博士論文

動画像における顕著性のあるイベント検出に関する研究

A Study on Salient Event Detection in Videos

国立大学法人　横浜国立大学
大学院環境情報学府

菅沼　雅徳
Masanori SUGANUMA

2017年9月

あらまし

近年では防犯を目的として，駅や空港などの公共施設だけではなく，マンションやビルなどの一般的な場所においても多くの監視カメラが設置されている．また，病院の手術室などにおいても，術後評価などの目的で大量の動画像の記録が行われている．その一方で，これら監視映像や術中映像などの多くは人間の目視によって確認作業が行われているため，その活用に膨大な労力が必要となっている．例えば，防犯のためには複数の監視映像を長時間監視し続けなければならない．また，術後評価のためには，術中における重要な医療行為を動画像から検出し，解析を行う必要があるが，大量の動画像から特定の医療行為をすべて検出することは大きな労力を要する．このように大量の動画像の記録は行われているが，それらを最大限有効活用できていないのが現状である．そのため，このような大量の動画像を適切に扱うために，映像内から特定の事象（イベント）を自動検出する，イベント検出の技術が求められている．特にイベントの中でも，映像内で出現頻度が低いイベントや重要なイベントを自動検出することは，上述した防犯や動画像解析を行う上で必要とされるため重要である．本論文では，映像内での出現頻度が低いイベントや重要なイベントを顕著性のあるイベントと呼び，これらを自動検出する方法を提案する．

イベント検出を行うモデルを構築する際に，用いる学習データに関して幾つかの条件が考えられる．まず，顕著性のあるイベント（負例）のラベルデータと顕著性のないイベント（正例）のラベルデータが利用可能な場合が挙げられる．この場合は，正例と負例を用いた教師あり学習によってイベント検出を行うモデルを構築するアプローチになる．次に，正例のラベルデータのみが与えられている場合が考えられる．このときは，正例データのみから正例モデルを構築し，そのモデルから逸脱するイベントを負例として検出するアプローチになる．最後に，正例および負例のラベルデータがともに与えられていない場合が考えられる．この場合は，映像を観測しながら正例と負例をそれぞれ定義し，イベント検出モデルを構築するアプローチになる．本論文では，上述した各条件におけるイベント検出を行う手法をそれぞれ提案し，先行研究との比較によって性能検証を行う．

まず，正例および負例が与えられている場合のイベント検出として，本論文では医用動画像からのイベント検出を行う．具体的には，覚醒下脳腫瘍摘出術と呼ばれる手術における重要な医療行為である電気刺激を動画像から自動検出する手法を提案する．提案手法では電気刺激を行う電極先端の検出と，電気刺激を行ったタイミングの検出を組み合わせることで，電気刺激を検出する．

次に，正例のみが与えられている場合のイベント検出として，多数の歩行者が登場する監視映像からのイベント検出を行う．ここでは，歩行者を正例，歩行者以外の物体である自動車や自転車を負例と定義する．提案手法では，Convolutional Autoencoder による復元誤差を利用してイベント検出を行う．UCSD Pedestrian dataset による性能検証を行い，先行研究と同等の性能を示すことができることを示す．

最後に，正例および負例が与えられていない場合のイベント検出として，監視映像からの侵入物体検出問題を扱う．ここでは，映像内での出現頻度が低い歩行者や車両などの物体を負例として定義して，性能検証を行う．本論文で提案する自己組織化ネットワークを用いることで，正例および負例が与えられていない場合でも，良好にイベント検出を行えることを示す．

Abstract

Surveillance cameras are installed in many public places such as airports, stations, and apartments for security. In addition, surgery videos are also recorded in operating rooms for postoperative evaluation. However, it takes a lot of effort to utilize these videos because there is a large number of videos and human have to visually check these videos. For example, it is required to have constant human monitoring of videos arriving at the control center for security. For postoperative evaluation, extracting specific medical practices from surgery videos is needed to analyze the videos, but it takes a lot of labor to extract specific medical actions from many videos. In light of this situation, event detection methods from videos are highly beneficial. Detecting events, such as important events and events which occur less frequently, are very useful for analyzing the videos. We refer to these events as salient events in this paper and propose methods to detect these salient events.

There are several conceivable conditions regarding training data when we construct event detection models. The first condition is to train models using label data of salient events and normal events. In the case of this situation, we construct models based on super-vised learning. The second condition is that only normal event data are available for training models. In the case of this situation, we construct normal models, and detect salient events using these models. The last condition is that we cannot use both label data of salient and normal events. In the case of this situation, it is needed to define events as salient or normal while observing videos, and we construct models based on these observations. In this paper, we propose event detection methods in each condition above, and demonstrate the effectiveness of our methods in comparative experiments with other methods.

In Section 3, we detect salient events in surgery videos using training data including salient and normal label data. Specifically, we detect electrical stimulation in videos of cotrical mapping in awake surgery. The proposed method consists of two phases: detection of a probe tip position and detection of electrical stimulation timings.

In Section 4, we detect salient events in surveillance videos using training data including only normal label data. In this section, pedestrians are defined as normal and other objects, such as cars and bike, are defiend as salient events. Our method detects these salient events using reconstruction error of convolutional autoencoder. To evaluate the proposed method, we test our method on the UCSD Pedestrian dataset.

Finally, in Setction 5, we deal with intrusion detection tasks in surveillance videos in the case using training data including neither salient nor normal label data. In this section, we define pedestrians and cars as salient events for evalution. We show that our method, self-organizing network, can detect salient events in videos.

目 次

図 目 次

表 目 次

第1章　序論

1.1　背景と目的

近年では防犯を目的として，駅や空港などの公共施設だけではなく，マンションやビルなどの一般的な場所においても多くの監視カメラが設置されている．また，病院の手術室などにおいても，術後評価などの目的で大量の動画像の記録が行われている．その一方で，これら監視映像や術中映像などの多くは人間の目視によって確認作業が行われているため，その活用に膨大な労力が必要となっている．例えば，防犯のためには複数の監視映像を長時間監視し続けなければならない．また，術後評価のためには，術中における重要な医療行為を動画像から検出し，解析を行う必要があるが，大量の動画像から特定の医療行為をすべて検出することは大きな労力を要する．このように大量の動画像の記録は行われているが，それらを最大限有効活用できていないのが現状である．そのため，このような大量の動画像を適切に扱うために，映像内から特定の事象（イベント）を自動検出する，イベント検出の技術が求められている．特にイベントの中でも，映像内で出現頻度が低いイベントや重要なイベントを自動検出することは，上述した防犯や動画像解析を行う上で必要とされるため重要である．本論文では，映像内での出現頻度が低いイベントや重要なイベントを顕著性のあるイベントと呼び，これらを自動検出する方法を提案する．

イベント検出を行うモデルを構築する際に，図 1.1 に示すように用いる学習データに関していくつかの条件が考えられる．

まず，顕著性のあるイベント（負例）のラベルデータと顕著性のないイベント（正例）のラベルデータが利用可能な場合（条件 1）が挙げられる．この場合は，正例と負例のラベルデータを用いた教師あり学習によってイベント検出を行うモデルを構築するアプローチになる．

次に，正例のラベルデータのみが与えられている場合（条件 2）が考えられる．現実的には，検出対象であるすべての負例パターンや多くの負例データを集めることが困難であることが多いため，正例データのみを用いたイベント検出に関する研究は盛んに行われている．この条件の場合，正例データのみから正例モデルを構築し，そのモデルから逸脱するイベントを負例として検出するアプローチになる．

		負例	
		ラベルあり	ラベルなし
正例	ラベルあり	条件1	条件2
	ラベルなし	×	条件3

図 1.1: イベント検出モデルを構築する際の学習データに関する条件

最後に，正例および負例のラベルデータがともに与えられていない場合（条件 3）が考えられる．現実世界での運用を考えると，事前に正例および負例の定義ができない場合や，時間の経過に伴う環境変化によって正例および負例の定義が変わってしまう場合（例えば，天候による照明変化など）や，学習データに存在しなかった正例データが出現する可能性なども考えられる．そのため，事前に正例および負例データを用意することが困難であったり，条件 1 や条件 2 のように事前に構築した固定のイベント検出モデルの適用だけでは不十分である場合も考えられる．したがって，より汎用性の高いイベント検出を行うためには，事前にラベルデータが与えられていなくても，適用中の動画像に応じてモデルの構築および更新が行われる手法が望まれる．この条件の場合，動画像を観測しながら正例と負例をそれぞれ定義し，イベント検出モデルの構築および更新を行うアプローチになる．

本論文では，上述した各条件におけるイベント検出を行う手法をそれぞれ提案し，先行研究との比較によって性能検証を行う．

1.2 本論文の構成

本論文の構成は次の通りである．まず第 2 章で，上述した各条件におけるイベント検出に関する先行研究について述べる．第 3 章では，正例および負例のラベルデータが与えられている場合のイベント検出として，覚醒下脳腫瘍摘出術と呼ばれる手術における重要な医療行為である電気刺激を動画像から自動検出する手法を提案する．提案手法では電気刺激を行う電極先端の検出と，電気刺激を行ったタイミングの検出を組み合わせることで，電気刺激を検出する．第 4 章では，正例のラベルデータのみが与えられている場合のイベント検出として，多数の歩行者が登場する監視映像からのイベント検出を行う．ここでは，歩行者を正例，歩行者以外の物体である自動車や自転車を負例と定義する．提案手法では，Convolutional Autoencoder による再構築誤差を利用してイベント検出を行う．UCSD Pedestrian dataset による性能検証を行い，先行研究と同等の性能を示すことができることを示す．第 5 章では，正例および負例のラベルデータが与えられていない場合のイベント検出として，監視映像からの侵入物体検出問題を扱う．ここでは，映像内での出現頻度が低い歩行者や車両などの物体を負例として定義して，性能検証を行う．本論文で提案する自己組織化モデルを用いることで，正例および負例のラベルデータが与えられていない場合でも，良好にイベント検出を行えることを示す．最後に第 6 章で，本論文のまとめと今後の課題について述べる．

第2章　関連研究

本章では，本研究と関連の深い動画像からのイベント検出に関する先行研究について述べる．

まず，顕著性のあるイベント（負例）と顕著性のないイベント（正例）のラベルデータを用いたイベント検出として，本論文では覚醒下脳腫瘍摘出術において重要な医療行為である電気刺激を動画像記録から自動検出する方法を提案する．そのため，本章ではまず本論文が対象とする覚醒下脳腫瘍摘出術について説明する．その後，動画像記録を用いた手術工程解析に関する先行研究について述べる．

続いて，正例のラベルデータのみを用いたイベント検出として，本論文では監視映像からの異常検知タスクを扱うため，正例データのみを用いた異常検知に関する先行研究について述べる．ここでは，顕著性のあるイベントを異常，顕著性のないイベント（正例）を正常と呼ぶこととする．

最後に，本論文で扱う負例および正例のラベルデータを用いないイベント検出に関する先行研究について述べる．

2.1　覚醒下脳腫瘍摘出術について

覚醒下脳腫瘍摘出術とは，脳機能を脳機能を最大限に温存しつつ，脳腫瘍を可能な限り摘出するために患者を目覚めさせた状態で行う脳腫瘍摘出術のことである．運動や言語などの脳機能の位置には患者によって個人差があり，また脳における腫瘍領域と正常領域の境界が曖昧であることが多い．そのため，術後の脳機能を最大限に温存するためには，腫瘍周囲の脳の機能野を正確に同定しながら腫瘍を切除する必要がある［5］．

機能野同定の工程は皮質マッピングとよばれており，術者が患者の脳表面に電極で直接電気刺激を行い，その際の言語タスクおよび運動タスクに対する患者の反応から機能野を同定する．言語タスクの例として，モニタに映っている物体の名称呼称や動詞生成が挙げられる．患者の言語野を電気刺激した場合，患者は発話ができなくなるため，術者はその反応を観察することで言語野の位置を同定することができる．

正確な機能野の同定には，電気刺激位置，電気刺激強度，電気刺激時のタスクの 3 変数の適切な組み合わせが必要であるが，これら 3 変数のさまざまな組み合わせを時間的制約のある術中に網羅的に行うことは不可能であるため，効率的なマッピングを行うことが必要である．また，正確な機能野の位置同定が患者の術後障害リスク減少や生存率向上につながるため，マッピング工程の洗練化や効率的なマッピング方法の解析は重要である．

2.2　正例および負例のラベルデータを用いた手術工程解析に関する先行研究

手術の動画像記録やシステムのセンサ情報を用いて術中のイベント検出や解析を行うことは，手術技術の洗練化や客観的評価，手術工程の効率化において重要である．また，近年導入が進んでい

図 2.1: 異常シーンの例（文献 [1] より引用）

るコンピュータ支援システムにおいても，術中に発生したイベントや正確な手術工程の理解は必要であるため，動画像記録などを用いたイベント検出，映像解析はますます必要とされている．

そのため近年では，術中の動画像記録をもとに，手術工程の自動認識や分割，動画像内の手術器具検出の研究が盛んに行われている [6–8]．例えば，腹腔鏡手術を対象に手術工程の認識を行った研究では，画像内での色特徴や使用されている手術器具の有無といった情報と hidden Markov model（HMM）[9] を用いた手法 [10–12] や，色特徴による画像内の領域分割，手術器具の追跡，組織領域の変形などの複数の画像特徴とベイジアンネットワークを用いた手法 [13] などが提案されている．

また，手術室内全体を映した動画像記録を用いた研究も行われている [14]．文献 [14] ではあらかじめ共通する 4 つの手術室内の状態を定義し，動画像内の画像特徴を用いて手術室内の状態を同定している．文献 [1] では，立体高次局所自己相関（Cubic higher-order local auto-correlation; CHLAC）特徴 [15] を用いた異常シーンの検出を行っている．この手法では，通常シーンと検出対象シーンにおける CHLAC 特徴の部分空間を比較することで，異常シーンかどうかを判定している．異常シーンの例として，図 2.1 内の手前に示すように，ある人が何かを拾うためにほか人の足元で腰を下ろしているシーンなどが挙げられる．また，Suzuki らは動画像記録のファイルサイズに着目した手術シーンの変化を検出する手法を提案している [16]．これは術中にトラブルが生じた場合や手術シーンが変化したときに，映像内の動きの量が増加しファイルサイズが変化するため，ファイルサイズという指標がシーン検出に利用可能であるという仮定に基づいている．

さらに，動画像内の画像特徴を用いた手術器具の検出や追跡を行った研究も報告されている [17, 18]．例えば，Raphael らは網膜手術における手術器具の検出手法を提案している [19]．この手法では，画像内の勾配情報を利用した追跡と学習によって得られた識別器を用いて，手術器具の存在確率を算出することで手術器具の検出を行っている．これら動画像記録を用いた研究では，解析対象の手術記録に適した画像特徴を用いた手法が提案されている．

また動画像内の画像情報だけでなく，ロボット支援システムや手術器具に取り付けたセンサから得た器具の位置情報や運動学的情報を用いた解析も行われている．これらの情報はおもに術者の手術動作の解析や技術評価に用いられている [20–24]．

動画像記録を用いたアプローチは，手術器具にセンサを装着して位置情報を取得するアプローチ [22] などに対して，センサの導入や既存のコンピュータ支援システムとの同期，手術のルーチン

