 市全体での強盗発生確率は，0.2（＝80 件/400 店舗）であることから，店舗の立地条件により，その発生確率が変化することが分かる．

３．強盗発生確率

上記１，２の時間帯別及び店舗別の強盗発生確率を用いて，店舗 i におけるある時間帯の強盗発生確率は，pq_i により求めることができる．

また，A 市においては，「11 月の土曜日の午前 0 時～6 時の時間帯に，立地環境が悪く，用途地域の指定のない場所に立地している店舗において，強盗発生確率が最も高くなることから，

$$p_{max}q_{max} = 0.028 \times 0.283 = 0.008$$

となる．

【確信度の算出】

○ 前兆シーン認識

ある店舗において，ある時間帯に，図 5.4 に示すような，不審人物が認識されたものとする．また，そのときの前兆シーン認識によるクラス事後確率 r は，

$$r = 0.953$$

であったとする．

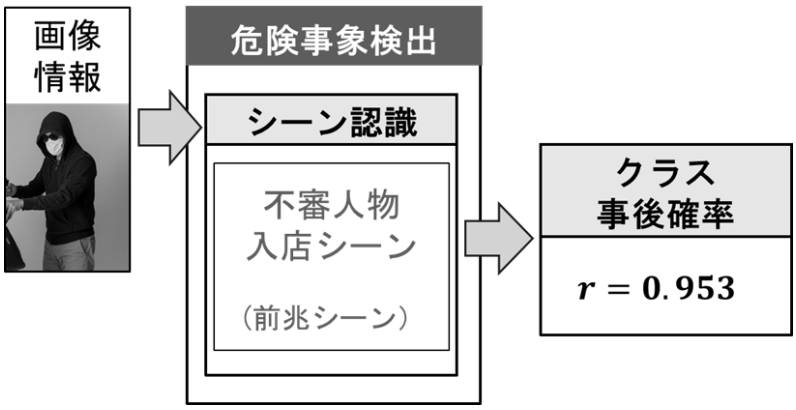


図 5.4 コンビニ強盗における前兆シーン認識

例 1

不審人物が認識されたのは，１月の金曜日，時間は午前１０時で，店舗 i の立地環境は良好で，住宅街に立地している．

以上のことから，統計データに基づいた強盗発生確率は，

$$pq_i = 2.734 \times 10^{-5} \times 0.160 = 4.374 \times 10^{-6}$$

確信度 Z_i は，

$$Z_i = 0.953 \times \frac{4.374 \times 10^{-6}}{0.008} = 5.211 \times 10^{-4}$$

となる．

このことから，危険予測においては，前兆シーン認識では「不審者」と認識されたが，確信度から，この人物がコンビニ強盗犯ではないと判断するのが自然であり，コンビニ強盗が発生する危険性はほとんどないと予測できる．したがって，この予測結果に基づき，行動計画では，店内巡回などの軽微な防犯対策を講じるか，あるいは防犯対策を取らないという選択肢も考えられる．

例 2

不審人物が認識されたのは，１１月の土曜日，時間は午前 3 時で，店舗 j の立地環境は悪く，用途指定のない地域に立地している．

以上のことから，統計データに基づいた強盗発生確率は，

$$pq_j = 0.028 \times 0.283 = 0.008$$

確信度 Z_i は，

$$Z_j = 0.953 \times \frac{0.008}{0.008} = 0.953$$

となる．

このことから，危険予測においては，前兆シーン認識で「不審者」と認識された人物は，確信度から，コンビニ強盗犯である可能性が非常に高くコンビニ強盗が発生する危険性が非常の高い予測され，十分に警戒する必要があると言える．したがって，この予測結果に基づき，行動計画では，店員による不審者への声掛けなどの厳重な防犯対策を講じる必要があると考えられる．

5.5 むすび

本章では，危険予測に用いる，危険事象が発生す可能性がある場合にその発生前に起こる事象（前兆現象）を検出するための前兆シーン認識の手法について述べた．しかしながら，前兆シーンがどの程度確からしいかという前兆シーン認識によって得られるクラス事後確率の推定のみでは，危険予測は難しい場合があり，これまで発生した危険事象に関する統計データを合わせて用いることにより，リアルタイムで精度の高い危険予測が可能であることを示した．

ただし，このような統計データが得られない，あるいは存在しない場合は，危険予測が可能でない場合とするか，前兆シーン認識のクラス事後確率をヒ

ューリスティックに判断し危険予測に用いるか，など，システム設計者が判断することになると考えられる．いずれにしても，本システムでは，統計データが存在しないことは想定していない．

危険予測では，前兆シーン認識によるクラス事後確率と統計データから得られる危険事象発生確率に基づき，確信度と呼ぶ危険事象発生の危険性の度合いを求めることにより，行動計画において，その確信度に応じた対策を講じることが可能となることを数値例を示しながら明らかにした．

従来，危険予測に画像認識を用いる場合は，画像認識の精度を向上させることによって予測精度を向上させようとする試みが行われてきた．また，統計モデルを用いた場合は，前節の例で示したとおり，ある特定の時間帯や場所などにおける危険事象発生確率を求めることにより危険予測を行うことが可能であるが，その予測には時間的幅などがあり，ピンポイント（不審者の入店したその瞬間など)でリアルタイムに危険予測を行うことは難しかった．

しかしながら，本システムにおいては，統計データ分析による危険予測の弱点をシーン認識のリアルタイム性によって補完することにより，信頼性の高い危険予測を実現している．

第６章
安全を確保するための行動計画

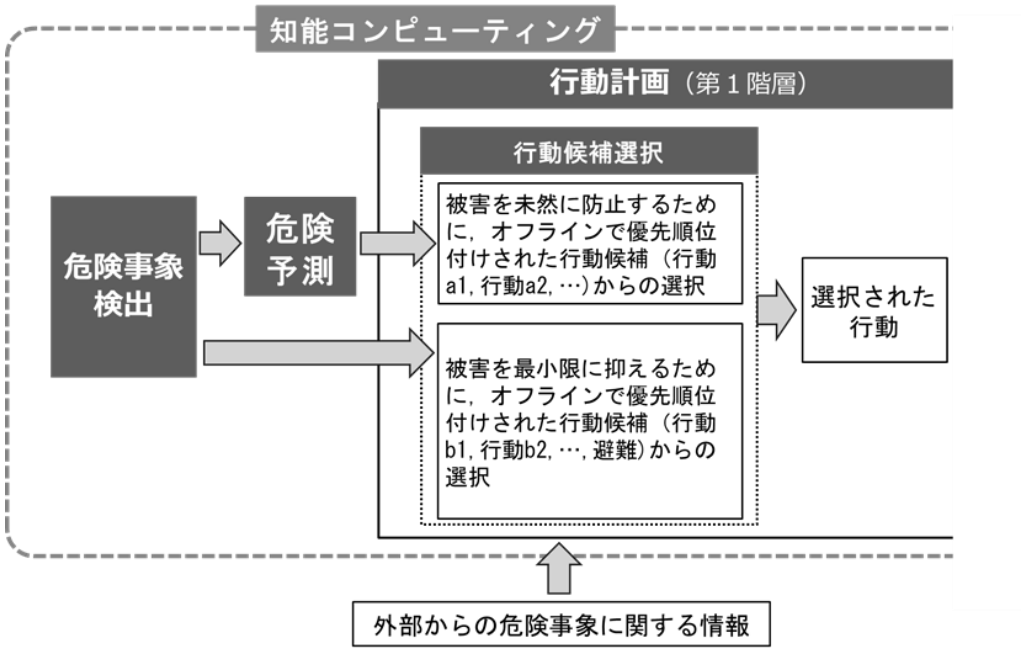
6.1 まえがき

高安全知能コンピューティングシステムおける行動計画では, 身体的損傷や経済損失などの被害をできるだけ小さくするための行動を決定し, 支援するという重要な役割を果たすものであり, 危険事象が発生した場合, あるいはその発生が予測される場合には, ユーザが危険を回避し安全な状態を確保するための行動を, この行動計画に基づいて取ることが可能となる. そのためには, 様々な危険事象に対して共通に利用できる高安全知能コンピューティングシステムにおける行動計画のプラットフォームを実現する必要がある.