ワークの変更が不必要であるという利点が挙げられる．また，現在までに膨大な量の動画像記録が蓄積されている場合，これらの動画像記録を手術の工程解析などに活用できるという利点もある．

2.3 正例のラベルデータのみを用いた動画像からのイベント検出に関する先行研究

本論文の 4 章では顕著性のあるイベント（負例）のラベルデータを用いずに，顕著性のないイベント（正例）のラベルデータのみを用いた監視映像からのイベント検出タスクを扱う．正例データのみを用いて正常モデルを構築し，そのモデルを用いて異常（イベント）検出を行う先行研究がこれまでに数多く提案されている．これらの先行研究は本論文とも関わりが深いため，本節で紹介する．なお，ここでは顕著性のあるイベント（負例）を異常，顕著性のないイベント（正例）を正常と呼ぶこととする．

正常データのみを用いた異常検知に関する先行研究では，映像内での出現頻度が低いパターンや，正常パターンと特徴が大きく異なるパターンを異常と定義して，それらを検出する手法が多い［25–27］．そのため，多くの手法では事前に異常を含まない正常データから映像内の正常モデルを構築し，その正常モデルにおける生起確率が低いパターンや，モデルを逸脱するパターンを異常として検出している．また，異常検知モデルを適用するシーンとして，人物が多数存在する混雑シーンとそれ以外のシーン（本章では通常シーンと呼ぶこととする）の 2 つに大別される．

2.3.1 通常シーンにおける異常検知

Adam らは，シンプルだが映像内の速度に関する異常を検出する手法を提案している［25］．この手法では，画像内の各局所領域でオプティカルフローを観察し，これまでに観測した速度情報とは異なる領域を異常として検出する．実験では，地下鉄のホームや街頭での人や車両などの間違った方向への移動を良好に検知できることが示されている．

Xiang らは屋内と屋外で撮影された動画像内での人物の行動パターンについてのモデルを学習し，学習データにはなかった行動パターンを異常として検出する手法を提案している［28］．Kim らはオプティカルフローと Markov Random Fields に基づいた手法によって異常検知を行っている［29］．

南里らは立体高次局所自己相関特徴（Cubic High-order Local Auto-Correlation; CHLAC）特徴［15］を用いて，人物の歩行動作を正常動作，走行動作と転倒動作を異常動作と定義した実験において，複数の人物が映る動画像から異常動作を検出することに成功している［30］．文献［30］では，通常動作と定義した「歩く」動作のデータから CHLAC 特徴を算出し，算出した CHLAC 特徴空間から主成分分析によって通常動作特徴の部分空間を構築し，その部分空間からの距離を異常値として異常動作検出を行っている．CHLAC 特徴は顔画像などの 2 次元データの認識に対して有効な高次自己相関（High-order Local Auto-Correlation; HLAC）特徴［31］に時間成分を加えて 3 次元に拡張したものである．CHLAC 特徴は HLAC 特徴と同様に次式 (2.1) によって定義される．

$$x_f^{(N)}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N) \triangleq \int f(\boldsymbol{r})f(\boldsymbol{r}+\boldsymbol{a}_1)\ldots f(\boldsymbol{r}+\boldsymbol{a}_N)d\boldsymbol{r} \tag{2.1}$$

$\boldsymbol{r}$ はデータ中の注目点，$\boldsymbol{a}_i(i=1,\ldots,N)$ は N 個の局所変位を表しており，画像内の 2 次元座標と時間の 3 次元ベクトルである．CHLAC 特徴は画像を時系列順に並べたボクセルデータに対して各点の位置と時間方向の局所的な自己相関特徴を算出し，これをボクセルデータ全体に対して積分することで得られる．局所変位の組み合わせは二値動画像からの位置変位だけの場合，図 2.2 に示すように 0 次は 1 個，1 次は 4 個，2 次は 20 個の計 25 個（次元）となる．CHLAC 特徴の場合は時間方向も加えた 3 次元の変位になるため，図 2.3 のような立方体上で表現することができる．二値動画像からの CHLAC 特徴の局所変位の組み合わせは 0 次が 1 個，1 次が 13 個，2 次が 237 個の計

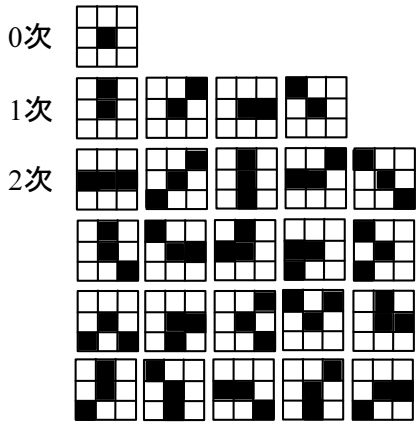


図 2.2: 高次局所自己相関の位置変位パターン

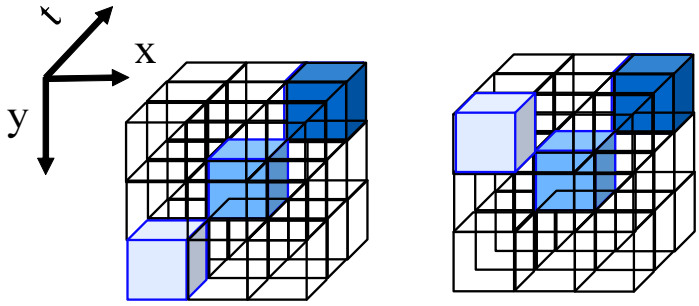


図 2.3: 立体高次局所自己相関の変位パターン例

251 個（次元）である．CHLAC 特徴を用いた異常検出では，一つの部分空間で通常動作部分空間を近似しているため，通常動作が複数存在する環境への適用が困難であると述べられている［30］．また，環境の変化に伴い，通常動作が変化する場合においても，部分空間の構築が困難であると考えられる．

そのほかに，映像内の各物体の動きを解析することで，異常検知を行う手法も提案されている．このとき各物体の動きを解析するために，追跡手法がしばしば用いられている．各物体を追跡することで，各物体の移動軌跡を得ることができるため，それらの移動軌跡を用いた異常検知手法が提案されている［32–35］．また，追跡によって得られた特徴を用いて物体の典型的な活動パターンをモデル化し，異常を検出するアプローチも提案されている［36］．

これら上述した手法は，人物が多数存在する混雑シーンでは，異常検知性能が低下してしまうと考えられる．例えば，混雑シーンでは正常シーンが多数存在するため，文献［30］で述べられているように南里らの手法では適用が困難である．また，混雑シーンでは物体のオクルージョンが多発するため，追跡が正確に行えないことが考えられる．そのため，追跡による特徴を用いた手法も混雑シーンにおける適用は不向きであると考えられる．

図 2.4: UCSD pedestrian dataset（Ped1，Ped2）の例

2.3.2 混雑シーンにおける異常検知

図 2.4 に混雑シーンの例を示す．図 2.4 は Mahadevan らによって公開されている UCSD pedestrian と呼ばれるデータセットで，Ped1 と Ped2 の 2 つのシーンで構成されている［37］．このデータセットでは，歩行者以外の自転車や自転車，スケートボーダなどが異常として定義されている．

このような映像内の環境が複雑な混雑シーンにおいて，高精度な異常検知が可能となれば，より正確な映像解析や幅広い領域での防犯システムの運用が可能となり，有用性が期待できる．そのため，近年では人物が多数存在する混雑シーンにおける異常検知に関する研究が盛んに行われており［38–46］，これまで局所および大局特徴量の分布に基づいた手法や，スパースコーディングや Autoencoder などような画像の再構築に基づいた手法などが提案されている．

Mahadevan らは mixture of dynamic textures（MDT）［47］を用いて時間情報と空間情報の両方を考慮した正常モデルを構築し，混雑シーンにおいて異常検知を行う手法を提案している［37, 48］．時間情報に関する正常性は MDT でモデル化され，空間情報に関する正常性は MDT に基づいた saliency detector［49］で定義している．Mehran らは，オプティカルフローに基づいた social force モデルを提案することで混雑シーンにおける正常パターンを表現し，異常検知を可能にしている［50］．この手法ではオプティカルフローを用いているが，物体の追跡は行っておらず，オプティカルフローに基づいた social force モデルによってシーンのモデル化を行っている．Kratz らは映像内の局所領域の動き情報と HMM に基づいたアプローチを提案している［51］．

近年ではスパースコーディングに基づいた手法も数多く提案されており，高精度な結果を示している．Cong らは，異常検知のための指標として Sparse reconstruction cost（SRC）を提案している［52］．これは正常データにおける基底ベクトルによる再構築に基づいた異常度を表す指標である．また，この手法では画像内の局所的な異常検知と画像全体に関わる大局的な異常検知の両方に対応可能なことも示されている．一般的にスパースコーディングに基づいた手法は計算コストが高く，リアルタイムで実行するための処理速度が十分でないことが多いが，Lu らは，高速で高精度なスパースコーディングベースの手法を提案している［53］．文献［53］では，再構築に用いる基底ベクトルの組み合わせを学習段階で複数獲得しておき，適用する際にはそれら獲得された組み合わ

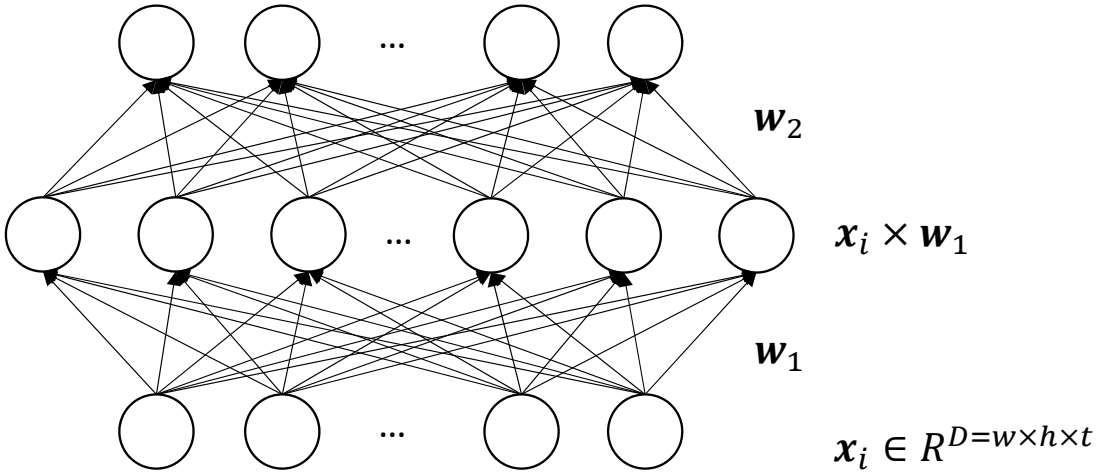


図 2.5: Autoencoder による特徴表現の例

せの中から最も再構築誤差が小さいものを選択することで高速化を図っている．

また，Autoencoder に基づいた異常検知手法も提案されており，混雑シーンにおける異常検知で優れた性能を示している．Sabokrou らは，大局的な特徴と局所的な特徴のそれぞれ用いて異常検知を行う手法を提案している［54］．このとき，大局的な特徴を Autoencoder によって表現している．図 2.5 に Autoencoder による特徴表現の例を示す．図 2.5 に示すように，入力情報 $\boldsymbol{x}_i$ を重み行列 $\boldsymbol{w}_1$ との内積によって変換し，変換後の特徴量 $\boldsymbol{x}_i \times \boldsymbol{w}_1$ を新たな特徴量とする．このように Autoencoder によって変換された特徴表現は，元の特徴表現より良い特徴が得られることがある．局所的な特徴は画像内の周辺領域との SSIM［55, 56］に基づいた類似度によって表現している．SSIM は画質評価に用いられる指標である．また，Xu らはより深い構造の Autoencoder を用いた異常検知手法を提案している［2］．この手法では図 2.6 に示すように，画像内の形状情報（image paches）と速度情報（optical flow patches），それらを統合した情報（joint patches）をそれぞれ Autoencoder に入力し特徴変換を行い，変換後の特徴量を One-Class SVM［57］に入力することで異常検知を行う．Hasan らは，Convolutional Autoencoder（CAE）と Autoencoder を用いた異常検知手法を提案している［58］．CAE の学習では画像の情報をそのまま入力し，入力画像を復元するように学習を行う．Autoencoder の学習では，画像から算出した Histogram of oriented gradients（HOG）特徴［59］と Histogram of optical flow（HOF）［60］を入力情報として学習を行っている．この手法では，これら入力情報との再構築誤差を指標として，異常検知を実現している．

また，近年高い性能で注目を集めている Convolutional Neural Network（CNN）［61］などの深層学習手法を利用した異常検知手法も提案されている［62, 63］．文献［3］では，学習済みの CNN を用いて抽出した特徴とオプティカルフロー特徴によって高精度に異常検知が行えることが示されている．文献［3］の手法の概要を図 2.7 に示す．図 2.7 内の BFCN は CNN の最終層にバイナリマップを出力する層を追加した Binary Fully Convolutional Network を表し，この BFCN に入力画像を入力することでバイナリマップを得る．そして，得られたバイナリマップから動きのパターンを抽出し，オプティカルフロー特徴と統合することで異常検知を行っている．

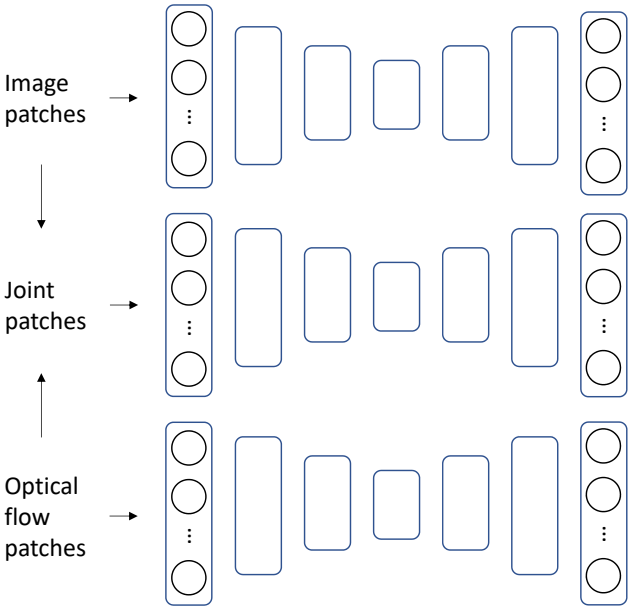


図 2.6: Xu らの手法 [2]による特徴表現の例

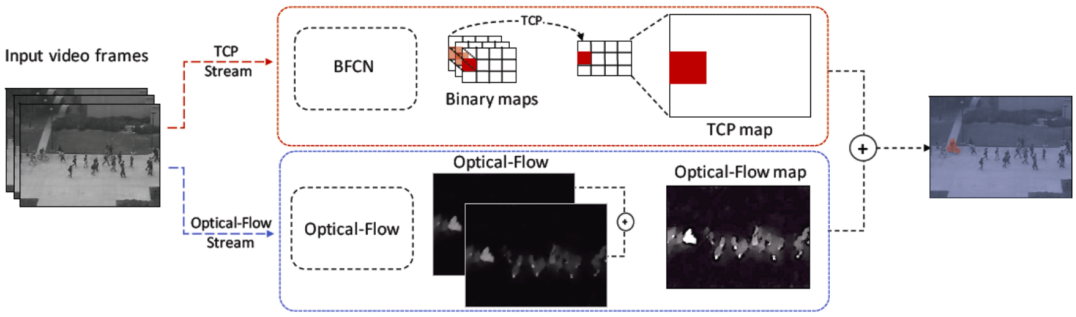


図 2.7: 文献 [3]の手法の概要（文献 [3]より引用）

2.4 正例および負例のラベルデータを用いない動画像からのイベント検出に関する先行研究

本論文の 5 章では，正例および負例のラベルデータを用いないイベント検出問題を扱うが，この問題設定の場合，事前に正例や負例データを用いることができないため，映像を観測しながらイベント検出モデルを構築する必要がある．このように観測した情報を用いてイベント検出モデルを構築もしくは更新し，そのモデルを用いてイベント検出を行う先行研究がこれまでにいくつか提案されている［64–66］．基本的には前節で述べた手法と同様に環境の正常モデルを構築し，その正常モデルから逸脱するパターンを検出するアプローチであるが，イベント検出を実行しながら正常モデルの更新が行われる点で大きく異なる．これらの研究では，性能検証の際にはモデル構築や評価のために正例，負例のラベル情報を利用しているが，実際に使用するときには正例や負例の定義を行わなくてもイベント検出が可能となっている．なお，これらの先行研究は主に異常検知問題を扱っているため，本節でも顕著性のあるイベント（負例）を異常，顕著性のないイベント（正例）を正常と呼ぶこととする．

上述したように，前節で説明した多くの先行研究では事前に構築された正常モデルが固定であり，異常検知の適用中にそのモデルが更新されることはない．しかし，現実世界での運用を考えると，天候によって照明変化が生じることで正常パターンが変化したり，学習データには出現しなかった正常パターンが出現する場合など，事前に構築した固定の正常モデルだけでは正確に映像内の正常性を表現することが難しいと考えられる．したがって，異常検知やイベント検出をより頑健に行うには，適用中の環境に応じてモデルの構築および更新が行われる環境に適応的な手法が望まれる．本節ではこれら環境変化に適応的な手法について述べる．

2.4.1 Grow When Required ネットワーク

環境に適応的な異常検知手法の例としては Masland らの Grow When Required（GWR）ネットワークが挙げられる［67–69］．文献［67］では，GWR ネットワークは異常を含まない環境の通常状態とみなすデータを用いてネットワーク構造を学習し，その結果として学習時には現れなかったデータを異常として検出できることが示されている．GWR ネットワークの概略図を図 2.8 に示す．GWR ネットワークはノードとエッジによって構成されている．各ノードには重みベクトルと発火係数が与えれており，入力ベクトルと最も距離が近い重みベクトルをもつノードが代表ノードとし

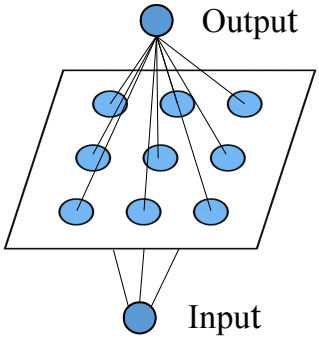


図 2.8: GWR ネットワークの概略図

て選択され，代表ノードがもつ発火係数の値がネットワークの出力値となる．代表ノードが選択されるたびに代表ノードの発火係数の値を次第に減少させていくことで，環境で頻繁に観測される特徴ベクトルに対して馴化していく．また，入力ベクトルに対する代表ノードとの特徴空間におけるユークリッド距離と発火係数がそれぞれ設定されたしきい値以下の場合は，新たなノードをネットワークに追加することで環境に適応していく点が GWR ネットワークにおける特徴である．さらに，GWR ネットワークでは発火係数の減少方法に馴化モデルが用いられている．馴化モデルには GWR ネットワークで用いられている Stanley モデル［70］や Wang-Arbib モデル［71, 72］などの生物の馴化現象を説明するモデルが提案されており，これら馴化モデルを環境に適応的なシステムのパラメータ更新に利用する研究が多く存在する．

Marsland らは GWR ネットワークを移動ロボットに搭載し，ソナーセンサを用いた異常検出実験を行っている．実験は大学構内の 10[m] 程度の通路において，ドアの開放の有無などを異常とした 3 つの異なる環境で行われている．実験の結果，学習環境では次第に環境からの入力に対して馴化していくことが確認された．また，学習環境とは異なる未知環境に対してこれまでに学習した GWR ネットワークを適用することで，学習環境では現れなかったドアの開放といった異常を検出することに成功している．

さらに，Nehmzow らは GWR ネットワークの入力にソナーセンサではなく，画像を用いた異常検出実験を行っている．実験では，周囲を壁で囲まれた閉空間でロボットによる壁伝い行動を行わせている．その結果，未知環境において学習時には存在しなかったボールを異常として適切に反応したことが確認されている［73–75］．

しかし，GWR ネットワークを用いた異常検出のアプローチでは，入力画像全体からの特徴ベクトルを用いているため，入力画像内のどの領域に異常が存在するかといった空間情報が欠落している．また，GWR ネットワークは単純な構造のため，環境に変動が含まれるような場合に対して適用することは難しいと考えられる．

2.4.2 刺激の選択性を用いた領域検出ネットワーク

武田らは 3 層のネットワーク構造を用いて環境の通常状態を表現する手法を提案している［4］．このネットワークモデルの概略図を図 2.9 に示す．このモデルは入力層，リージョン層，パターン層の 3 層構造で構成されており，動画像を 1 フレームごとに入力画像として扱う．入力層は入力画像を格子状に区分けした複数の矩形領域から構成されており，それぞれの矩形領域内の 3×3 画素の 2 次元パターンが入力パターンとなっている．リージョン層は入力層の各矩形領域に 1 対 1 に対応したリージョンノードによって構成されている．パターン層は入力画像から得られる入力パターンを記憶するパターンノードによって構成されている．リージョンノード間，リージョンノードとパターンノード間にはいずれも結合関係が形成され，結合荷重が存在する．入力に応じてそれらの結合荷重が更新されることで環境の正常性を記憶するネットワーク構造を獲得する．実験では車両や歩行者を検出すべき対象とした侵入物体検出問題にネットワークモデルを適用し，樹木の揺れといった環境変化への反応を抑制しつつ，検出対象に適切に反応を示すことが確認されている．

2.4.3 適応的背景モデル

動画中の背景は環境で最も頻繁に観測される正常状態と捉えることができるため，背景モデルの研究は本研究とも密接に関係している．背景モデルの構築に関する研究は数多く行われており，樹木や水面の揺れ，天候による照明条件の変化などの環境変化に対して頑健な背景モデルの構築方法

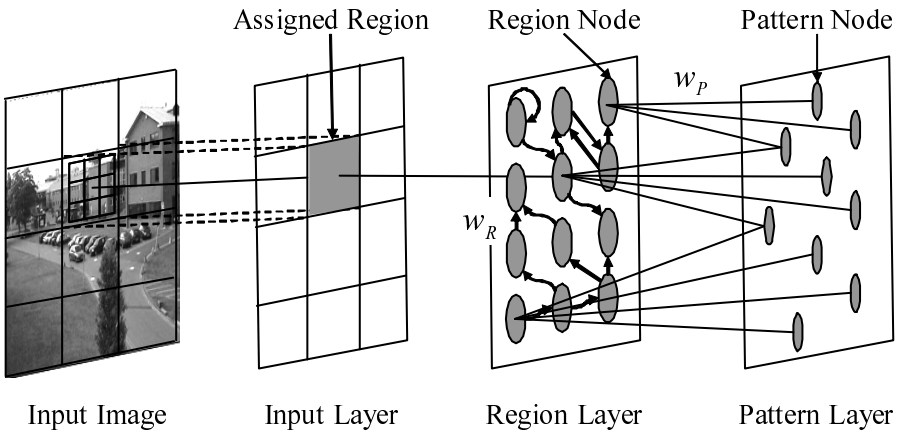


図 2.9: 刺激の選択性を用いた領域検出ネットワークの概略図（文献 [4] より引用）

が提案されている．本節では環境の観測に応じてパラメータを更新することで環境変化に頑健な適応的背景モデルについて述べる．

混合ガウス分布を用いた背景モデル構築

画像中の画素ごとに混合ガウス分布を用いて背景モデルを構築する研究が数多く提案されており，照明条件の変化などの背景変化に対して頑健に対処できることが示されている．Stauffer ら [76–78]，KaewTraKulPong ら [79] は 1 つの画素を混合ガウス分布によって表現することで，環境変化に適応可能な背景モデルを構築する手法を提案している．また，Han ら [80]，Zivkovic ら [81, 82] は画素ごとにガウス分布の数を変えることで環境変化が大きい場合にも対応可能な背景モデルを提案している．

ここで，混合ガウス分布を用いて背景モデルを構築する最も代表的な手法の 1 つである Stauffer らの手法について説明する．

まず，時刻 t における画像中の位置 (x, y) における画素値を I_t とすると，時刻 t までの画素値 $(I_0, \ldots, I_t)$ は K 個の混合ガウス分布 η を用いてモデル化でき，I_t の確率分布は次式 (2.2) で表される．

$$P(I_t) = \sum_{k=1}^{K} w_{k,t}\eta(I_t, \mu_{k,t}, \Sigma_{k,t}) \tag{2.2}$$

$w_{k,t}$ は k 番目のガウス分布の重み，$\mu_{k,t}$ は k 番目のガウス分布の平均値，$\Sigma_{k,t}$ は k 番目のガウス分布の共分散行列である．なお，計算の簡略化のため，$\Sigma_{k,t} = \sigma_{k,t}^2\mathbf{I}$ とする．

次に，新しく観測された画素値 I_t に対して K 個の分布の中で最も一致する分布を探す．分布 k の平均値からある標準偏差 κ 以内に画素値 I_t が存在する場合，一致していると判定する．次に，K 個のガウス分布の重みを次式 (2.3) によって更新する．

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha M_{k,t} \tag{2.3}$$

α は学習率，$M_{k,t}$ は I_t と一致した分布があれば 1，そうでなければ 0 の値をとる．各分布の重みを更新した後に重みの総和が 1 となるように正規化する．さらに I_t と一致した分布 k の平均値，分散を次式によって更新する．

$$\mu_{k,t} = (1 - \rho)\mu_{k,t-1} + \rho I_t \tag{2.4}$$

$$\sigma_{k,t}^2 = (1-\rho)\sigma_{k,t-1}^2 + \rho(I_t - \mu_{k,t})^T(I_t - \mu_{k,t}) \tag{2.5}$$

ρ は学習率であり，次式 (2.6) で算出される．

$$\rho = \alpha\eta(I_t|\mu_{k,t}, \sigma_{k,t}) \tag{2.6}$$

そして，K 個のガウス分布をそれぞれ w/σ の降順に並べ変えて，次式 (2.7) を満たす B 個のガウス分布を背景モデルと定義する．

$$B = \arg\min_b \left(\sum_{k=1}^{b} w_k > T \right) \tag{2.7}$$

T は背景モデルの決定に関わる定数であり，T が小さい場合は背景となるガウス分布数が少なくなるため，環境変化に対して反応しやすくなる．一方，T が大きい場合は複数のガウス分布が背景に割り当てられるため，照明変化や木々の揺れなどの環境変化を考慮した背景モデルが構築される．

2.4.4 スパースコーディングに基づいたオンライン学習手法

Zhao らは通常動作の一連のフレーム集合を Histograms of oriented gradients（HOG）[59] と Histograms of optical flow（HOF）[60] でベクトル化を行い，それらを通常動作の辞書として保持し，スパースコーディングと組み合わせることで人物の異常動作を検知する手法を提案している [83]．このとき異常かどうかの判定は次式（2.8）で算出される正常度で判定される．

$$J(\mathbf{X}_i, \boldsymbol{\alpha}_i, \mathbf{D}) = \frac{1}{2}\sum_j \|\mathbf{X}_i^j - \mathbf{D}\boldsymbol{\alpha}_i^j\|_2^2 + \lambda_1 \sum_j \|\boldsymbol{\alpha}^j\|_1 + \lambda_2 \sum_{j,k} \mathbf{W}_{jk}\|\boldsymbol{\alpha}_i^j - \boldsymbol{\alpha}_i^k\|_2^2 \tag{2.8}$$

$\mathbf{D}$ は辞書，$\mathbf{X}_i$ は入力情報（イベント），$\boldsymbol{\alpha}_i$ は重みベクトル，$\mathbf{W}_{jk}$ は正則化項である．正常データに対しては式（2.8）の値が小さくなるため，値がしきい値を超えた場合に入力情報が異常であると判定される．この手法では，与えられた $\mathbf{X}_i$ に対して式（2.8）を最小化する重みベクトル $\boldsymbol{\alpha}_i^*$ と辞書 $\mathbf{D}^*$ を学習する．具体的には，まず入力情報 $\mathbf{X}_t$ と辞書 $\mathbf{D} = \mathbf{D}_{t-1}$ を用いて，式（2.9）を解くことで重みベクトル $\boldsymbol{\alpha}_t$ を学習する．

$$\min_{\boldsymbol{\alpha}_i^1, \boldsymbol{\alpha}_i^{n_i}} \frac{1}{2}\sum_j \|\mathbf{X}_i^j - \mathbf{D}\boldsymbol{\alpha}_i^j\|_2^2 + \lambda_1 \sum_j \|\boldsymbol{\alpha}^j\|_1 + \lambda_2 \sum_{j,k} \mathbf{W}_{jk}\|\boldsymbol{\alpha}_i^j - \boldsymbol{\alpha}_i^k\|_2^2 \tag{2.9}$$

続いて，式（2.10）を解くことで辞書 $\mathbf{D}$ を更新する．

$$\min_{D \in C} \frac{1}{2t}\sum_{i=1}^{t} \sum_{j=1,\ldots,n_i} \|\mathbf{X}_i^j - \mathbf{D}\boldsymbol{\alpha}_i^j\|_2^2 \tag{2.10}$$

ここで，$C = \{\mathbf{D} \in \mathbb{R}^{m \times k} : \mathbf{d}_j^T\mathbf{d}_j \leq 1, \forall j = 1, \ldots, k\}$ である．このように Zhao らの手法では，新しいフレームが入力されるたびに重みベクトルと辞書を更新することで，新しい環境に対しても適応可能としている．また，監視映像に対する異常検知実験から，辞書の更新を行った方が行わなかったときよりも性能が優れていることが確認されている．

2.5 まとめ

本章では，本研究に関連する先行研究として，手術動画像記録からのイベント検出，監視映像および混雑シーンからのイベント検出に関する研究について述べた．

手術動画像からのイベント検出では，手術工程の解析のために術中の医療行為や術者の行動をイベントとして検出する手法について説明した．手術工程の解析には，動画像記録を利用するアプローチと手術器具などに取り付けたセンサ情報を利用するアプローチが挙げられるが，前者のアプローチはセンサ等の導入コストがかからない点や既に蓄積されている動画像記録に適用できる点で，後者のアプローチより優れている．