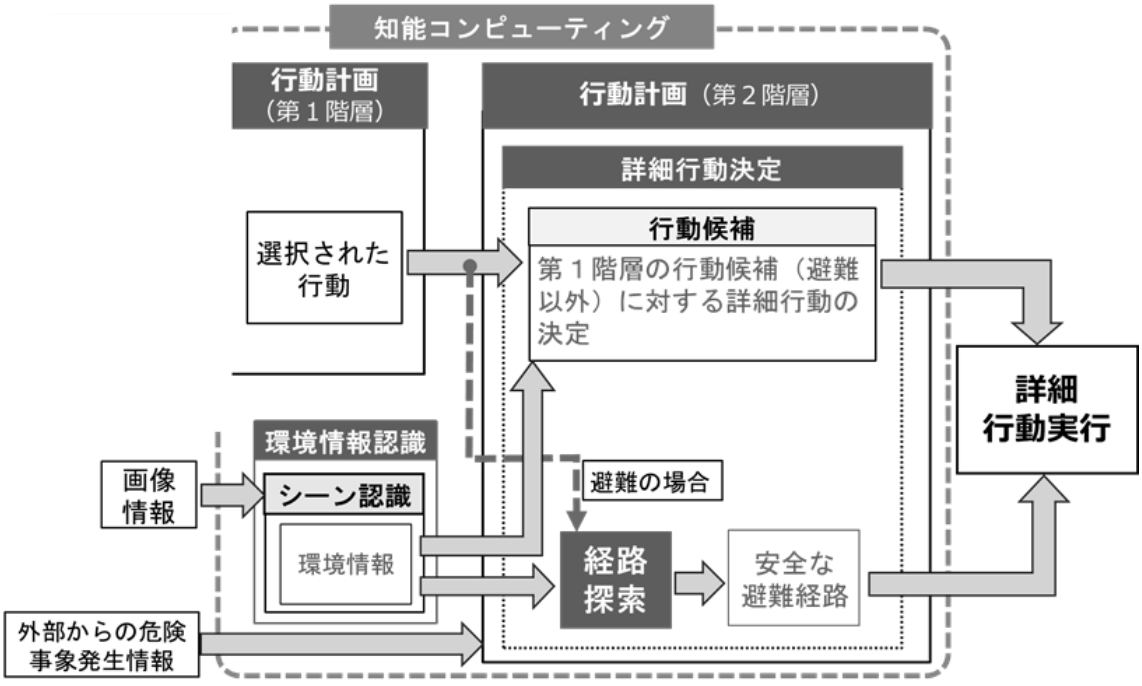
本章では, コンビニ強盗, 銀行強盗, スーパーでの万引き, テロ, 暴力事件, 急病, 火災, 津波, 洪水, 崖崩れを含む自然災害などの多様な危険事象に対して, 利用場所, 設置環境及びユーザ固有の条件に合わせて人間の安全を支援する行動計画のプラットフォームの実現方法を提案する.

6.2 高安全知能コンピューティングシステムにおける行動計画の構成

本システムにおける行動計画をプラットフォームとして機能させるため, その知能処理の階層化を提案している. 行動計画は 2 階層に分類でき, 優先順位付けされたグローバルな行動候補を選択する第 1 階層とその詳細行動を決定する第 2 階層からなる. これを図 6.1 に示す.



(a)　行動計画の第 1 階層での行動候補の選択



(b)　行動計画の第 2 階層での詳細行動選択，避難

図 6.1　行動計画における知能情報処理の階層化

第 1 階層では，警告，通報，避難などの危険事象の発生を未然に防止するため，あるいは，被害を最小限に抑えるための複数種類の行動（以下，行動候補と呼ぶ.）を列挙しておき，さらに，これらの行動候補の優先順位をあらかじめ定めておく. リアルタイムで危険事象が検出されたとき，行動候補の中から最も安全が確保されると思われる行動を選択する. ただし，危険予測が可能な場合には，その予測に基づき得られる危険性の度合いである確信度に応じてどの行動候補を優先的に実行するかを選択する. 行動候補の優先順位付けについては，特に危険予測に対する行動の効果を示す十分な統計データなどの入手が困難なため，予測される確信度を優先順位付けの基準に含めた階層分析法を用いて主観的に行う.

第 2 階層では，第 1 階層で選択された行動候補を実行するために必要となる詳細行動や初期状態から出発し安全な最終状態へ遷移するために，シーン認識により得られた環境条件の制約下での詳細行動を決定する. 特に第 1 階層で選択された行動の中で危険を回避するための行動として最も重要となるのは避難行動であり，その詳細行動である避難経路の決定は，実世界環境のリアルタイム情報が常に変化するため，あらかじめ決定しておくことが難しい. したがって，リアルタイムで到来する経路情報に基づき，被害にあわないような安全な避難経路をリアルタイムで探索し決定しなければならない. このような避難経路決定問題に対しては，時々刻々変化する経路情報をシーン認識等により捉え，安全な経路を決定していく.

なお，行動計画に基づき，危険事象の発生を未然に防止する行動を実行したにもかかわらず危険事象の発生防止に失敗した場合や前兆シーン認識が不可能であると共に危険事象が発生してしまった場合は，被害を最小限に抑えるための行動が選択される. また，津波などの危険予測や危険事象検出はシーン認識のみでは困難であり，外部たとえば気象庁などからの危険事象情報を取得し，被害を最小限に抑えるための避難などの行動を選択する.

6.3　行動計画の各階層における知能情報処理

行動計画では，危険事象の検出情報，危険事象発生時の環境情報，あるいはインターネットからの外部情報に基づき，第 1 階層において複数種類の行動候補の中から最も安全が確保されると思われる行動を，図 6.1(a)に示すようにリアルタイムで選択する．また，危険予測が可能な場合には，確信度に応じ，被害を未然に防止する行動が選択される．行動候補については，あらかじめオフラインでその優先順位を決定しておく．それらの行動候補には，たとえば，強盗などの犯罪が発生した場合の防衛行為や通報，けが人や急病人などが発生した場合の応急措置，火災や津波からの避難などがあげられる．

図 6.1(b)の第 2 階層では，第 1 階層で選択された行動に対する詳細行動が決定される．詳細行動についても，行動候補と同様にあらかじめオフラインで決定しておく．詳細行動には，たとえば，防衛行為では防犯器具使用の手順教示など，応急措置では周囲の人々による可能な救命処置などがあげられる．しかしながら，行動計画の第 1 階層で避難行動が選択された場合，経路情報が時々刻々変化するため，避難経路をあらかじめオフラインで決定しておくことは困難である．たとえば，大地震が発生し津波の到来が予想された場合，津波到達時間や道路の被害情報などに基づき，安全な避難経路をリアルタイムで探索し決定しなければならない．このような避難経路決定問題に対しては，図 5.1(b)に示すとおり，変化する避難経路の状況をシーン認識などにより捉え，その情報に基づき経路探索を行うことで安全な経路を決定していく必要がある．

以上のように，図 6.1 に示す行動計画は，種々の危険事象に対するための知能情報処理に共通する構成となっている．したがって，危険事象の種類が異なっても図 2.1 の構成に対応させて高安全知能コンピューティングシステムを実現できるという意味で，共通に利用できるプラットフォームとなっている．以下に行動計画の適用例を示す．

6.3.1　コンビニ強盗発生時における行動計画の適用例

○問題設定

シーン認識により，コンビニ強盗の発生が検出された．数名の客がいたが，けが人は発生していない．また，行動計画で示される安全行動は，客や従業員の安全を最優先に考えるものとする．

図 6.2 に示すとおり，行動計画においては，以上の情報から次の知能処理が行われる．

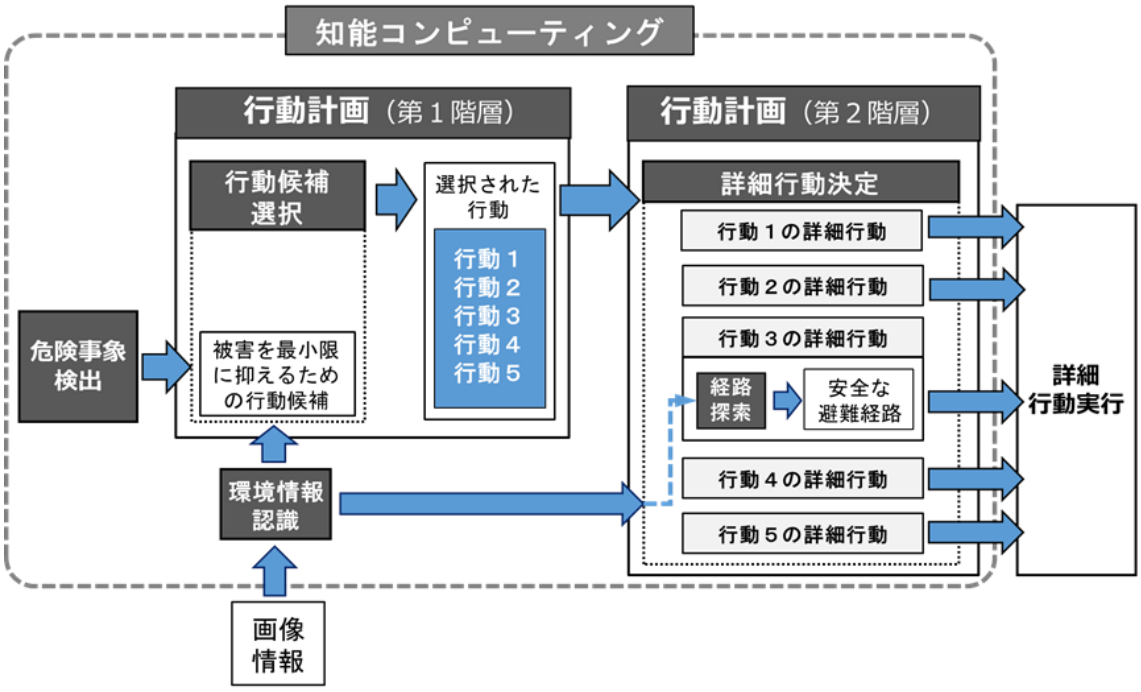


図 6.2　コンビニ強盗発生時における行動計画の処理フロー

①　第１階層において，優先順位付けされた行動候補を選択する．

・行動１：警報

・行動２：通報

・行動３：避難（客）

・行動４：防衛

・行動 5：防犯設備の使用

犯人への抵抗は行わず，安全を確保できる以上の 5 種類の行動（行動に付された番号が優先順位とする.）が選択される.

② 第 2 階層において，行動 1～5 を実行するための詳細行動を決定する.

・行動 1 の詳細行動：従業員に警報器（ベル，パトランプなど）の起動を指示，起動スイッチの場所を教示．あるいは自動起動.

・行動 2 の詳細行動：従業員に通報設備の起動を指示，起動スイッチの場所を教示．あるいは自動起動（この場合は，警備会社へ通報→警備会社が監視カメラで強盗発生を確認→警察へ通報となる.）.

・行動 3 の詳細行動：従業員に客の避難誘導を指示．店内に設置された複数台のカメラの映像による環境情報認識（シーン認識）に基づき，犯人の位置情報に応じて犯人と接触しない安全な避難経路を探索･教示. なお，避難経路探索の詳細な適用例については，6.5.2 において述べる.

・行動 4 の詳細行動：従業員にレジから離れるよう指示．必要があれば防犯器具（さすまたなど）の使用を指示，器具の設置場所を教示.

・行動 5 の詳細行動：従業員に犯人の退店時に防犯設備を使用する（犯人へカラーボールを投げつける.）よう指示，設備の設置場所を教示.

なお，犯人の退店後の逃走方法や方向などを確認するよう指示する.

このように，本システムを導入したコンビニエンスストアでは，コンビニ強盗の発生にともない，冷静さを失いパニック状態になっていると推測される従業員や客に対し，安全を確保するための適切な指示，支援を行うことが可能となり，ひいては，早期の犯人検挙へもつながる．

6.3.2　津波発生時における行動計画の適用例

○問題設定

大きな地震が発生し，大津波警報が発表された．防災無線，テレビ，インターネットなどの外部から危険事象発生情報として，津波到達時間，浸水区域などの情報を入手した．システムユーザの位置情報（海沿いの自宅）及びシーン認識により得られる環境情報から，周囲にはまだ津波が到達していないこと，また，屋外の安全な場所への避難が可能であるこのも確認できた．なお，システムユーザは本システムの端末（スマートフォンなど）を携帯しているものとする．更に以下の（1），(2)の状況を設定する．

(1)システムユーザが健常者の場合

ケガ等もしておらず，普段どおり行動が可能である．

(2)システムユーザが高齢等で自力での移動が容易ではない場合

システムユーザは，事前に，主観的な自身の体調，津波到達時間などの情報を考慮した行動候補の優先順位付けをしおく．なお，主観的な行動候補の優先順位付け手法については，6.4 において詳述する．

以下に体調「普通」，津波到達時間「避難に十分な時間あり」などの場合の優先順位付けの例を示す．（行動候補に付された番号が優先順位とする．）

行動候補 1：避難支援等関係者（近隣住人等）の支援を受け避難

行動候補 2：自力で安全な場所まで避難

行動候補 3：自宅に留まる（垂直避難を含む）

行動計画においては，以上の情報から図 6.3 に示す知能処理が行われる．

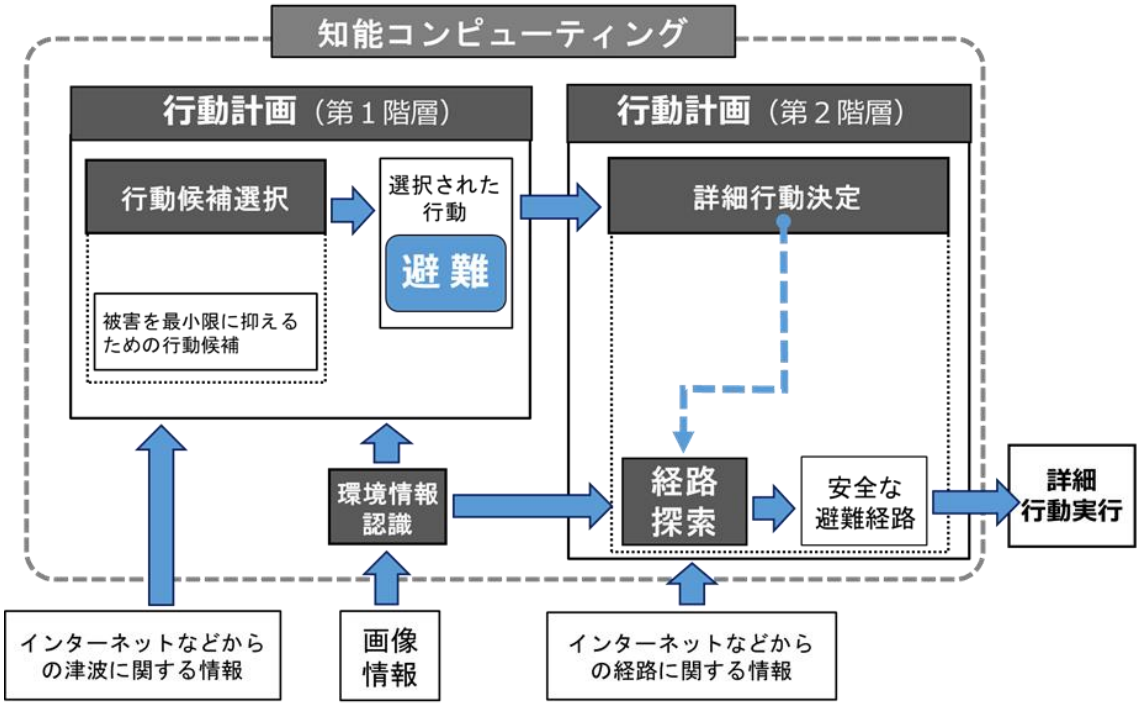


図 6.3　津波発生時における行動計画の処理フロー

①　第 1 階層において，優先順位付けされた行動候補を選択する．

問題設定の(1)の場合

行動：津波浸水区域外への避難

津波に関する外部からの情報と津波が現在地周辺まで到達していないこと（環境情報認識）から津波浸水区域外への避難を選択．

初期設定は徒歩避難（負傷した場合など，ユーザの状態に合わせてリアルタイムで変更可）．

津波浸水区域外への避難を最優先としている．ただし，第 2 階層における詳細行動決定において，津波浸水区域外への避難が困難となった場合は，近くの高いビル（津波避難タワーを含む）に避難することを優先し，更にそれも不可能な場合は，現在地周辺の頑丈な建物に避難することを教示する．

問題設定の(2)の場合

行動：近隣住人等の支援を受け津波浸水区域外への避難

警報発令時の体調（普通），津波到達時間（時間あり）などのリアルタイム情報から行動候補 1 により，車いすを押すなどの支援を受けながら津波浸水区域外へ避難するが選択される．