監視映像からのイベント検出として，人物などが多数存在する混雑シーンにおけるイベント検出と混雑シーンではないシーンにおけるイベント検出に関する先行研究について述べた．特に，近年では混雑シーンにおけるイベント検出タスクが注目を集めており，映像内の特徴量の分布を利用した手法や，スパースコーディングや Autoencoder などのような画像の再構築に基づいた手法などが提案されており，高い性能を示している．

また，映像を観測しながらイベント検出モデルの構築および更新が行われる，より汎用性の高い適応的なアプローチについても述べた．映像を観測しながらイベント検出モデルを構築するため，正例および負例のラベルデータが事前に与えられていなくても，イベント検出が可能である．また，これら適応的な手法では，天候による照明変化などに対応可能となるなどの利点が挙げられる．

第3章　正例および負例のラベルデータを用いたイベント検出

3.1　はじめに

近年，カメラなどのセンサの増加による手術室の高度化に伴い，これらセンサから取得した術中情報を用いた手術工程の解析が注目されている．手術工程の解析は手術技術の洗練化や客観的評価，手術工程の効率化の実現において重要である．また近年，導入が進んでいるコンピュータ支援システムにおいても正確な手術工程の理解は必要不可欠であるため，手術工程の解析はますます必要とされている．

しかし，得られる膨大な術中情報を人手で解析することは，大きな労力を必要とするため，計算機による自動解析の実現が求められている．このため，近年では動画像や音声など様々なセンサから得られた術中情報を用いた手術工程の自動解析が盛んに行われている．例えば，動画像記録内の画像情報を用いた手術工程の認識や特定の手術器具の検出などが挙げられる．本研究はこれらの取り組みのひとつであり，覚醒下脳腫瘍摘出術における皮質マッピング工程の動画像記録を用いた手術工程の自動解析の実現を目指している．また，皮質マッピング工程の自動解析の研究は文献を渉猟した限り行われていないため，自動解析が可能となれば覚醒下脳腫瘍摘出術の洗練化や手術工程の効率化が期待できる．

3.2　皮質マッピング工程における電気刺激位置の自動検出

提案する電気刺激位置の検出方法の概要を図 3.1 に示す．まず，電極先端位置の検出では，(1) 電極全体の形状と色特徴に基づいた検出，(2) 電極先端の形状と色特徴に基づいた検出，(3) 電極先端位置の追跡，の 3 手法による検出結果を統合することで最終的な電極先端位置を検出する．3 手法で検出を行う理由は，ある検出手法が検出に失敗しても他の手法で結果を補うことで，1 手法で行うよりも安定した検出を実現するためである．1 つの検出手法が誤検出を起こしやすい場面に対して，良好に検出を行える他の手法を用いることで精度向上が見込まれ，提案手法では特性の異なる 3 つの検出手法を用いることで検出精度の向上を行う．電極先端位置の検出では電気刺激が行われたタイミングに関係なく，すべてのフレームで電極先端位置の検出を行う．しかし，皮質マッピング工程の解析には電気刺激位置の取得，すなわち電気刺激を行ったタイミングでの電極先端位置の検出が必要である．そこで本研究では，検出した電極先端位置の周囲のオプティカルフローのヒストグラムを用いて電気刺激時特有の脳表面のへこみや電極の動きを表現する．このオプティカルフローのヒストグラムを特徴ベクトルとして SVM に入力することで，電気刺激終了タイミングを検出する．本研究では，検出した電気刺激終了タイミング時での電極先端位置を電気刺激位置とする．

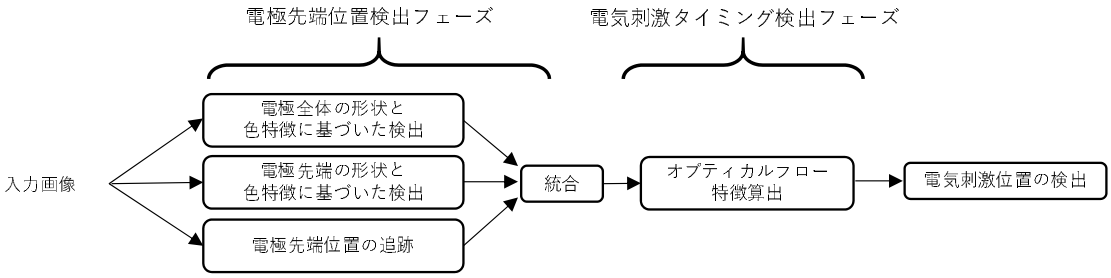


図 3.1: 提案手法の概要

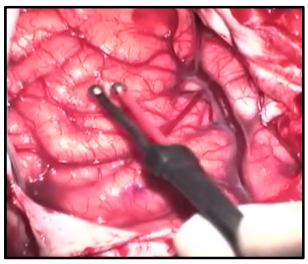


図 3.2: 皮質マッピングで用いる電極の例

3.2.1 電極先端位置の検出

電極全体の形状と色特徴を用いた検出

図 3.2 に示すように皮質マッピング工程で用いられている電極は柄の部分が黒く，その先端に電気刺激を行う電極先端部分が位置している．そこで電極全体の形状に基づいた検出手法では，この特徴に着目して電極先端位置を検出する．まず，背景差分処理によって移動物体である電極全体を抽出する．次に，抽出した移動領域から判別分析法によって二値化することで電極の黒領域を抽出する．抽出した黒領域の先端部分に電極の先端が位置していると考えられるため，抽出した黒領域の主軸を次式（3.1）によって算出し，主軸の先端を黒領域の先端とする．

$$y = x \cdot \tan\theta + a \tag{3.1}$$

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2m_{1,1}}{m_{2,0} - m_{0,2}}\right) \tag{3.2}$$

$$a = y_c - x_c \cdot \tan\theta \tag{3.3}$$

$m_{2,0}$，$m_{0,2}$ はそれぞれ抽出した黒領域の x 軸と y 軸方向の分散，$m_{1,1}$ は xy 軸方向の共分散，x_c，y_c は黒領域の重心位置である．これらのモーメントは次式によって算出される．

- 0 次モーメント

$$m_{0,0} = \sum\sum x^0 y^0 I(x,y) \tag{3.4}$$

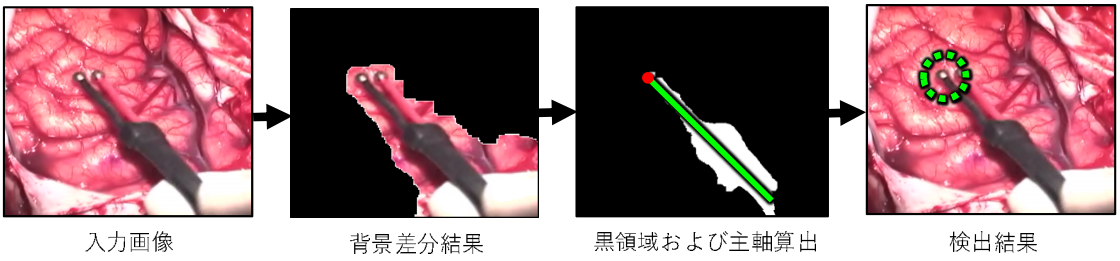


図 3.3: 電極全体の特徴に基づいた検出の処理結果例

- 1 次モーメント

$$m_{1,0} = \sum\sum x^1 y^0 I(x,y) \tag{3.5}$$

$$m_{0,1} = \sum\sum x^0 y^1 I(x,y) \tag{3.6}$$

- 重心からの 2 次モーメント

$$m_{2,0} = \sum\sum (x-x_c)^2 y^0 I(x,y) \tag{3.7}$$

$$m_{0,2} = \sum\sum x^0 (y-y_c)^2 I(x,y) \tag{3.8}$$

$$m_{1,1} = \sum\sum (x-x_c)^1 (y-y_c)^1 I(x,y) \tag{3.9}$$

- 重心

$$(x_c, y_c) = (\frac{m_{1,0}}{m_{0,0}}, \frac{m_{0,1}}{m_{0,0}}) \tag{3.10}$$

$I(x,y)$ は抽出した黒領域の画素値を表しており，ここでは $I(x,y) = 1$ としている．また，主軸角度 θ は x 軸に対する角度を表している．

図 3.2 に示すように，電極先端は照明の反射によって光り輝いている．そのため，黒領域の先端位置から 40 pixel の範囲内で最も近く，輝度値が 200 以上の位置を電極先端位置とする．図 3.3 にこれらの処理で得られる一連の結果画像例を示す.

電極先端の形状と色特徴を用いた検出

ここでは，histograms of oriented gradients（HOG）特徴 [59] と L*a*b*カラーヒストグラムを入力特徴量として Boosting の 1 手法である Real Ada Boost [84] を用いて電極先端を検出する識別器を構築する．

Boosting はアンサンブル学習法の 1 種で，性能の低い識別器（弱識別器）を複数組み合わせて 1 つの強力な識別器（強識別器）を構築する手法である．Boosting の 1 種である AdaBoost [85] では，検出対象データと対象以外のデータから構成される学習データに対して，それらを 2 クラスに識別

す弱識別器を複数構築する．このとき，前の弱識別器が正しく識別できなかった学習データの重みを高く，正しく識別できた学習データの重みは低くするように更新を行う．この弱識別器の選択と重みの更新を繰り返し行い，各弱識別器の信頼度に応じた重み付き線形和をとることで強識別器を構築するのが AdaBoost である．

本章で用いる Real AdaBoost は AdaBoost を拡張した手法である．AdaBoost は重み更新の際，学習データに対して正誤の 2 値判定を用いて更新が行われるため，弱識別器でどの程度学習データの識別が可能であるかの判定ができない．そこで，弱識別器の出力を特徴量の分布に応じて実数値化し，効果的な重みの更新を可能にしたのが Real AdaBoost である．人や車検出において従来の AdaBoost に比べ，少ない弱識別器で高精度な検出を可能としている．次に Real AdaBoost のアルゴリズムを示す．

STEP 1： N 個の各学習データ x_i に対象画像と対象以外の画像に対応したラベル $y \in \{+1, -1\}$ をつける．

STEP 2： 各学習サンプルの重み $D_t(i)$ を次式 (3.11) によって初期化する．

$$D_1(i) = \frac{1}{N}, (i = 1, 2, \ldots, N) \tag{3.11}$$

ここで，$D_1(i)$ は学習回数 1 回目の各学習データの重み，N は学習データの総数を表している．

STEP 3： 用いる特徴量をビンに変換し学習データの重みを足し合わせて，確率密度分布を作成する．1 つのある特徴量の確率密度分布は図 3.4 に示すように離散的に表現される．

$$W_+^j = \sum_{i: j \in J \wedge y_i = +1} D_t(i) \tag{3.12}$$

$$W_-^j = \sum_{i: j \in J \wedge y_i = -1} D_t(i) \tag{3.13}$$

ここで，J は特徴量の集合，$D_t(i)$ は学習回数 t 回目の各学習データの重みを表している．

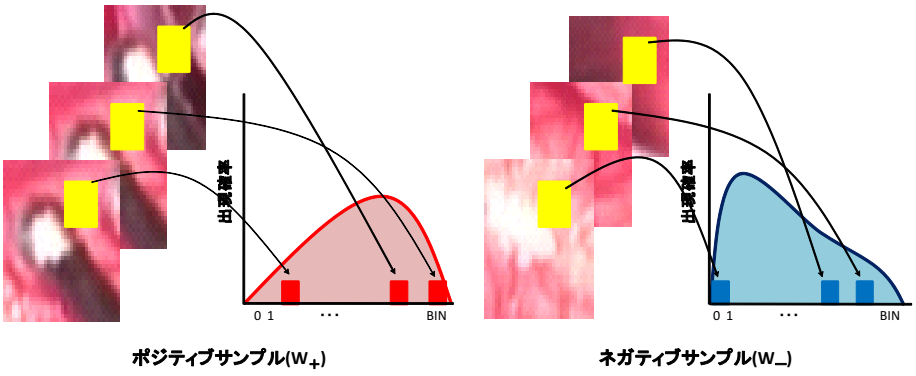


図 3.4: 確率密度分布

STEP 4： STEP3 で算出した 2 つの確率密度分布から Bhattacharyya 距離で類似度を算出する．図 3.5 に示すように，2 つの分布の類似度が低い場合は 2 つの分布は分離しやすく，対象物体と非対象物体の識別精度の高い特徴と判断することができる．逆に，図 3.6 に示すように 2 つの分布の類似度が高い場合は，対象物体と非対象物体の識別精度が低い特徴と判断すること

ができる．識別精度の評価値 z は次式 (3.14) によって算出される．この評価値 z が最も高い特徴を学習回数 t 回目における弱識別器とする．

$$z = 1 - \sum_j \sqrt{W_+^j W_-^j} \tag{3.14}$$

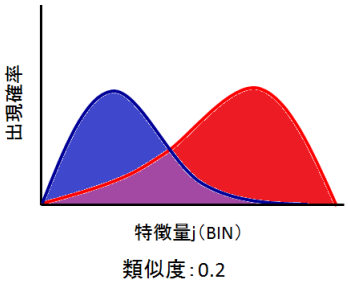


図 3.5: 分離しやすい分布

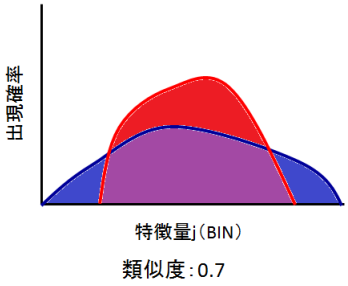


図 3.6: 分離しにくい分布

STEP 5： 従来の AdaBoost では STEP4 で選択した弱識別器が正しく分類できなかった学習データに対して重みを高くし，次の弱識別器では正しく分類されるように更新が行われる．そこで，Real AdaBoost では，それぞれのデータに対してどの程度の識別が可能であったかを STEP4 で選択された弱識別器の出力値を参照することによって把握し，効果的な重み更新を行う．各学習データに対する弱識別器の出力は式 (3.15) で表され，本論文では，$\epsilon = 1.0\times10^{-7}$ としている．

$$h_t(x_i) = \frac{1}{2} \ln \frac{W_+^j + \epsilon}{W_-^j + \epsilon} \tag{3.15}$$

弱識別器の出力値を用いた Real AdaBoost における重み更新は次式 (3.16) で表される．その後，式 (3.17) によって正規化する．

$$D_{t+1}(i) = D_t(i) \exp(-y_i h_t(x_i)) \tag{3.16}$$

$$D'_{t+1}(i) = \frac{D_{t+1}(i)}{\sum_{n=1}^{N} D_{t+1}(n)} \tag{3.17}$$

STEP 6： STEP3〜STEP6 を学習回数 T だけ繰り返し行い，使用する弱識別器を選択する．最終的に使用する強識別器は，STEP4 で選択された弱識別器の出力値の線形和で表される．強識別器 $H(x)$ は次式 (3.18) で表され，$H(x)$ がしきい値 λ より高ければ対象物体，低ければ非対象物体と判定される．

$$H(x) = \mathrm{sign}(\sum_{t=1}^{T} h_t(x) - \lambda) \tag{3.18}$$

本手法で用いる HOG 特徴は局所領域における輝度値の勾配方向をヒストグラム化した特徴量であり，大まかな物体形状を表現することが可能である．しかし，HOG 特徴はグレースケールの特徴量であるため，電極先端の色特徴を表現できない．そのため本手法では L*a*b*カラーヒストグ