② 第 2 階層において，避難を実行するための詳細行動を決定

環境情報認識（シーン認識）により，通行不能個所などの避難経路情報から，安全に避難可能な最短経路をリアルタイムで探索・決定し，システムユーザへ教示する．なお，避難経路探索の適用例については，6.5.2 において詳述する．

このように，シーン認識を用いてリアルタイムで避難経路情報を得ることにより，安全に最短で目的地まで到達可能となる．

6.4 行動候補の優先順位付け

第 1 階層における行動候補の選択において問題となるのは，各行動候補にあらかじめ設定しておく必要のある優先順位である．危険事象に対応する安全対策の効果を最大限に得るためには，この行動候補の優先順位付けが重要であることは言うまでもない．

優先順位付けについては，過去に発生した危険事象に対する行動やその発生を未然に防止する行動の効果を示すデータが多く存在する場合は，統計モデルリングや機械学習により，その対策の効果を分析することが可能となる．これに基づき安全性が高いと推定される行動候補の優先順位をオフラインであらかじめ決定しておくことは容易である．たとえば，強盗，窃盗などの危険事象発生後の事例においては，優先順位が最も高い行動候補は「関係機関へ通報する」となり，次に優先順位が高い行動候補は「防犯設備を稼働させる」となる．また，急病人やけが人の発生後の事例おいては，優先順位が最

も高い行動候補は「救急へ通報する」となり，次に優先順位が高い行動候補は「必要に応じて周囲の人々が AED などにより救命処置を行う」となる．

一方，危険事象に対する行動やその発生を未然に防止する行動の効果などを示すデータが少ない，あるいは存在しない場合は，統計モデルリングや機械学習により行動候補の優先順位をあらかじめ決定することが困難である．たとえば，前兆シーン認識において不審者と認識された人物の何パーセントが犯罪行為を実行するのか．また，不審者と認識された人物に対し取った犯罪を未然に防止する行動の効果がどれほどあったのか．あるいは，その人物が犯罪の実行を考えていたのか．といったことをインタビューして確認することは不可能であり，それらのデータを得ることが困難である場合が多いと考えられる．

このような場合は，行動候補の立案者となるシステム設計者が専門家等の知見を生かし，費用対効果なども考慮しながら，危険事象に対し効果が高いと思われる行動候補を検討せざるを得ない．したがって，その優先順位付けについては，システム設計者が多基準意思決定手法を用い主観的に決定する必要がある．本研究では，最も基本的なモデルである階層分析法（Analytic Hierarchy Process：AHP）[49，50]を用い数理的に意思決定を行う手法を提案する．

この提案手法の有効性を示すため，以下に，危険予測においてコンビニ強盗が高い確率で発生すると予測された場合，すなわち，コンビニ強盗の確信度が非常に高い場合の AHP によるコンビニ強盗未然防止行動の優先順位付けの適用例を示す．

6.4.1　AHP を用いた行動候補の優先順位付け手法

（１）優先順付けのためのＡＨＰ

行動候補の優先順位付けするための 3 階層からなる AHP の階層構造を図 6.4 に示す．

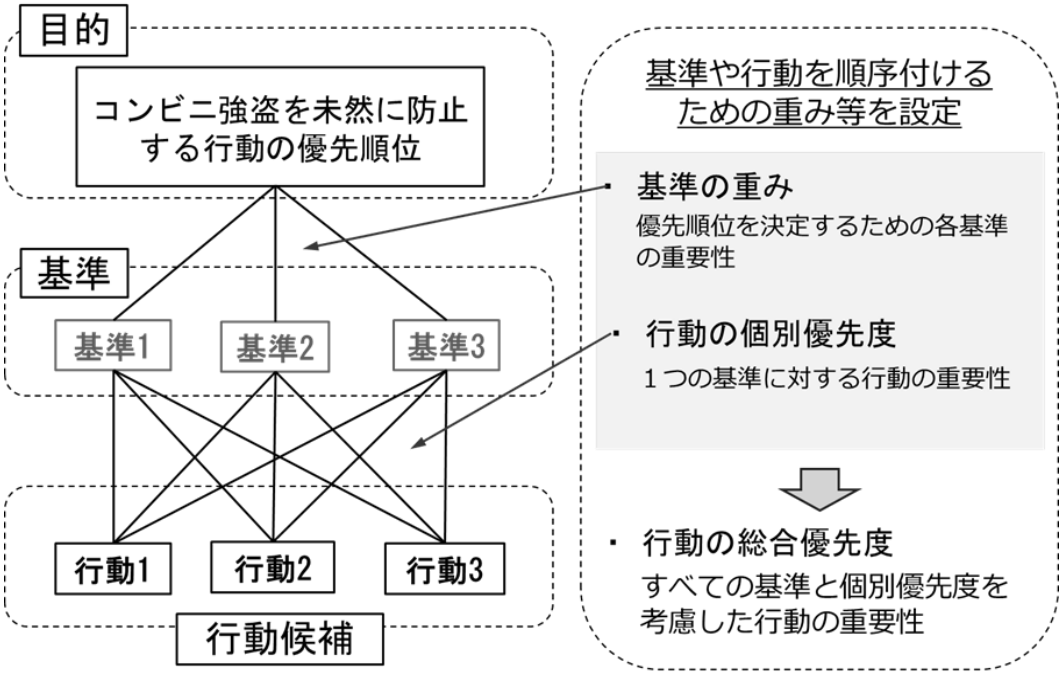


図 6.4　コンビニ強盗発生予測（危険予測）における AHP の階層構造

第 1 階層の「目的」は，コンビニ強盗発生を未然に防止する行動候補の優先順位付けである．

第 2 階層の目的達成のための「基準」は以下のとおりとした．

①基準 1 :強盗発生の危険性（確信度）

②基準 2 ：容易に行動することが可能（容易性）

③基準 3 :少ない経費で行動することが可能（経済性）.

各基準は,行動の優先順位を決定するための重要性に応じ重み付けされる．また．１つの基準に対する行動の重要性に応じ，個別の優先度を付ける．

第 3 階層の「行動候補」は，次のとおりとする．

・犯行を思いとどまらせることを主眼に置いた防犯行動を取ることとする．

・防犯器具（さすまたなど）の準備などの防衛行動は，店員への負担や危険性を配慮し，積極的には取らないものとする．

行動 1 ：待機店員によるレジ応援

行動２：防犯用の店内放送

行動３：屋外の回転灯の点灯

また，これらの防犯行動は統計データの分析結果を活用することが重要である[48]．たとえば，行動１については，強盗発生時の従業員数の統計データを参考としている．データでは従業員が２名以下の時もっと多く強盗が発生している．

なお，第３階層は一般に代替案と呼ばれているが，ここでは行動候補と呼ぶこととする．すべての基準の重みと行動の優先度を考慮した総合的な優先度を求め，行動の優先順位を付ける．

（２）優先順位の決定方法

基準の重みと行動の個別優先度については，一対比較法により決定する．一対比較に用いられる重要性の数値尺度とその意味については，表 6.1 に示すとおりである．

表6.1　AHPの重要性の比較尺度

定　義	数値尺度
同じくらい重要（equal importance）	1
少し重要（weak importance）	3
かなり重要（strong importance）	5
非常に重要（very strong importance）	7
きわめて重要（absolute importance）	9

※ 2，4，6，8は中間的な判断に用いる．

この比較尺度に基づき, 基準と行動における $n \times n$ の一対比較行列を生成し, 固有値法を用い比較行列の固有ベクトルを基準の重みあるいは個別の行動の優先度とする. これらの重みベクトルと個別優先度ベクトルをかけ合わせることで, 行動の総合的な優先度を求め, これにより, 行動の優先順位を決定する.

まず, 一対比較により基準あるいは行動に対する比較行列 $\boldsymbol{A}_i$ を生成する. 本適用例では, 確信度は非常に高く, コンビニ強盗が発生する危険性が高い場合を想定している.

表 6.2「基準」における比較行列

	確信度	容易性	経済性
基準 1 ：確信度	1	7	5
基準 2 ：容易性	1/7	1	1/3
基準 3 ：経済性	1/5	3	1

式にすると,

$$\boldsymbol{A}_0 = \begin{pmatrix} 1 & 7 & 5 \\ 1/7 & 1 & 1/3 \\ 1/5 & 3 & 1 \end{pmatrix}$$

$\boldsymbol{A}_0$ の基準に対する比較行列では, たとえば, 1 行 2 列目の「7」は, 確信度が非常に高い場合, コンビニ強盗発生未然防止行動を容易に取れることよりも, 強盗が今まさに起きようとしているという危険性を, 行動候補の優先順位を付けにおいては, 非常に重視しているということを表している.

同様にして, 各基準に対する行動の比較行列を生成する.

表 6.3「確信度」に対する行動の比較行列

	行動 1	行動 2	行動 3
行動 1	1	3	2
行動 2	1/3	1	1/3
行動 3	1/2	3	1

$$\boldsymbol{A_1} = \begin{pmatrix} 1 & 3 & 2 \\ 1/3 & 1 & 1/3 \\ 1/2 & 3 & 1 \end{pmatrix}$$

表 6.4「容易性」に対する行動の比較行列

	行動 1	行動 2	行動 3
行動 1	1	1/9	1/5
行動 2	9	1	3
行動 3	5	1/3	1

$$\boldsymbol{A_2} = \begin{pmatrix} 1 & 1/9 & 1/5 \\ 9 & 1 & 3 \\ 5 & 1/3 & 1 \end{pmatrix}$$

表 6.5「経済性」に対する行動の比較行列

	行動 1	行動 2	行動 3
行動 1	1	1/9	1/9
行動 2	9	1	1
行動 3	9	1	1

$$\boldsymbol{A}_3 = \begin{pmatrix} 1 & 1/9 & 1/9 \\ 9 & 1 & 1 \\ 9 & 1 & 1 \end{pmatrix}$$

次に，比較行列から基準の重みや行動の個別の優先度を計算する．その手法には，固有値法，近似法などがあるが，ここでは，固有値法を用いる．

固有値法では，比較行列 $\boldsymbol{A}_i$ の最大固有値と対応する固有ベクトル $\boldsymbol{w}_i$ を求める．この固有ベクトルを重み等として使用可能かどうかの判断については，一貫性比率（CR）を指標として用い，その値が 0.1 より小さければ，その固有ベクトルを重みあるいは個別優先度とすることができる．

比較行列 $\boldsymbol{A}_0$における固有ベクトル及び CR は，

$$\boldsymbol{w}_0 = \begin{pmatrix} 0.731 \\ 0.081 \\ 0.188 \end{pmatrix}, \quad CR = 0.062$$

同様にして，

$$\boldsymbol{w}_1 = \begin{pmatrix} 0.528 \\ 0.140 \\ 0.333 \end{pmatrix}, \quad CR = 0.052$$

$$\boldsymbol{w}_2 = \begin{pmatrix} 0.063 \\ 0.672 \\ 0.265 \end{pmatrix}, \quad CR = 0.028$$

$$\boldsymbol{w}_3 = \begin{pmatrix} 0.053 \\ 0.474 \\ 0.474 \end{pmatrix}, \quad CR = 0.000$$

となり，いずれの CR も 0.1 より小さく固有ベクトルを重み等として利用できる．

最後に，これらの固有ベクトルを用いて，各行動の総合優先度を求める．

$$\begin{pmatrix} 0.528 & 0.063 & 0.053 \\ 0.140 & 0.672 & 0.474 \\ 0.333 & 0.265 & 0.474 \end{pmatrix} \begin{pmatrix} 0.731 \\ 0.081 \\ 0.188 \end{pmatrix} = \begin{pmatrix} 0.401 \\ 0.246 \\ 0.354 \end{pmatrix}$$

結果は，行動 1 の優先度が 0.401，行動 2 の優先度が 0.246，行動 3 の優先度が 0.354 となる．したがって，行動の優先順位は，総合優先度が高い順に，行動 1，行動 3，行動 2 の順番となる．

6.4.2　優先順位付けされた行動候補の実行

確信度にしきい値が設定されている場合には，あらかじめ，以上のような手法により，確信度が高い場合の優先順位付けされたコンビニ強盗を未然に防止する行動が行動計画の第 1 階層に設定され，また，同様にして，確信度が低い場合，あるいは確信度が中間的な値の場合にも，それぞれ確信度に応じ優先順位付けされた未然防止行動が行動計画の第 1 階層に設定される．

これにより，危険予測により，リアルタイムで得られる確信度に基づき，対応する行動計画の第 1 階層の行動候補が選択されることとなる．

本システムを導入している店舗では，前兆シーン認識により不審者が検出され，危険予測により高い確信度が得られた場合には，不審者を刺激しないようその行動を注視しながら，行動計画の指示に従い，行動 1，3，2 の順番でそれぞれの未然防止行動に対応する詳細行動を実行して行く．店員は不審者が店を出て行くまでその態勢を維持する．万が一，それらの未然防止行動が成功せず，コンビニ強盗が発生してしまった場合は，コンビニ強盗が発生してしまった場合の行動計画に基づき警察あるいは警備会社へ通報し，防犯設備の起動などを実行することとなる．

このような犯罪対策のシステムにおいては，標準化され取り扱いが簡単なシステムであること．また，定型化された防犯行動を指示できることなどが求められている[51]．本システムにおいては，様々な危険事象に対応した防犯行動をシステム設計者等の知見に基づき行動計画にあらかじめ具体的に定

めておくため,本システムにより自律的に適切な行動選択とユーザへの指示が可能となる.

6.5 詳細行動決定問題

図 6.1(b)に示すとおり,行動計画の第 2 階層では第 1 階層で選択された行動を実行するための詳細行動を決定する. この詳細行動は, 第 1 階層において選択される大まかな行動が決まれば, ほとんどの場合, 容易に決定が可能であると考えられる. たとえば, 第 1 階層において防犯器具による防衛行動が選択された場合, その器具の最適な取り扱い手順はマニュアル化されていることが一般的であり, システムユーザに対する指示・支援は容易である. (なお, 当該事案の適用例については, 6.3.1 において詳述しているので参照されたい.)

しかしながら, 第 1 階層において, 危険を回避するための最も容易で典型的な行動である「避難」が選択された場合の詳細行動である避難経路の決定は難しい問題であるため, 第 2 階層における例題として, 通行可能な経路を通過し安全な目的地まで最短で到達するための避難経路の探索を取り上げることとする.

6.5.1 状況の変化に対応した避難経路探索の条件

従来の一般的な避難経路探索では, 想定される被害の規模や地理情報システムなどを用い狭隘道路での渋滞などを考慮した避難場所までの到達可能性を評価することにより, 事前に避難経路の危険性を分析しておき避難経路を決定する手法が取られている[52].

本システムでは，シーン認識による環境情報認識に基づいて，リアルタイムで通行不能個所を検出し，動的に経路を探索することにより，最も安全と思われる避難経路をシステム利用者に提供するための手法を提案している．

したがって，動的に避難経路が変化することがあるため，ユーザは本システムの端末を携帯する必要がある．また，道路網や建物内の構造などの避難経路探索に用いる基本情報は，危険事象発生前にシステムに入力・設定されているものとする．また，交差点など道路や建物内の主要な場所に設置された監視カメラからの映像が得られ，端末では動画シーン認識により，環境情報たとえば，道路の損壊，火災，道路への障害物流入，道路の冠水，車の渋滞，超過密歩行困難状態などを認識する．これに基づき通行可能な経路は動的に変化するが，それを通過して安全な目的地まで最短で到達するための経路を探索する．

6.5.2　津波避難行動計画への適用例

（1）問題設定

・地震発生により津波警報が発表された．津波浸水区域の情報を取得できるものとする．

・浸水が予想される区域から避難する．

・システム利用者の現在地の情報は端末の GPS により取得される．

（2）基本情報の設定

・地図情報：道路網，監視カメラの設置場所，避難場所の情報が配信されている．

・避難場所：浸水域外にある安全となる地点

・避難方法：徒歩

・経路情報：通行可能あるいは通行不可

（３）避難経路探索の方法

地図情報から，監視カメラの設置場所をノードとし，ノード間の重みをそれらの間の経路長（非負）とする．ダイクストラ法により現在地点と目的地の最短経路の道のり（パス）を求める．ただし，シーン認識により得られた経路情報が通行不可の場合は，そのノード間の重みを 0 にして再探索を行い，最短経路のパスが得られる．