(a) 電極先端画像　　(b) 非電極先端画像

図 3.7: 電極先端の学習画像例

ラムも用いることで電極先端の色特徴を表現する．識別器の構築に用いた学習画像例を図 3.7 に示す．学習画像のサイズはすべて 40×40 pixel である．識別器の構築には電極先端画像を $7,000$ 枚，非電極先端画像を $20,000$ 枚用いた．構築した識別器はウィンドウサイズを 30 pixel から 45 pixel の範囲で変えながらラスタスキャンを行うことによって適用される．これにより電極先端位置が複数のウィンドウによって検出されるため，Mean-Shift 法［86］を用いて周辺のウィンドウと統合する．この際の統合数が 3 以上かつ最大のウィンドウ位置を電極先端位置とする．

電極先端位置の追跡

これまでに述べた 2 手法はフレーム間の情報，すなわち時間情報を用いていない．本手法は動画像記録を対象としているため，時間情報は電極先端位置の検出に有効であると考えられる．本手法では，前フレームでの検出結果位置の周辺で電極先端を探索し，追跡を行う．探索は現フレームの追跡領域範囲内の各画素に HOG 特徴による重みを与え，Mean-Shift 法を用いることで行われる［87］．HOG 特徴量から求めた重み分布を用いて，現フレームでの追跡領域中心位置から重みの大きい位置へ移動するように以下の計算を繰り返す．

1. 重み分布の計算
   追跡中心 x の周辺画素 $\mathbf{x}_i (i = 0, \ldots, N)$ の HOG 特徴量と $\mathbf{x}$ における参照用モデル $\mathbf{v}$ との距離から重み $w(\mathbf{x}_i)$ を次式（3.19）によって求める．

$$w(\mathbf{x}_i) \quad = \quad \exp(-D(\mathbf{x}_i)^2) \tag{3.19}$$

$$D(\mathbf{x}_i) \quad = \quad \|\mathrm{HOG}(\mathbf{x}_i) - \mathbf{v}\| \tag{3.20}$$

   $\mathrm{HOG}(\mathbf{x}_i)$ は座標 $\mathbf{x}_i$ における HOG 特徴量を算出する関数である．

2. 移動量の算出
   求めた重み $w(\mathbf{x}_i)$ を用いて移動量 $\Delta\mathbf{x}$ を求める．前フレームの追跡領域周辺に注目した探索を行うために，式（3.21）で示すカーネル関数 $K(\mathbf{x})$ を用いて，追跡中心 $\mathbf{x}$ の移動量 $\Delta\mathbf{x}$ を式（3.22）によって求める．

$$K(\mathbf{x}) = 1 - \mathbf{x} \tag{3.21}$$

$$\Delta\mathbf{x} = \frac{\sum_{i=0}^{N} K(\mathbf{x}_i - \mathbf{x}) w(\mathbf{x}_i)(\mathbf{x}_i - \mathbf{x})}{\sum_{i=0}^{N} |K(\mathbf{x}_i - \mathbf{x})| w(\mathbf{x}_i)} \tag{3.22}$$

   ステップ 1，ステップ 2 の処理を $|\Delta\mathbf{x}| < 1$ を満たすまで繰り返し行い，追跡中心座標 $\mathbf{x}$ を求める．

3. 追跡の失敗判定
   移動後の座標 **x** での HOG 特徴を求め，式（3.20）によって求めた参照用モデルとの距離 $D(\mathbf{x})$ が閾値 d_m 以上の場合，追跡に失敗したと判定する．また，閾値が d_s 以下の場合は移動後の座標 **x** で参照用モデル **v** を更新する．

3 手法による検出結果の統合

電極先端位置の検出方法では，ここまでに述べた 3 手法による検出結果を Mean-Shift 法［86］によって統合し，その際の統合数が最大の位置を最終的な電極先端位置とする．統合に用いた Mean-Shift 法では，ある注目ウィンドウとその周囲のウィンドウ群との重心を求め，その重心に注目ウィンドウを移動させ，注目ウィンドウから 40 pixel の範囲内に存在する周囲のウィンドウ群と統合を行う．提案手法では最終的な電極先端位置をある 1 つの手法による検出位置にするのではなく，3 手法を考慮した検出位置にするために Mean-Shift 法を統合方法として採用した．

3.2.2 電気刺激終了タイミングの検出

電気刺激時には，脳表面には電極との接触によるへこみや照明変化が生じる．また，電気刺激の間，電極はほとんど動かず，刺激終了時に脳表面から離れる際に素早く動くといった特徴がみられる．本手法では，これらの特徴がみられる電気刺激の終了タイミングをオプティカルフローのヒストグラムによって表現し，SVM に入力することで電気刺激の終了タイミングを検出する．電気刺激の間，電極は静止しているため，刺激開始時と刺激終了時における電極の位置に変化は少ない．そのため，電気刺激の終了タイミングを検出することで，工程解析に必要な電気刺激位置を検出することが可能である．電気刺激終了タイミングの検出手法の概要を図に示す．まず，3.2.1 節で検出した電極先端位置を中心とする 100×100 pixel の矩形領域内の各点におけるマッチングコスト d_{SAD} を次式（3.23）に従って算出する．

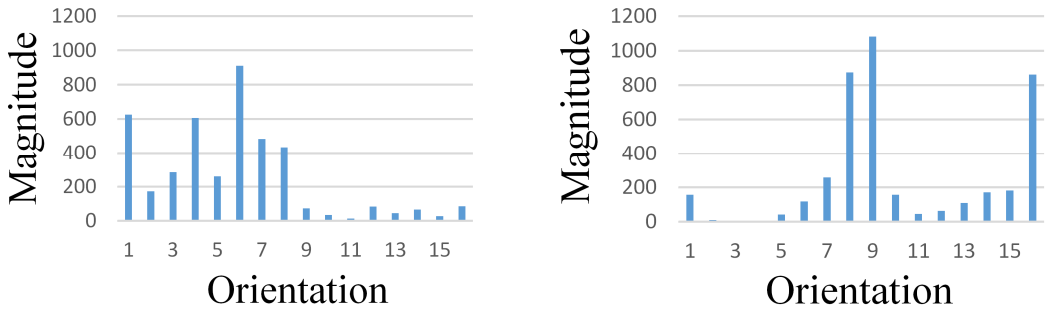
$$d_{SAD} = \sum_{(i,j)\in W} |I_t(x+i, y+j) - I_{t+1}(x+i, y+j)| \tag{3.23}$$

W は各点のマッチングコストを算出するためのウィンドウサイズ，I_t，I_{t+1} はそれぞれ t フレーム目での輝度値，$t+1$ フレーム目での輝度値を表している．各点のオプティカルフローは次フレームとのマッチングコスト d_{SAD} が最小の位置への動きベクトルである．次に，求めたオプティカルフローの方向 θ と移動距離の総和 M に関するヒストグラムを次式（3.24），（3.25）に従って作成する．

$$\theta(dx, dy) = \tan^{-1}\left(\frac{dy}{dx}\right) \tag{3.24}$$

$$M(dx, dy) = \sqrt{dx^2 + dy^2} \tag{3.25}$$

dx, dy はそれぞれオプティカルフローの x 軸方向の動きベクトル，y 軸方向の動きベクトルを表している．本手法では，方向 θ は 45° ごとに 8 方向に量子化するため，1 フレームにおけるオプティカルフローのヒストグラムは 8 次元の特徴ベクトルとなる．電気刺激終了時の特徴的な動作である静止状態から素早い動きへの切り替わりを表現するために，本手法では現フレームから前 10 フレームのオプティカルフローのヒストグラムも用いる．前 10 フレームという値は実験的に決定した．前 10 フレームのオプティカルフローのヒストグラムは方向ごとに移動距離の総和 M の値を平



(a) 電気刺激終了タイミング　(b) 通常

図 3.8: オプティカルフロー特徴の例

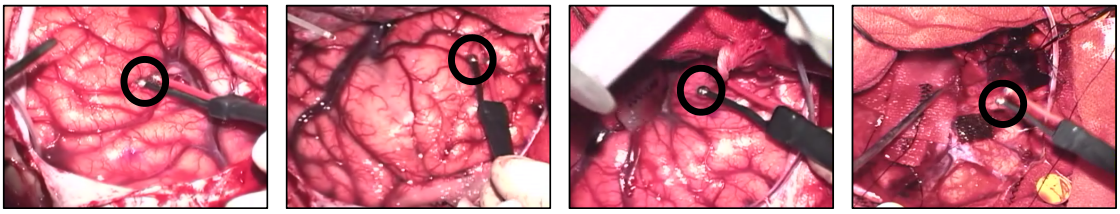


図 3.9: 電極先端位置検出の結果例（統合後）

均して用いる．そのため，各フレームにおけるオプティカルフロー特徴は 16 次元の特徴ベクトルとなる．現在のフレームから算出される 8 次元のオプティカルフロー特徴が電極と脳表面の激しい変化を表現し，前 10 フレームから算出されるオプティカルフロー特徴が刺激中の静止状態を表現することをねらいとしている．電気刺激終了タイミング検出に用いたオプティカルフロー特徴例を図 3.8 に示す．これらの特徴ベクトルを SVM へ入力することで，現在のフレームが電気刺激の終了タイミングであるか否かを識別する．

3.3　電気刺激位置の自動検出実験

3.3.1　概要

本手法の性能を検証するため，6 症例の皮質マッピング動画像記録（フレームレートは 30fps），計 11,000 フレームに対して (1) 電極先端位置の検出，(2) 電気刺激終了タイミングの検出，(3) 電極先端位置検出と電気刺激終了タイミング検出による電気刺激位置検出の 3 つの実験を行った．実験では 5 症例を学習データ，残りの 1 症例をテストデータとする 6 分割交差検定を行った．実験のためすべてのフレームに対して電極先端位置と電気刺激終了タイミングの正解ラベル付けを手作業で行った．(2) の実験では (1) で電極先端位置をすべて正しく検出できたフレーム群を用いて電気刺激終了タイミングの検出を行った．実験に用いた SVM には，$\gamma = 10.0$，$C = 10.0$ のガウシアンカーネルを使用した．本実験で用いたデータでは行われた電気刺激の回数は計 154 回であった．

表 3.1: 電極先端位置の検出結果

	統合	全体特徴	先端特徴	追跡
再現率	0.8201	0.7938	0.7263	0.7535
適合率	0.9516	0.8486	0.8194	0.8826
F 値	0.8810	0.8203	0.7700	0.8129

表 3.2: 刺激終了タイミングの検出結果

再現率	0.6964
適合率	0.7127
F 値	0.7045

表 3.3: 電極刺激位置の検出結果

再現率	0.5928
適合率	0.8977
F 値	0.6806

性能評価には次式で示される再現率 R，適合率 P，F 値 (F) を用いた．

$$R = \frac{correct}{S}, P = \frac{correct}{ALL}, F = \frac{2RP}{R+P} \tag{3.26}$$

$correct$ は正しく検出された数，S は検出すべきサンプル数，ALL はすべての検出数である．電極先端位置の検出実験においては，検出ウィンドウ内に電極先端がある場合を正しく検出されたとしている．電気刺激終了タイミングの検出実験では，電気刺激の終了タイミングとして検出したフレームの前後 5 フレーム以内に，正解ラベルをつけた電気刺激終了時のフレームが存在した場合を正しく検出されたと判定した．これは 5 フレーム以内の誤差であれば，工程解析に支障はないことが実験的に確認されているためである．

3.3.2 実験結果

まず，6 症例に対して行った電極先端位置の検出結果の平均評価値を表 3.1 に示す．表 3.1 では，3 手法を統合した結果（統合），電極全体の形状と色特徴に基づく検出だけを用いた結果（全体特徴），電極先端の形状と色特徴に基づく検出だけ用いた結果（先端特徴），電極先端位置の追跡だけを用いた結果（追跡）を示している．再現率，適合率，F 値のいずれも 3 手法を統合した結果が最も精度が高いことがわかる．3 手法を統合した後の検出結果例を図 3.9 に示す．照明条件や他の手術器具などの環境変化がある場合でも，正しく電極先端位置を検出できている．また各種パラメータを変更して検証を行ったところ，検出精度に大きなバラつきはみられなかった．

次に，電気刺激終了タイミングの検出実験で，6 症例に適用した平均評価値を表 3.2 に示す．再現率，適合率はそれぞれ 0.6964，0.7127 であった．

最後に，電極先端位置検出と電気刺激終了タイミング検出を組み合わせて，電気刺激位置の検出実験を行った際の平均評価値を表 3.3 に示す．再現率が 0.5928 と検出漏れが少し目立つ結果となった．これは電極先端位置は正しく検出できているが，電気刺激終了のタイミング検出で検出漏れが生じていることが原因である．しかし，適合率が 0.8977 であり，過検出は抑えられた結果となった．各評価値が 0.8 以上ならば，皮質マッピング工程の解析において有効性があると医師から評価を受けている．そのため，電極先端位置の検出においては有効性が示された．電気刺激終了タイミング検出については検出精度の向上が必要である．

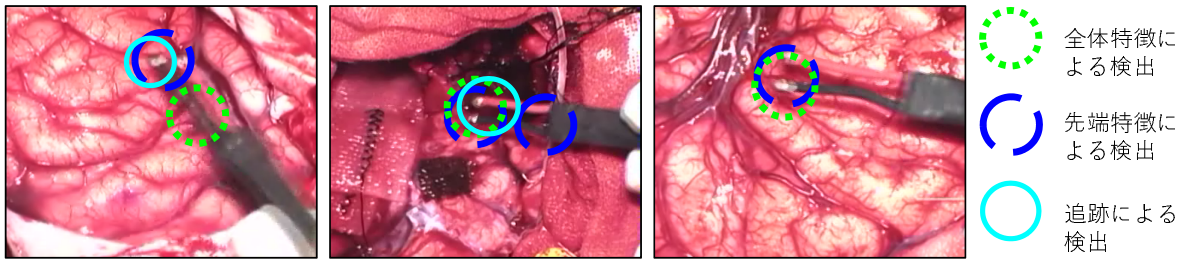
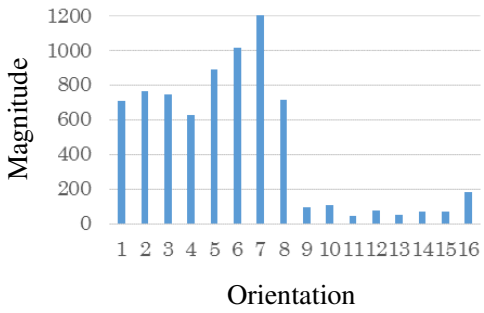


図 3.10: 電極先端位置検出の結果例．(a) 電極全体の特徴による検出失敗例．(b) 電極先端の特徴による検出失敗例．(c) 追跡失敗例．

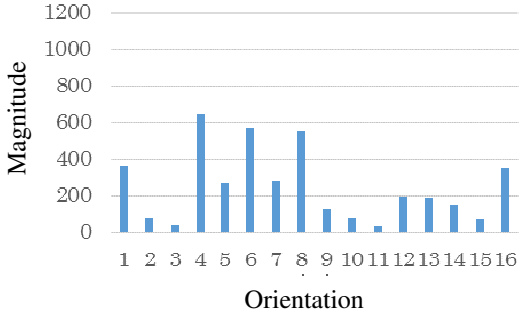
3.3.3 考察

まず，電極先端位置検出について，図 3.10 に 3 つの検出手法のうち 1 つの検出手法が検出に失敗したが，残りの 2 手法の結果を統合することで正しく検出できた例を示す．図 3.10（a）は電極の黒領域がぶれてしまっていることが原因で電極全体の特徴を用いた検出がうまく行えなかった例である．一方，他の 2 手法は電極先端の特徴を考慮した検出手法であるため，黒領域のぶれに影響されずに正しく電極先端位置を検出している．このため，最終的な統合結果は正しく検出できている．図 3.10（b）は電極先端の特徴を学習した識別器が過検出を起こしている例である．今回用いた HOG 特徴量はエッジベースの特徴量であるため，電極先端付近のエッジと似た領域で過検出が生じている．図 3.10（b）では，3 つのウィンドウが電極先端付近にあり，1 つのウィンドウが電極の柄の部分に出現しているが，本手法では Mean-Shift 法による統合数が最大のウィンドウ位置を最終的な検出結果とするため，正しい検出結果が得られる．図 3.10（c）は電極のフレーム間の動きが速く，前フレームの位置情報を使用している追跡が失敗してしまった例である．このようにフレーム間の動きが大きいとトラッカーの探索範囲に電極先端位置が存在しないため，追跡が行えない．しかし，他の 2 手法は前後のフレーム情報を用いないため，電極の動きがフレーム間で大きい場合でも電極先端位置を検出することが可能である．提案手法では，3 つの検出手法はそれぞれ異なった特徴を用いて検出を行っており，また他のある 1 手法が誤検出を起こしてしまう場面に対して，他の検出手法は正しく検出が行える特徴を用いている．そのため，上記で例を挙げたようにある 1 つの検出手法が誤検出を起こしてしまっても，他の 2 手法が正しく検出を行えている場面が多くみられ，3 手法を統合することで精度が向上したと考えられる．このように各手法が結果を補完し合うことで安定した検出を実現している．

図 3.11 に正しく電気刺激終了タイミングを検出できた時，検出漏れ時のそれぞれのオプティカルフロー特徴を示す．図 3.11（a）に示すように正しく検出できた例では，ヒストグラムのビン番号が 1 ~ 8 のうち複数の方向成分の移動距離の総和が高い値を示し，後半の 9 ~ 16 の成分は低い値を示している．これは前半の高い値を示した方向成分が電極や脳表面での大きな変化を表現し，後半の低い値を示した方向成分が刺激中の静止状態を適切に表現していると考えられる．一方，図 3.11（b）は電気刺激終了タイミング時のオプティカルフロー特徴であるが，ビン番号 1 ~ 8 の各方向成分の移動距離の総和の値が低いため，検出漏れとなってしまった例である．このフレームでは，脳表面のへこみや照明変化が弱く，脳表面上の形状変化をオプティカルフローで正確に記述することができなかったことが原因であると考えられる．また脳表面は類似した色情報をもつ領域が多く存在するため，オプティカルフローを求める際のマッチングコストを正確に算出できなかった



(a) 正しく刺激終了タイミングを検出できた例



(b) 検出漏れの例

図 3.11: オプティカルフロー特徴の例

ことも考えられる．

3.4 まとめ

本章では，覚醒下脳腫瘍摘出術における皮質マッピング工程の動画像記録から，工程解析に重要である電気刺激位置の自動検出手法を提案し，実験によって性能の検証を行った．提案手法は電極先端位置の検出と電気刺激終了タイミング検出の 2 段階から構成される．電極先端位置の検出では，電極全体の形状と色特徴に着目した検出，電極先端の形状と色特徴に着目した検出，電極先端位置の追跡，の 3 手法による検出結果を統合することで高精度な電極先端位置検出を実現した．また電気刺激終了タイミング検出では，検出した電極先端周辺のオプティカルフローの分布を特徴量とし，SVM へ入力することで電気刺激終了タイミングの検出を行った．これら電極先端位置検出と電気刺激終了タイミング検出を組み合わせることで電気刺激位置の検出を行い，F 値 0.6806 の識別率を示した．

今後の課題として，高精度な電気刺激終了タイミングの検出手法の提案が挙げられる．特に，本手法における電気刺激終了タイミングの検出では，電極先端周辺のオプティカルフローしか特徴量として用いていないため，電極全体のオプティカルフローの分布を用いるなど，他の特徴量の追加の必要があると考えている．また皮質マッピング工程の解析には，電気刺激位置の脳表面での三次元位置が必要である．そのため，今後は動画像内の特徴的な血管やしわなどの情報を用いて三次元の脳モデルとのマッチングを行い，検出した二次元位置を脳表面での三次元位置に変換する必要がある．さらに今後は，症例数を増やして実験を行うことで，より信頼性のある検証を行う．

第4章　正例のラベルデータのみを用いたイベント検出

4.1　はじめに

本章では，検出対象である顕著性のあるイベント（負例）のラベルデータを用いずに，顕著性のないイベント（正例）のラベルデータのみを用いたイベント検出を行う．本章では，ニーズの高い監視映像からの異常検知問題を扱う．そのため，以降では顕著性のあるイベントを異常，顕著性のないイベントを正常と呼ぶこととする．

近年では防犯を目的として，駅や空港などの公共施設だけではなく，マンションやビルなどの一般的な場所においても多くの監視カメラが設置されている．このように多くの監視映像が存在するため，監視者がすべての映像を監視することは不可能に近い．また，少数の映像の監視だとしても，長時間の監視を行うには膨大な労力を必要とする．そのため現状では，映像をみながら異常が発生した際にアラートを行うのではなく，異常が発生した後に異常イベントについての調査や確認の目的で監視映像が使用されていることが多い．

そこで，異常が発生した際にリアルタイムでアラートを行うための異常検知に関する研究が盛んに行われている．異常検知を行うモデルを構築する際に，事前に多くの異常サンプルを集めることが困難であることから，先行研究の多くは正常サンプルのみから正常モデルを構築し，その正常モデルにおける生起確率が低いパターンや，モデルから逸脱するパターンを異常として検出するアプローチが多く提案されている．しかし，2 章で述べたように多くの先行研究では，異常検知モデルの構築に用いる特徴量を人手で決定しており，その特徴量が最適であるとは限らない．また，歩行者などが多く登場する混雑シーンでは，事前に特徴量を設計することは困難であると考えられる．

そこで本章では，事前に明示的に特徴量を与えることなく，高精度に異常検知が可能なより汎用性の高い異常検知モデルの提案を行う．実験では，歩行者が多く登場する混雑シーンにおける異常検知実験において，先行研究と比較を行うことで性能検証を行う．

4.2　Convolutional Autoencoder による異常検知

本手法では，映像内の正常性を表現するために，Convolutional Autoencoder（CAE）を用いる．正常データのみを用いて学習を行った CAE は，正常データに対しては小さな再構築誤差，異常データに対しては大きな再構築誤差を示すことが期待できる．そのため，本手法では CAE による再構築誤差を指標として，異常検知を行う．また，本章では全結合層を含まない Fully CAE を用いることで，空間情報が失われることを防ぐ．これは入力画像を再構築する際に，空間情報が必要となるためである．

本手法で用いる CAE の構造を図 4.1 に示す．符号化部分は複数の畳込み層と Pooling 層によって構成されており，復号化部分は畳込み層と Unpooling 層によって構成されている．CAE の入力に

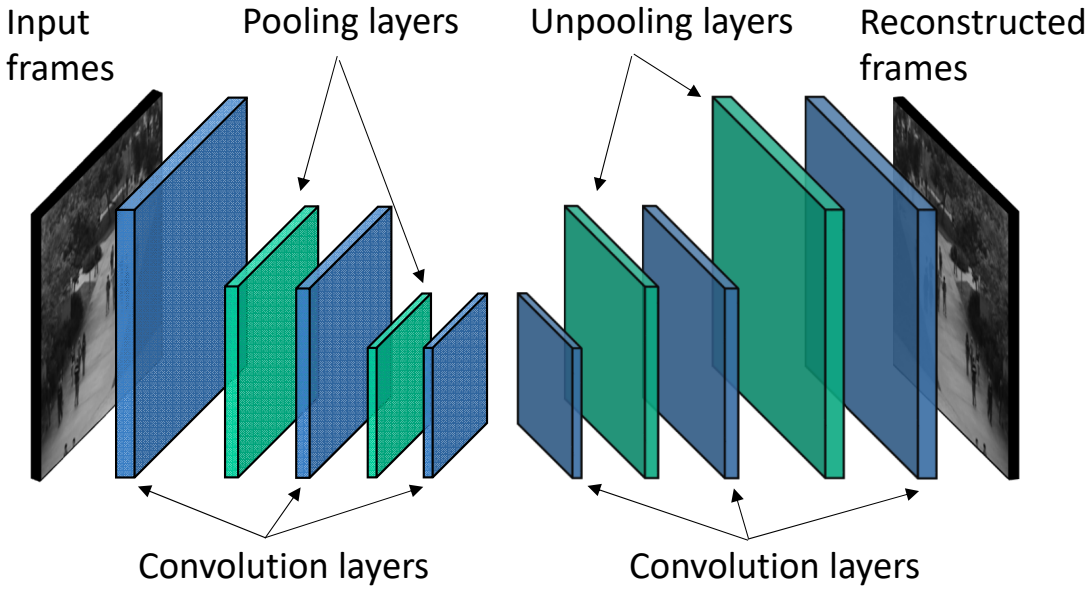


図 4.1: 本研究で用いる CAE の構造

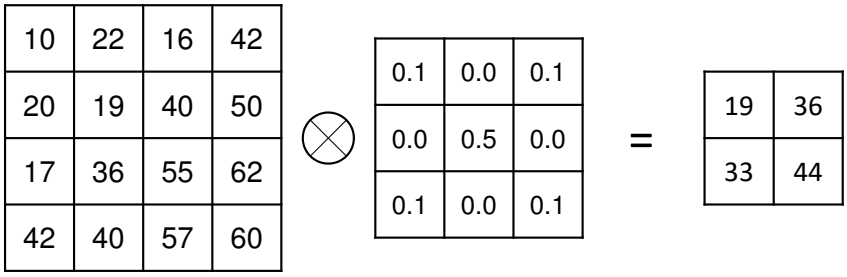


図 4.2: 画像サイズ 4×4 画素の入力画像とサイズ 3×3 画素のフィルタの畳込みによって生成される画像の例

は，映像内から時間方向のスライドウィンドウによって切り出したパッチを用いる．次節以降で，本手法で用いる CAE の構造や学習データ，学習の流れについて詳しく述べる．

4.2.1 CAE の構造

畳込み層

画像の畳込み処理は，画像とフィルタ間で行われる次式（4.1）の積和計算を表す．

$$u_{i,j} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i-p,j-q} h_{pq} \tag{4.1}$$

ここで，$x_{i,j}$ は $W \times H$ 画素の画像 $\mathbf{x}$ における画素 (i, j) の画素値，h_{pq} は $H \times H$ 画素のフィルタにおける画素 (p, q) の画素値を表す．画像の畳込み処理は，適用するフィルタの濃淡パターンと類似

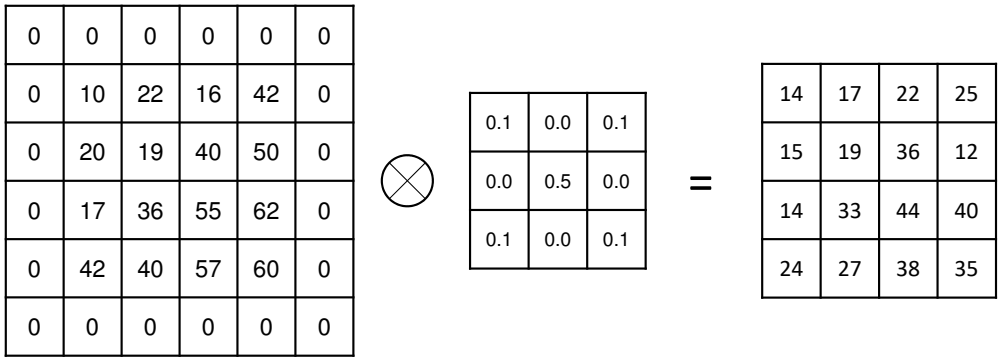


図 4.3: 画像サイズ 4 × 4 画素の入力画像にゼロパディングを行った後に，サイズ 3 × 3 画素のフィルタの畳込みを行って生成された画像の例

したパターンが入力画像内のどの部分に存在するかを検出する働きがあるため，フィルタがもつ特徴的なパターンを入力画像内から抽出することが可能である．畳込み処理の例を図 4.2 に示す．図 4.2 に示したように，一般的に入力画像とフィルタが重なり合うように積を求めるため，フィルタ適用後の画像のサイズは入力画像より小さくなる．このとき，フィルタの適用前後で画像サイズを変更させない方法として，ゼロパディングと呼ばれる方法がよく用いられている．ゼロパディングでは，入力画像のまわりに画素値 0 を挿入してから，フィルタによる畳込み処理を行うことで画像サイズの変更を防ぐ．図 4.3 にゼロパディングを行った際の例を示す．また，フィルタを入力画像に適用する際のずらし幅のことをストライドと呼び，ストライドを s としたときの出力画像の画素値 (i, j) は次式（4.2）で表される．

$$u_{i,j} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{si-p,sj-q} h_{pq} \tag{4.2}$$

この式から，出力画像サイズは入力画像サイズに対して $1/s$ 倍になることがわかる．

畳込み層は上述した畳込み処理を行う層である．実際の畳込み層では 1 枚の入力画像に対して 1 つのフィルタを適用するのではなく，複数チャネルの入力画像に対して複数個のフィルタを適用することが一般的である（図 4.4）．図 4.4 の例では，K チャネルの入力画像に対して 3 個のフィルタを適用し，3 チャネルの出力画像を出力している．3 種類の各フィルタは入力画像と同じチャネル数 K をもち，それぞれ 1 チャネルの特徴マップ u_{ijc} を次式（4.3）に従って出力する．

$$u_{ijc} = \sum_{k=0}^{K} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} v_{i-p,j-q,k} h_{pqkc} + b_{ijc} \tag{4.3}$$

b_{ijc} はバイアスを表す．続いて，式（4.3）によって算出された特徴マップに対して活性化関数 f を適用し，適用後の値が畳込み層の最終的な出力となる．活性化関数には，Rectified Linear Units（ReLU）[88]や tanh などがよく用いられている．ゼロパディングの適用およびストライド 1 の場合，$H \times H \times K$ の入力画像は畳込み層によって，$H \times H \times C$ の特徴マップを出力することになる．ここで C は畳込み層のフィルタの種類の数を表す（図 4.4 の場合，$C = 3$）．