（４）津波避難のための経路探索

上記（１）から（３）に基づき，以下に示す条件 1～4 による避難経路探索の例を図 6.5 に示す．

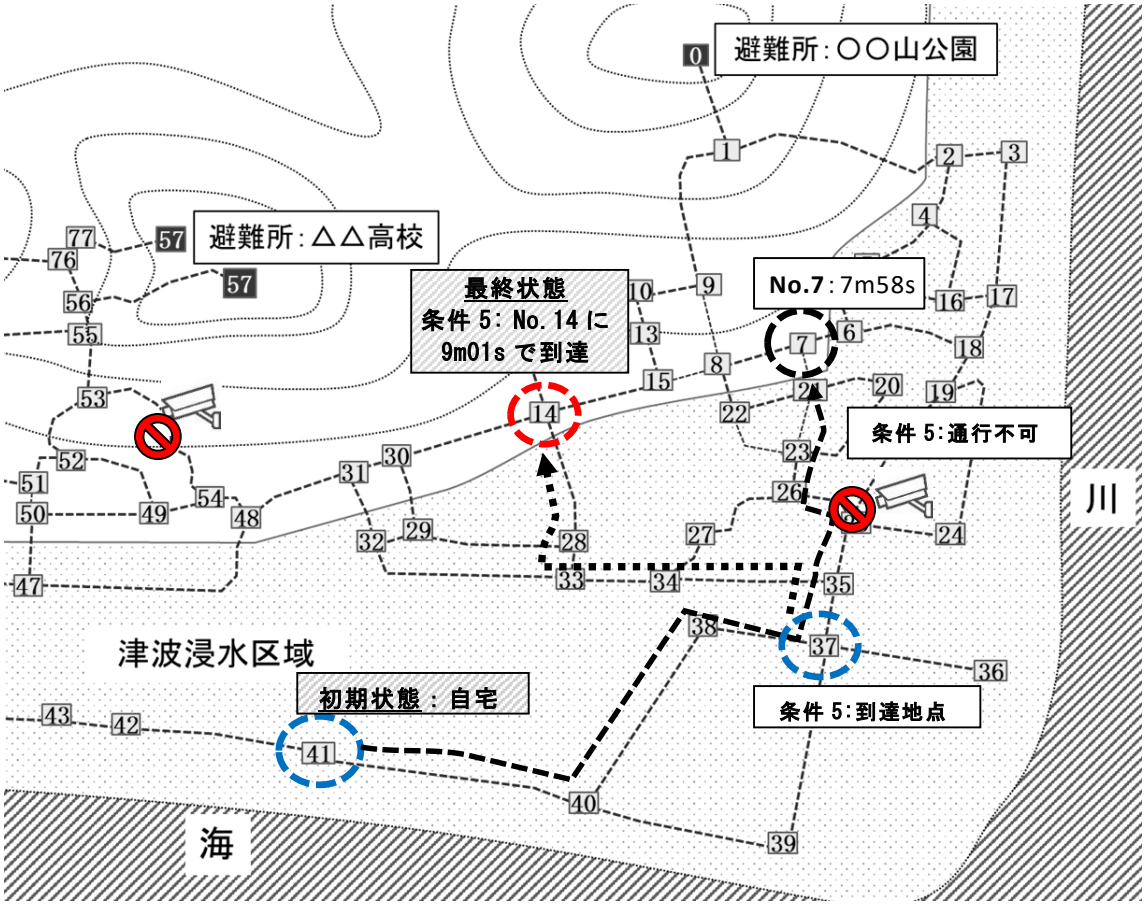


図 6.5　津波避難のための経路探索例

図中の数字はノード番号，ノード間の破線は経路である．また，津波浸水区域情報を表示している．

条件１：自宅はノード番号 41 にある．

条件２：自宅からの避難開始より 10 分で津波が到達する．

条件３：避難開始時に監視カメラの画像からシーン認識により，ノード番号 53 において崖崩れが発生し，通行不可であることを認識した．

条件４：転倒し足に軽傷を負ったため，歩行速度は 3 km/h である．

条件１～４に基づいて，自宅を出発した時点での浸水区域外に最短で到達可能な避難経路を探索して得られた，ノード番号のパスは以下のようになる．

41→40→38→37→35→25→26→23→21→7

ノード番号 7 への到達時間は 7 分 58 秒と見積もられる．

ノード番号 37 に到達した時点で余震等のため，条件５のような経路情報の変化が起こったとする．

条件５：ノード番号 25 において家屋が倒壊し，通行不可となった．

条件５に基づき避難経路を再探索すると，ノード番号 37 からのパスは以下のようになる．

37→35→34→33→28→14

ノード番号 14 への到達時間は，9 分 1 秒であり，条件２の津波到達時間前に浸水区域外に避難可能であることがわかる．浸水区域外に避難後は，更に高台のノード番号 0 あるいは 57 の避難所に，同様の経路探索手法を用いて避難することとなる．

なお，ノード番号 25 に到達した時点でそこが通行不可であると判明し，迂回して当初の目的地であるノード番号 7 へ向かおうとすると 10 分 35 秒かかってしまい途中で津波に巻き込まれてしまう．

このように，シーン認識を用いてリアルタイムで経路情報の変化を認識することにより，安全な目的地まで最短で到達することが可能となる．

6.6 むすび

高安全知能コンピューティングシステムでは，危険事象検出・予測用カメラあるいはその周辺に設置された環境認識用カメラの画像情報から危険事象が検出された場合や予測された場合には，被害を最小限に抑えるため，あるいは被害を未然に防止するための優先順位付けされた複数種類の行動候補を自律的に提示することにより，システムユーザの安全を支援する．行動候補として特に，危険から回避するための避難は最も典型的である．その詳細行動を決定するため，周辺環境中の避難経路候補に設置されたカメラからの画像情報を利用して通行不能かどうかを認識して最短で安全な目的地に到達できる避難経路を探索する．

以上のような行動計画問題は，優先順位付けされたグローバルな行動候補を選択する第 1 階層とその詳細行動を決定する第 2 階層からなる．第 1 階層においては行動候補の優先順位付けがあらかじめ必要となる．行動の効果を示す十分なデータの入手が困難なため，統計モデルリングや機械学習により優先順位付けすることが困難である場合，システム設計者が専門家等の知見や費用対効果なども考慮しながら，主観的に優先順位付けを行わなければならない．このため，危険予測から得られる確信度と AHP を用い数理的に意思決定を行う手法を提案した．

次に第 2 階層では，初期状態から出発して，環境条件の制約下において利用可能な行動を選択し，安全が保証される最終状態へ遷移するための具体的

な行動を決定する．一例として津波避難を取り上げ，シーン認識により得られる時々刻々変化する経路情報に基づき，通行不能な個所を検出し動的に経路探索することにより，最も安全と思われる避難経路をシステム利用者に提供できることを示した．

本システムは，提案手法を用いた行動計画により設置環境やユーザ固有の条件に合わせ，多様な危険事象に対して共通に利用できるプラットフォームとして，その有用性は高い．

第 7 章
結　言

本論文では, 第 2 章から第 6 章にわたり, 様々な危険事象に共通に活用可能なプラットフォームとしての高安全知能コンピューティングシステムの実現に必要となる，システムの構成及び知能情報処理について述べた.

第 2 章では, 高安全知能コンピューティングシステムが対象とする危険事象の範囲と本システムが危険事象検出, または危険予測及びこれらの処理結果に基づいた行動計画の 3 つの知能情報処理によって構成されることを示した. この構成により, 対象とする危険事象が変更されても, それに合わせてシステムの構成を変更する必要がないため, 本システムがプラットフォームとして機能することが明らかとなった.

第 3 章では, 危険検出のための高精度シーン認識の手法を提案した. この手法では, シーン認識には直接用いることができない物体カテゴリ認識のために学習させた CNN を画像の大域的な特徴抽出器として用い, そこから得られた CNN 特徴ベクトルを SVM で学習/識別することにより, 高精度の静止画を用いた単一シーン認識が可能であることを評価実験により示した.