本手法で用いる CAE の畳込み層では，ストライド 1 の畳込み処理の後，Batch Normalization [89] を行い，Rectified Linear Units（ReLU）を適用する．Batch Normalization は，学習を行う際にミニバッチごとに正規化を行うことで，Deep Neural Network（DNN）の学習を安定化させるための手

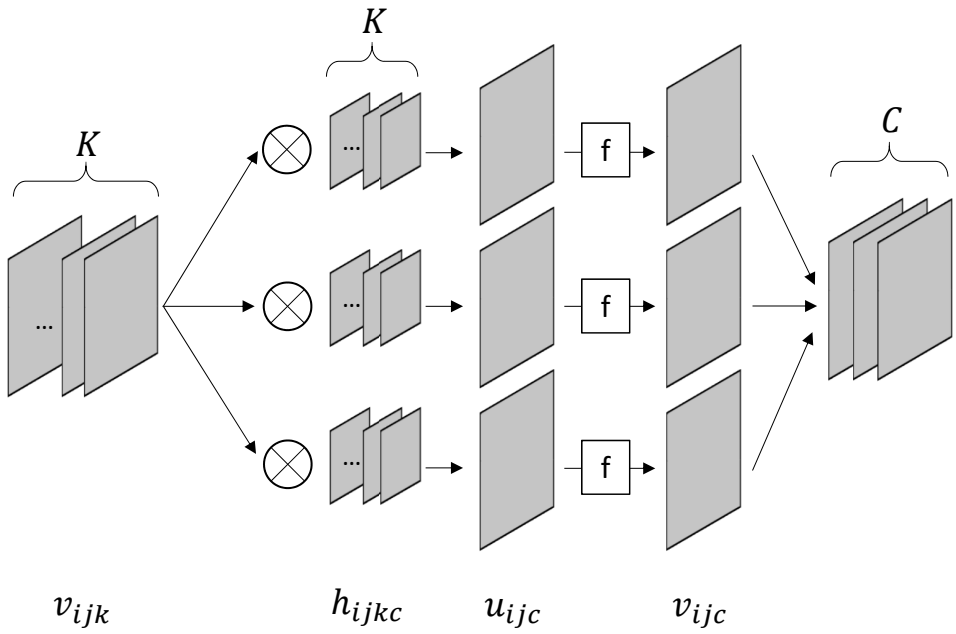


図 4.4: 畳込み層の概要

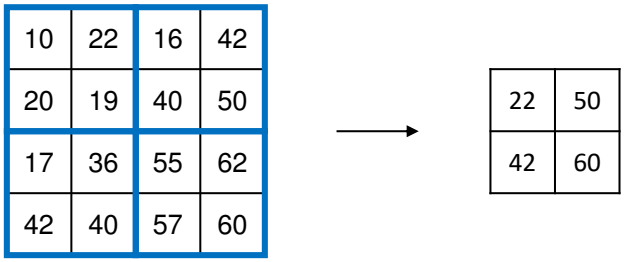


図 4.5: Max pooling の例

法である．ReLU は次式（4.4）で表される活性化関数であり，勾配消失を防ぐことが可能である．

$$f(x) = \max(0, x) \tag{4.4}$$

また，畳込み処理の前に，特徴マップに対してゼロパディングを行うことで，処理の前後で特徴マップの大きさが変わらないようにする．つまり，$M \times N \times K$ の入力特徴マップは，畳込み層によって $M \times N \times C$ の出力となる．ここで，M，N，K は入力特徴マップの幅，高さ，チャネル数をそれぞれ表し，C は出力チャネル数を表す．

Pooling 層

プーリング処理では，畳込み層などで抽出された特徴を局所ごとにまとめることで，特徴マップ内での平行移動に対する不変性を獲得することができる．また，一般的にはプーリング処理によって特徴マップのサイズが小さくなるため，DNN のパラメータ数を減らす役割も果たす．図 4.5 に

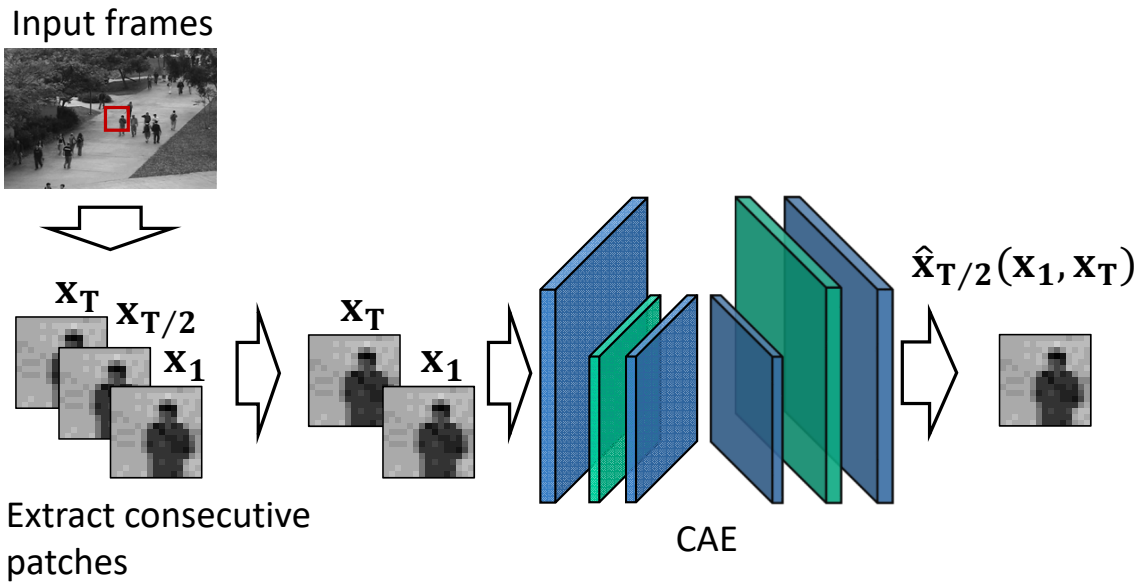


図 4.6: 提案手法における学習方法

最大プーリング（Max pooling）の例を示す．Max pooling では，特徴マップ内の局所領域での最大値を出力する．図 4.5 の例では，2×2 にフィルタをストライド 2 で適用しているため，プーリング後の特徴マップの幅と高さのサイズはそれぞれ 1/2 になる．Unpooling はプーリング処理の逆の処理になる．そのため，2×2 のフィルタをストライド 2 で適用した場合，Unpooling 後の特徴マップの幅と高さはそれぞれ 2 倍になる．

本手法で用いる Pooling 層では，2×2 のフィルタをストライド 2 で Max pooling を適用する．そのため，$M \times N \times C$ の入力特徴マップは，Pooling 層によって $M' \times N' \times C$ の出力となる．ここで，$M' = \lfloor M/2 \rfloor$，$N' = \lfloor N/2 \rfloor$ をそれぞれ表す．また，Unpooling 層では，2×2 のフィルタをストライド 2 で適用する．そのため，$M \times N \times C$ の入力特徴マップは，Pooling 層によって $M'' \times N'' \times C$ の出力となる．ここで，$M'' = 2M$，$N'' = 2N$ をそれぞれ表す．

4.2.2 CAE の学習方法

入力データと出力目標

CAE による異常検知の先行研究では，入力画像のチャネル数と出力画像のチャネル数は同一であり，かつ入力画像を再構築するように最適化を行っている．しかし，本手法では，CAE における入力画像のチャネル数と出力画像のチャネル数が異なる．また，入力画像と同一の画像を出力目標とするのではなく，入力画像に存在しないチャネルの画像を出力目標とする．具体的には，図 4.6 に示すように，入力動画から抽出した T フレーム分のパッチのうち，1 および T フレーム目のパッチを CAE に入力し，間の $T/2$ フレーム目のパッチを再構築するように CAE の学習を行う．異常検知においては，映像内の物体の形状情報と速度情報が重要であるため，先行研究では HOG 特徴やオプティカルフローなどの特徴量を用いることが多い．しかし，これらの特徴量が対象とする映像に対して最適であるとは限らない．そこで本手法では，上述したように入力には存在しないフ

レーム情報を出力目標として CAE を学習を行うことで，映像内の物体の形状と速度に関する特徴量を明示的に与えることなく，学習によって獲得することをねらっている．これにより，あらかじめ学習に用いる特徴量を人手で設計することなく，end to end で異常検知が行える柔軟な手法になると考えている．

CAE の最適化

CAE の関数には次式（4.5）で示される平均二乗誤差を用いる．

$$E(\mathbf{w}) = \frac{1}{2N}\sum_{i} \|\mathbf{x}^i_{T/2} - f_W(\mathbf{x}^i_{1,T})\|^2 \tag{4.5}$$

N はミニバッチのサイズ，$\mathbf{x}^i_t$ は i 番目の入力データで t フレーム目のパッチ，f_W は CAE をそれぞれ表す．ゆえに本手法では，式（4.5）を最小化する f_W（CAE）を学習することを目的とする（式（4.6））．

$$\hat{f}_W = \arg\min_W \frac{1}{2N}\sum_{i} \|\mathbf{x}^i_{T/2} - f_W(\mathbf{x}^i_{1,T})\|^2 \tag{4.6}$$

CAE の最適化には，Momentum stochastic gradient descent（Momentum SGD）を用いる．Adam 法 [90]や AdaGrad [91]，RMSProp [92]などのオプティマイザも適用したが，実験的に Momentum SGD を使用することを決定した．Momentum SGD の慣性項は 0.9，ミニバッチサイズは 128，weight decay は 0.0005 で計 250 epoch の学習を行った．学習率は 0.01 で学習を開始し，5 epoch 目で 0.1，125 epoch 目で 0.01，200 epoch 目で 0.001 に設定した．

異常判定

提案手法では，CAE による再構築誤差を異常判定の指標とする．具体的には，式（4.7）に算出される二乗誤差がしきい値 ϵ を超えた場合に，異常であると判定する．

$$E(\mathbf{x}) = \|\mathbf{x}^i_{T/2} - f_W(\mathbf{x}^i_{1,T})\|^2 \tag{4.7}$$

ここで f_W は学習済みの CAE を表す．

4.3 混雑シーンにおける異常検知実験

4.3.1 データセット

提案する CAE による異常検知の有効性を検証するために，公開されているデータセットである UCSD pedestrian dataset に提案手法を適用し，先行研究との比較を行う．UCSD pedestrian dataset は，多数の歩行者や車両が登場する混雑シーンの映像であり，Ped1 と Ped2 の 2 種類の映像を含んでいる．図 4.7 に Ped1 および Ped2 の映像例を示す．図 4.7 に示すように，Ped1 は道路に対して正面から撮影した映像，Ped2 は道路に対して横から撮影した映像になっている．このデータセットにおける異常例として，自転車，自動車，スケートボーダー，車椅子や，歩行者の異常行動（芝の中を歩く，道路を横切る）が挙げられる．Ped1 は 34 シーケンスの学習用データ，36 シーケンスのテスト用データで構成されており，各フレームサイズは 238×158 画素である．同様に，Ped2 は

図 4.7: UCSD pedestrian dataset（Ped1，Ped2）の例

16 シーケンスの学習用データ，12 シーケンスのテスト用データで構成されており，各フレームサイズは 360×240 画素である．各シーケンスは 120 – 200 フレームで構成されている．学習用データは，異常を含まない正常フレームのみで構成されており，学習用データを用いて異常検知モデルの学習を行う．テスト用データには，異常を含む異常フレームが約 3400 フレーム，異常を含まない約 $5,500$ フレーム分の正常フレームが存在する．

CAE の学習のために，Ped1 および Ped2 の学習用データからそれぞれパッチを抽出し，CAE の学習をそれぞれ行う．パッチは映像内をウィンドウを走査することで抽出される．パッチのサイズは 20×20 画素とし，$T = 5$ フレーム分のパッチを取得する．すなわち，1 フレーム目と 5 フレーム目のパッチを CAE の入力に，3 フレーム目のパッチを CAE の出力目標とした．この走査を学習画像のすべてについて行い，学習用のパッチを取得した．ウィンドウのストライドは 10 とした．

4.3.2 評価方法

このデータセットにおけるタスクは，テスト用データの各フレームに異常が存在するかどうかを検出することである．本章では，先行研究と同様に，各フレームに異常が存在するかどうかのフレーム単位での評価を行う．提案手法ではフレーム内のパッチごとに異常か正常かの判定を行うため，少なくとも 1 つ以上のパッチが提案手法によって異常と判定された場合，そのフレームを異常フレームと判定する．

評価指標として，Area Under the Curve（AUC）と Equal Error Rate（EER）を用いる．AUC は Receive Operating Characteristic（ROC）曲線下の面積を表し，値が 1.0 に近いほど異常検知性能が優れていることを表す．ROC 曲線は次式で算出される True Positive Rate（TPR）と False Positive Rate（FPR）によって作成される．

$$TPR = \frac{the\ number\ of\ true\ positive\ frame}{the\ number\ of\ positive\ frame} \tag{4.8}$$

$$FPR = \frac{the\ number\ of\ false\ positive\ frame}{the\ number\ of\ negative\ frame} \tag{4.9}$$

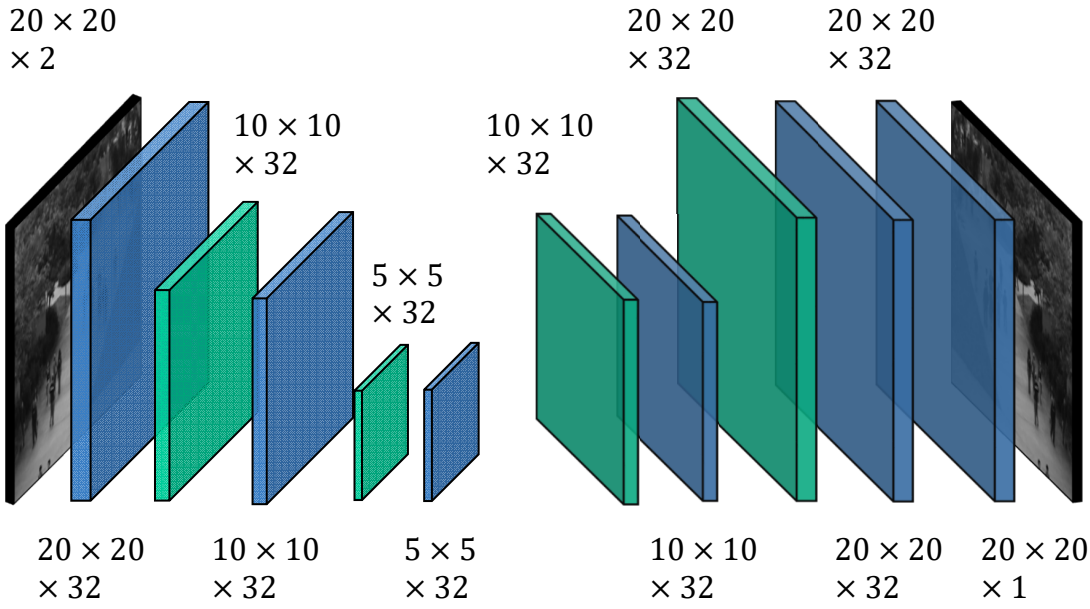


図 4.8: 本実験で用いた CAE の構造

なお，異常は positive，正常は negative として示している．EER は FPR= 1−TPR のときの誤検出率を表し，0.0 に近いほど異常性能が高いことを示す.