また, 複数種類のシーンが発生する可能性のある生活空間における高精度の多クラス分類を実現するための手法として識別器選択方式による多クラス分類手法を提案した. 評価実験では，強盗，急病，その他シーンの各テストデータセットを用い，提案手法と一般に SVM の多クラス分類に用いられる

one-vs-one 方式による実験を行い，その結果，提案手法は各単一シーンの認識結果に匹敵する高い正解率を示したが，one-vs-one 方式は大きく正解率が低下した．これにより，提案手法が複数シーンの認識に有効であることが明らかとなった．更に，感情価を用いた多クラス分類の手法を提案し，学習コストの低減と評価実験により，この手法が識別器選択方式による多クラス分類と同等の認識精度を得られることを示した．

第 4 章では，静止画によるシーン認識のみでは検知できない，人間などの物体の動きによりはじめて異常な動作等であると認識可能となるシーンが存在するため，動画を用いたシーン認識の手法を考案した．これには，フレームとオプティカルフローの重畳画像を用いる．また，動画シーン認識の場合は，複数枚の重畳画像の識別結果（正例/負例）からシーン認識結果を決定する必要があるため，2 種類のシーン認識結果決定法を評価実験により比較した．その結果，認識周期を設定し識別結果の多数決を取る方式に比べ，隠れマルコフモデルを用い状態系列を推定する方式がより高精度にシーン認識が可能であることが明らかとなった．この結果は，単純には比較できないが最先端の認識手法と同等の性能を有している．

第 5 章では，前兆シーン認識によって得られるクラス事後確率の推定のみでは困難な危険予測を，危険事象に関する統計データを合わせて用いることにより，精度の高い予測が可能であることを示した．この危険予測により，確信度を求めることにより，行動計画において，それに応じた対策を講じることが可能となることを明らかにした．これにより，統計データ分析による危険予測の弱点をシーン認識のリアルタイム性によって補完することにより，これまで難しかった，リアルタイムでの危険予測が可能となった．

第 6 章では，行動計画を階層化することにより，様々な危険事象に対し，容易に適用が可能なプラットフォームとして機能することを示した．この行動計画は，優先順位付けされたグローバルな行動候補を選択する第 1 階層とその詳細行動を決定する第 2 階層からなっている．行動候補の優先順位付け

問題については, 行動の効果を客観的に示す十分なデータの入手が困難なため, システム設計者が専門家等の知見や費用対効果なども考慮しながら, 主観的に優先順位付けを行わなければならない場合, 確信度と AHP を用い数理的に意思決定を行う手法を提案した. また, 第 2 階層では, 詳細行動を決定するが, 一例として津波避難を取り上げ, シーン認識により得られる時々刻々変化する経路情報に基づき, 通行不能個所を検出し動的に経路探索する手法を提案し, この手法により, 最も安全と思われる避難経路をユーザに提示できることを示した.

以上のことから, 様々な危険事象に共通に活用可能なプラットフォームとしての高安全知能コンピューティングシステムの実現が可能であることが明らかとなった.

今後の研究課題としては, 高安全知能コンピューティングシステムの更なる多様な危険事象への応用があげられる. 本システムにおける危険事象検出には画像情報を用いているが, これに加え, 音声, 温度, 圧力, 臭いなどの情報も入力としたマルチモーダル情報処理により, 本システムの応用範囲の拡大を図る必要があると考える. また, 人間の安全・安心を預かることとなる高安全知能コンピューティングシステムは, 危険事象の見落としや不適切な行動計画の提示などの誤動作を極限まで低減させる必要がある. したがって, システムの信頼性や有用性を検証するための実証実験が必要性であると考えられる.

参考文献

[1] 向殿政男, “安全性に関する将来展望”, 日本信頼性学会誌, 信頼性 Vol. 3(7), pp.346-351 (2011).

[2] S. Mohammadi, A. Perina, H. Kiani and M. Vittorio, “Angrycrowds: Detecting violent events in videos”, In ECCV, pp.3-18 (2016).

[3] Y. T. Liao, C. L. Huang and S. C. Hsu, “Slip and fall event detection using Bayesian belief network”, Pattern recognition 45(1), pp.24-32 (2012).

[4] D. Pathak, A. Sharang and A. Mukerjee, “Anomaly Localization in Topic-Based Analysis of Surveillance Videos,” in IEEE Winter Conference on Applications of Computer Vision, pp.389-395 (2015).

[5] D. Van Hamme, P. Veelaert, W. Philips and K. Teelen, “Fire detection in color images using Markov random fields”, In: International Conference on Advanced Concepts for Intelligent Vision Systems, pp.88-97 (2010).

[6] H. L. Chen, M. Tsai, and C. Chan, “A Hidden Markov model-based approach for recognizing swimmer’s behaviors in swimming pool”, In: 2010 International Conference on Machine Learning and Cybernetics, vol. 5, pp.2459–2465 (2010).

[7] Kenichi Takada and Michitaka Kameyama, “High-Accuracy Scene Recognition and Its Application to Highly-Safe Intelligent Systems”, IEEE International Conference on Systems, Man and Cybernetics, pp.2524–2529 (2018).

[8] Kenichi Takada and Michitaka Kameyama, “Moving Image Scene Recognition and Its Application to Highly-Safe Intelligent Systems", The International Conference on Information and Digital Technologies 2019, pp.486-491(2019).

[9] 内閣府・消防庁・気象庁共同調査，“平成 23 年東日本大震災における避難行動等に関する面接調査（住民）分析結果”，（2011） http://www.bousai.go.jp/kaigirep/chousakai/tohokukyokun/7/pdf/1.pdf

[10] 宮城県本部／全国消防職員協議会・栗原消防職員協議会　石川正紀，“人はなぜすぐに避難しないのか　―人は皆、「自分だけは死なない」と思っている―”，第 36 回宮城自治研集会，第 4 分科会（2016）.

[11] A. Oliva, A. Torralba, “Modeling the shape of the scene: a holistic representation of the spatial envelope”, International Journal of Computer Vision, Vol.42(3), pp.145-175 (2001).

[12] C. Farabet, C. Couprie, L. Najman and Y. LeCun, “Learning hierarchical features for scene labeling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.35, pp.1915-1929 (2013).

[13] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, “Learning deep features for scene recognition using places database”, In Neural Information Processing Systems(NIPS), pp.487–495 (2014).

[14] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, “Object detectors emerge in deep scene cnns”, International Conference on Learning Representations (ICLR) (2015).

[15] Caffe | Deep Learning Framework.
http://caffe.berkeleyvision.org/

[16] caffe/models/bvlc_googlenet at master-BVLC/caffe-GitHub
https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet

[17] ImageNet
http://www.image-net.org/

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, “Going deeper with convolutions”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1-9 (2015).

[19] A. S. Rezaian, H. Azizpour, J. Sullivan and S. Carlsson, “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition”, IEEE Conference on Computer Vision and Pattern Recognition, pp.806-813 (2014).

[20] A. Krizhevsky, et al., “ImageNet classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems 25, pp.1106-1114 (2012).

[21] V. Nair, G. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines”, ICML, pp807-814 (2010).

[22] Machine Learning Project at the University of Waikato in New Zealand
https://www.cs.waikato.ac.nz/~ml/index.html

[23] J. Pearl, “Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning”, Proc. of Cognitive Science Society (CSS-7) (1985).

[24] G. Cooper and E. Herskovits, “A Bayesian method for the induction of probabilistic networks from Data”, Machine Learning, Vol.9, pp.309-347 (1992).

[25] Yoav Freund, Robert E. Schapire, “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting”, Journal of Computer Sys Sci 55, pp.119–139 (1977).

[26] V. Vapnik and A. Lerner, “Pattern Recognition Using Generalized Portrait Method”, Automation and Remote Control, p774-780 (1963).

[27] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection”, Proceedings of the fourteenth Int. Joint Conference on Artificial Intelligence, San Mateo, pp.1137-1143 (1995).

[28] J. A. Russell and L. F. Barrett, “Core affect prototypical emotional episodes and other things called emotion: dissecting the elephant”, Journal of Personality and Social Psychology, Vol.76, No. 5, pp.805-819 (1999).

[29] D. Kahneman, “Thinking Fast and Slow”, Farrar, Straus and Giroux, New York (2011). 村井章子(訳), “ファスト＆スロー あなたの意思はどのように決まるか？ 上・下”, 早川書房 (2014).

[30] M. D. Lieberman, R. Gaunt, D. T. Gilbert and Y. Trope, “Reflexion and reflection: A social cognitive neuroscience approach to attributional inference”, Advances in experimental social psychology, Vol. 34, pp.199-249 (2002).

[31] Bruce D. Lucas and Takeo Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision”, International Joint Conference on Artificial Intelligence, pp. 674-679(1981).

[32] G. Farneback, “Two-Frame Motion Estimation Based on. Polynomial Expansion”, Image Analysis, Vol. 2749(2003).

[33] CRCV Center for Research in Computer Vision at the University of Central Florida
http://crcv.ucf.edu/data/UCF101.php

[34] K. Soomro, A. R. Zamir, and M. Shah. “UCF101: A dataset of 101 human action classes from videos in the wild”, CRCV-TR-12-01 (2012).

[35] Serre Lab » HMDB a large human motion database
http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/

[36] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “HMDB: A large video database for human motion recognition”, In ICCV, pp.2556-2568 (2011).

[37] K. Simonyan, and A. Zisserman, “Two-Stream Convolutional Networks for Action Recognition in Videos”, Proceedings of the Advances in Neural Information Processing Systems, pp. 568–576 (2014).

[38] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei, “Large-scale video classification with convolutional neural networks”, In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp.1725-1732 (2014).

[39] He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778 (2016).

[40] H. Wang and C. Schmid, “Action Recognition with Improved Trajectories”, International Conference on Computer Vision (ICCV), pp.3551–3558 (2013).

[41] S. Rolf, “Maximum likelihood theory for incomplete data from an exponential family”, Scandinavian Journal of Statistics 1, pp.49-58 (1974).

[42] L. E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state Markov chains”, The Annals of Mathematical Statistics, vol.37, no.6, pp.1554-1563 (1966).

[43] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”, in IEEE Transactions on Information Theory, vol.13, no.2, pp.260-269 (1967).

[44] J. C. Platt, “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods.” In Advances in Large Margin Classifiers, pp.61–74. MIT Press (2000).

[45] J. Nelder, R. Wedderburn, “Generalized Linear Models”, Journal of the Royal Statistical Society, Series A (General) 135(3), pp.370-384 (1972).

[46] M. D. Hoffman, A. Gelman, “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”, Journal of Machine Learning Research 15, pp. 1593–1623 (2014).

[47] 木下広章，柴田久，石橋知也，雨宮護，樋野公宏，“コンビニエンスストアにおける犯罪発生状況と防犯施策に関する考察”，日本都市計画学会　都市計画論文集，Vol. 51，pp.350-356 (2016).

[48] A. Gelman and D. B. Rubin, “Inference from Iterative Simulation Using Multiple Sequences”, Statistical Science, Vol.7, No.4, pp. 457–511 (1992).

[49] T. Saaty, “A scaling method for priorities in hierarchical structures,” Journal of Mathematical Psychology, Vol.15, pp.234-281 (1977).

[50] T. Saaty, “The Analytic Hierarchy Process”, McGraw-Hill, New York (1980).

[51] 佐取朗，“犯罪の現状とセキュリティシステム”，日本信頼性学会（編）信頼性ハンドブック，日科技連，pp.798-803（2014).

[52] 市川総子，阪田知彦，吉川徹，“建物倒壊および道路閉塞のモデルに化による避難経路の危険度を考慮した避難地への到達前能性に関する研究”，地理情報システム学会　GIS-理論と応用, Vol. 12, No. 1, pp. 47-56 (2016).

付　録

以下に，テストデータによる静止画単一シーン認識の評価実験に使用したプログラムコードを示す.

1．CNN 特徴ベクトル抽出処理

```
#!/usr/bin/python
# -*- coding: utf-8 -*-
# pyton 2.7,  chainer1.17.0

import numpy as np
import chainer
from chainer import cuda, Function, gradient_check, Variable
from chainer import optimizers, serializers, utils
from chainer import Link, Chain, ChainList
import chainer.functions as F
import chainer.links as L
from chainer.functions import caffe
from PIL import Image
#
print "load >> GoogleNet"
func = caffe.CaffeFunction('bvlc_googlenet.caffemodel')
print ">> Completed"
#
ygdata= np.zeros(1024).reshape(1,1024).astype(np.float32)
ybfr= np.zeros(1024).reshape(1,1024).astype(np.float32)
#
#
for ii in range(0,100):
      fn= "{0:0>4}".format(ii+1)
      image = Image.open('ファイル名' + fn + '.jpg').convert('RGB')
      fixed_w, fixed_h = 224, 224
```

```
        w, h = image.size
        if w < h:
                shape = (fixed_w * w / h, fixed_h)
        else:
                shape = (fixed_w, fixed_h * h / w)
        left = (shape[0] - fixed_w) / 2
        top = (shape[1] - fixed_h) / 2
        right = left + fixed_w
        bottom = top + fixed_h
        image = image.resize(shape)
        image = image.crop((left, top, right, bottom))

        x_data = np.asarray(image).astype(np.float32)
        x_data = x_data.transpose(2,0,1)
        x_data = x_data[::-1,:,:]

        mean_image = np.zeros(3*224*224).reshape(3, 224, 224).astype(np.float32)
        mean_image[0] = 103.0
        mean_image[1] = 117.0
        mean_image[2] = 123.0
        x_data -= mean_image
        x_data = np.array([ x_data ])

        ####   CNN 特徴ベクトル抽出
        x = chainer.Variable(x_data)
        y, = func(inputs={'data': x}, outputs=['pool5/7x7_s1'], train=False)
        for p in range(0,1024):
            ybfr[0][p] = y.data[0, p]
        ygdata = np.append(ygdata, ybfr, axis=0)
        print "FileNum: " + fn

#CSV ファイルに出力
np.savetxt('Feature'+str(ii+1)+'.csv', ygdata, delimiter=',')
```

２．SVM 識別器生成及びテストデータ（未知画像）による評価の処理

```
# -*- coding: utf-8 -*-

#学習してモデルを生成、保存。
# pyton 2.7,  scikit-learn 0.18.0

import numpy as np
from sklearn import svm
from sklearn.metrics import classification_report, accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.externals import joblib

#トレーニング,  クラスデータの読み込み
trainFeature = np.genfromtxt(open('トレーニングデータセット名.csv', 'r'), delimiter =
      ',')
trainClass = np.genfromtxt(open('トレーニングデータセットのクラスファイル名.csv',
      'r'), delimiter = ',' ,dtype=np.int64)

#識別器のパラメータ設定
clf = svm.SVC(kernel='poly', C=1.0, degree=2, probability=True)
#SVM の学習
clf.fit(trainFeature, trainClass)

#  テストデータによる評価 /
testFeature = np.genfromtxt(open('テストデータセット名.csv', 'r'), delimiter = ',')
test_class = np.genfromtxt(open('テストデータセットのクラスファイル名.csv', 'r'),
      delimiter = ',')
#  識別
test_pred = clf.predict(testFeature)
# テストデータを  0,  1  判定
t_pob= test_probability = clf.predict_proba(testFeature) # 0 に属するか,  1 に属するか
      確率表示
print "¥n テストデータ個別の識別結果："
ngyo, nrets = test_probability.shape
for num in range(0,ngyo) :
      if test_pred[num] == 0 :
```

```
        print "%d: 0 >> %0.3f (%0.4f) |  Correct Class [%d]"%
    (num+1,t_pob[num,0],t_dec[num],test_class[num]),
        if test_class[num] != 0 : print " miss",
    else:
        print "%d: 1 >> %0.3f ( %0.4f) |  Correct Class [%d]"%
    (num+1,t_pob[num,1],t_dec[num],test_class[num]),
        if test_class[num] != 1 : print " miss",
    print ""

#テストデータ（未知画像）の識別結果
print classification_report(test_class, test_pred)
print "Test Data Accuracy: %0.2f " % (accuracy_score(test_class, test_pred)*100) +"%"
print "¥n confusion_matrix"
print confusion_matrix(test_class, test_pred)

#/ SVM 識別器の保存 /
joblib.dump(clf,"モデル名_model")
```

謝辞

本論文は，著者が石巻専修大学大学院理工学研究科において行った研究を取りまとめたものであります．

本研究を行うにあたり，恩師亀山充隆東北大学名誉教授より終始熱心な御指導と御鞭撻を頂きました．先生から学ばせて頂いた原理を深く追求していく姿勢は，研究者として最も重要なものだと実感しております．加えて，先生の御人格と研究，教育に対する前向きな御姿勢と，御討論などでの有益なる御助言から多くを学ばせて頂いたことを銘記し，ここに改めて深く感謝の意を表しますとともに，常日頃からの先生のあたたかい御心遣いにより，充実した研究室生活を送ることができましたことに厚く御礼申し上げます．

また，本研究をまとめることができましたのも，指導教員として，研究環境を充実したものにしていただくとともに研究の励ましをいただき，終始懇篤な御指導と御鞭撻を賜りました工藤すばる教授の御厚情によるものであり，衷心より感謝申し上げます．

更に，本論文をまとめるにあたり，御専門の立場から有益な御意見，御助言をいただいた，佐々木慶文教授，阿部正英教授に心より感謝と御礼を申し上げます．

令和４年２月１８日