4.3.3 実験設定

図 4.8 に本実験で用いた CAE の構造を示す．図中の $w \times h \times c$ は，w，h，c がそれぞれ特徴マップの幅，高さ，チャネル数を表す．最終層以外には ReLU を活性化関数として用い，最終層は tanh を活性化関数とした用いる．また，ボトルネック部分（$5 \times 5 \times 32$）の畳込み層は 3×3 画素のフィルタを用い，それ以外の畳込み層は 5×5 画素のフィルタを用いる．各畳込み処理前にゼロパディングを行い，特徴マップのサイズが変わらないようにしている．

本手法は Chainer［93］（version 1.16.0）にて実装を行い，3.2 GHz CPU，32 GB RAM，NVIDIA GTX 1080 を搭載した計算機にて実験を行った.

4.3.4 実験結果

表 4.1 に提案手法および先行研究による UCSD pedestrian dataset における異常検知性能の結果を示す．提案手法をテスト用データに適用する際は，学習のときと同様にウィンドウをストライド 10 で走査してパッチを取得し，CAE に入力することで適用を行った．表 4.1 から，提案手法は Ped2 において最も性能が優れていることがわかる．特に，AUC および EER の両指標においてすべての先行研究より優れた結果を示している．図 4.9 に提案手法による異常検知結果の例を示す．図 4.9 に示すように，異常対象である自転車や車両，スケートボーダを良好に検出できていることがわかる．特に，スケートボーダは歩行者と形状特徴が類似しているため，速度情報を考慮しなければ検出できないと考えられるが，提案手法ではスケートボーダに対しても良好に反応を示せている．このことから，オプティカルフローなどのような速度情報を明示的に与えなくても，速度の違いを異

表 4.1: UCSD pedestrian dataset における定量評価結果

	Ped1(frame)		Ped2(frame)	
	AUC	EER	AUC	EER
Social force［94］	0.675	0.31	0.556	0.42
Social force + MPPCA［37］	0.668	0.32	0.613	0.36
MPPCA［29］	0.590	0.40	0.693	0.30
MDT［37］	0.818	0.25	0.829	0.25
TCP［3］	0.957	0.08	0.884	0.18
Conv-AE［58］	0.810	0.28	0.900	0.21
AMDN［2］	0.921	0.16	0.908	0.17
Proposal	0.720	0.34	0.934	0.14

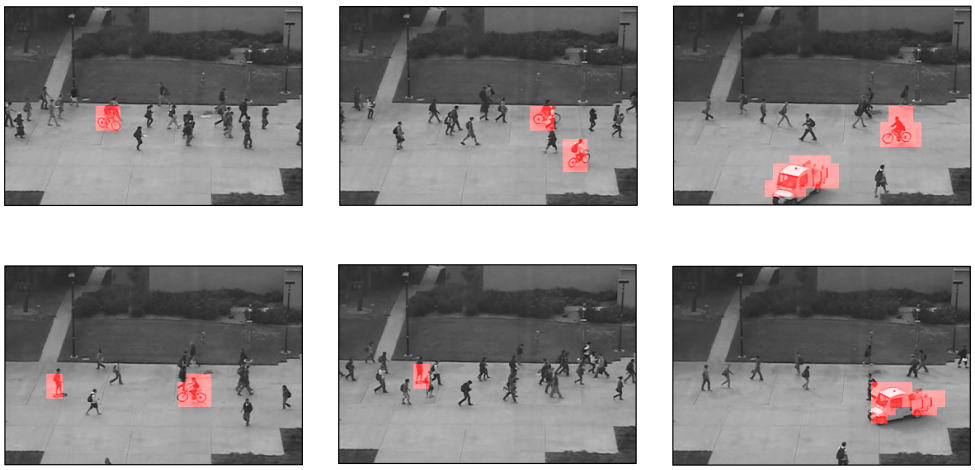


図 4.9: Ped2 における異常検知の結果例

常として適切に捉えることができることが確認できた．このように，提案手法では速度特徴を人手によって与えるのではなく，CAE によって学習を行うことで，より適切な速度特徴を獲得することができ，先行研究よりも高性能な結果が示せたと考えられる．

一方，表 4.1 の結果から，Ped1 において提案手法は最先端の先行研究と比べて優れた性能を示すことができていない．これは Ped1 では適切な速度情報を学習できていないことが原因として考えられる．Ped1 と Ped2 の違いとして，カメラに対する物体の移動方向が挙げられる．Ped1 ではカメラに対して手前から奥方向もしくは奥方向から手前に向かって物体が移動するシーンが多いが，反対に，Ped2 ではカメラに対して左方向から右方向もしくは右方向から左方向に向かって水平に物体が移動するシーンが多い．このとき，水平方向に移動している物体が多い Ped2 の方が，画像内の位置によって速度の捉え方が不変なため，速度特徴を学習しやすいと考えられる．逆に，Ped1 では，画像内の上領域と下領域では速度の捉え方が異なる（下領域の方が上領域に比べて 1 フレームあたりに移動する量が大きい）ため，速度特徴の学習がうまく行えずに，性能が低下してしまったと考えられる．そのため，今後は Ped1 のような縦構造の映像に対しても適切に速度特徴を学習

図 4.10: Ped1 における異常検知の結果例

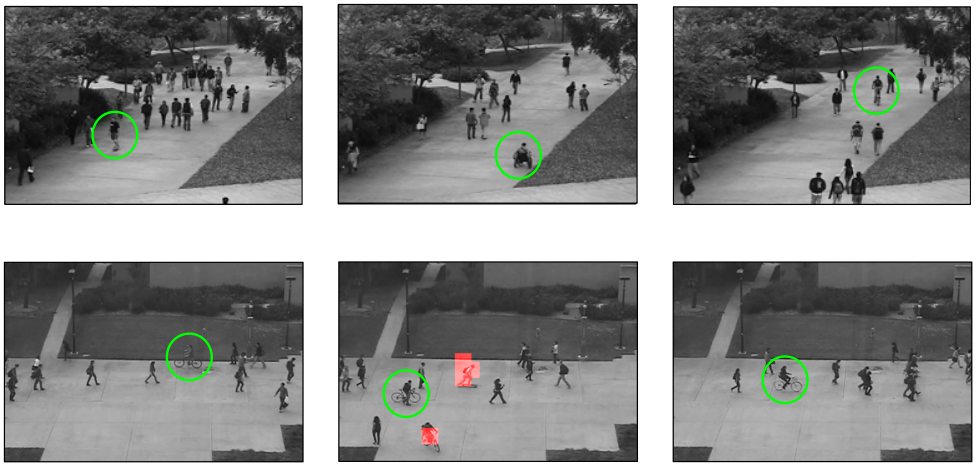


図 4.11: 提案手法による検出漏れの例

できるような改良が必要である．例えば，画像内の位置によってパッチ取得時のフレーム数 T を変更する，パッチサイズを可変にする，などが挙げられる．Ped1 において良好に異常検出できた例を図 4.10 に示す.

また，Ped1 および Ped2 で共通して検出漏れをしてしまっている例として，速度が歩行者と類似している自転車やスケートボーダ，車椅子が挙げられる．これら検出漏れの例を図 4.11 に示す．このことから，提案手法では先行研究と比べて形状特徴をうまく学習できていないことがわかる．特に，歩行者と速度が類似している車椅子を検出するには，形状特徴の違いを捉えるしかないと考えられる．本章では明示的に形状特徴を与えずに学習を行っているため，歩行者と速度が類似した異常対象に対して検出漏れが生じてしまったと考えられる．そのため，今後は速度特徴だけでなく，形状特徴も適切に学習可能な方法を検討する必要がある．

4.4 まとめ

本章では，検出対象である異常（負例）のラベルデータを用いずに，正常（正例）のラベルデータのみを用いた監視映像からの異常検知を行った．提案手法は Convolutional Autoencoder（CAE）の再構築誤差を利用して，異常検知を行う．先行研究では，異常検知に重要な形状特徴や速度特徴をあらかじめ人手によって与えている手法が多いが，本手法ではこれらの特徴を与えずに異常検知を行う，より柔軟な異常検知手法の提案を行った．UCSD pedestrian dataset（Ped1，Ped2）に提案手法を適用した結果，Ped2 のデータに対してはすべての先行研究よりも優れた性能を示した．このことから，あらかじめ特徴を与えなくても良好に異常検知が行えることが確認できた．一方，Ped1 のデータに対しては検出漏れが目立つ結果となり，Ped1 のような縦構造の映像に対する異常検知性能の向上が今後の課題として挙げられる．また，Ped1 および Ped2 で共通する課題として，速度特徴が類似している異常対象を検出漏れしてしまう点が挙げられる．そのため，今後は形状特徴もより適切に学習できるような枠組みを検討する必要がある．

第5章　正例および負例のラベルデータを用いないイベント検出

5.1　はじめに

本章では，検出対象である顕著性のあるイベント（負例）および顕著性のないイベント（正例）のラベルデータを用いない監視映像からのイベント検出を行う．前章で扱った問題設定では，事前に正例のラベルデータが用意されており，そのデータに基づいて構築した正常モデルを用いてイベント検出を行っている．また，構築した正常モデルは固定であり，イベント検出の適用中にそのモデルが更新されることはなかった．しかし，現実世界での運用を考えると，事前に正例および負例の定義ができない場合や，時間の経過に伴う環境変化によって正例および負例の定義が変わってしまう場合（例えば，天候による照明変化など）や，学習データに存在しなかった正例データが出現する可能性なども考えられる．そのため，事前に正例および負例データを用意することができない場合や，事前に構築した固定のイベント検出モデルの適用だけでは不十分であるという課題が挙げられる．したがって，イベント検出をより頑健に行うには，適用中の環境に応じてモデルの構築および更新が行われる環境に適応的な手法が望まれる．これまでに環境に適応的な手法が提案されており，異常検知問題に適用されている．本章でもこれら先行研究と同様に，監視映像からの異常検知問題を扱うため，本章以降では顕著性のあるイベント（負例）を異常，顕著性のないイベント（正例）を正常と呼ぶこととする．

環境に適応的な異常検知手法の例としては Masland らの Grow When Required（GWR）ネットワークが挙げられる［67–69］．GWR ネットワークでは頻繁に観測される入力刺激に対して，馴化モデルである Stanley モデル［70］を用いてネットワークの出力を次第に減少させていくことで環境の正常性を表現する．GWR ネットワークを移動ロボットに搭載し，ソナーセンサ情報を入力刺激として扱った実験では，ロボットがそれまで観測していた環境とは異なる環境に置かれるとネットワークが強い出力を示すことが確認されている．武田らは環境からの入力パターンに対する反応とそのパターンが生じる領域に対する反応の抑制によって，環境に変動を含む監視映像内の侵入物体を検知するネットワークモデルを提案している［4］．Staufer らは画素ごとに混合ガウス分布を用いて環境の背景のモデル化を行っている［76］．この手法では，混合ガウス分布を用いることで背景の揺らぎなどを多峰性の分布で表現しつつ，パラメータ更新によって環境変化への適応も実現している．これらの手法は固定カメラからの映像を用いた侵入物体検知問題に対して優れた性能を示している［4］．しかし，武田らの手法では映像内の位置関係を環境の正常性を表現するために利用していること，Staufer らの手法では複雑な入力パターンを多峰性の分布で表現することが困難なことから，映像内の背景が周期的に変動する旋回カメラなどで撮影された映像に対して適用することは困難が予想される．旋回カメラによる異常検知が可能となれば，固定カメラより広範囲な監視を行うことができたり，決まったルートを巡回する警備ロボットへの応用が行えるなど，より多種類の場面で適用することが可能となりさらなる有用性が期待できる．

そこで本章では，固定カメラと水平方向に等速旋回する旋回カメラから撮影された 2 種類の監視

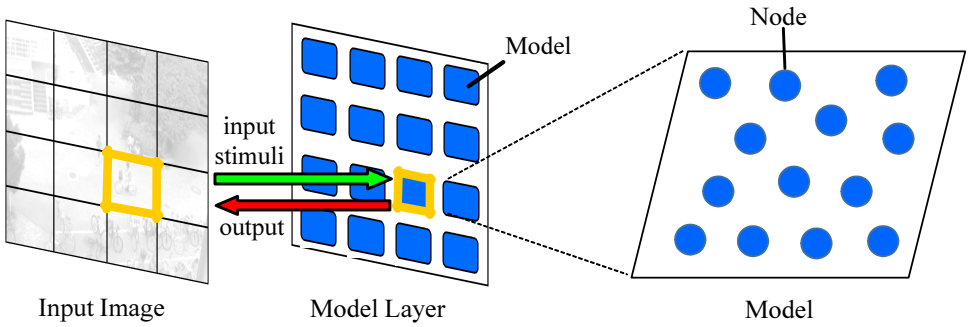


図 5.1: 提案手法の構造

映像を対象に異常検知を行う，より環境への適応性が高い自己組織化モデルを提案する．提案するモデルは，環境からの入力刺激に応じてノードの生成，削除，またノード状態の更新を行うことで環境の正常性を表現する．実験では固定カメラと水平方向に等速旋回する旋回カメラから撮影された 2 種類の監視映像を用いて，映像内に現れる歩行者と車両を検出対象とした侵入物体検知問題に提案手法を適用し，先行研究と比較することで性能の検証を行う．

5.2 自己組織化モデルによる異常検知

5.2.1 概要

提案する異常検知モデルの構造を図 5.1 に示す．提案手法では，同一構造の自己組織化モデルが入力画像中に格子状に整列している．本論文では $m \times n$ 画素に対して 1 つのモデルを配置しており，例えば入力画像サイズが $M \times N$ 画素の場合，入力画像中には $L = \frac{M}{m} \times \frac{N}{n}$ のモデルが配置される．

各モデルは複数のノードによって構成される．ノード i は，重みベクトル $\boldsymbol{v}_i$，時刻 t における馴化係数 $h_i(t)$，年齢 $age_i(t)$ をもち，重みベクトルは環境からの入力刺激と同次元数で各要素が $[0.0, 1.0]$ の実数値，馴化係数はノードがもつ重みベクトルと類似した入力刺激の出現頻度を表す $[0.0, 1.0]$ の実数値，年齢はノードの年齢を表す非負の整数値である．

提案手法では，入力画像中の各格子領域から算出した画像統計量を入力刺激として対応する各モデルに入力する．そして，モデル内でその入力刺激と類似した重みベクトルをもつノードの選択を行い，選択されたノードの年齢と馴化係数の更新を行う．さらに選択されたノードの入力刺激との類似度と馴化係数に応じてモデル内へのノード追加や，ノードの年齢に応じてノード削除を行うことで環境の正常性を表現する構造を構築していく．そして，ノードの類似度を用いて環境内の刺激に対する応答値をモデルごとに非負の実数値で出力することで異常検知を行う．

5.2.2 処理の流れ

図 5.2 に提案手法の処理の流れを示す．処理の詳細については次の通りである．

1 全てのモデルを次の手順で初期化する．

(a) モデルに I 個のノードを生成する．

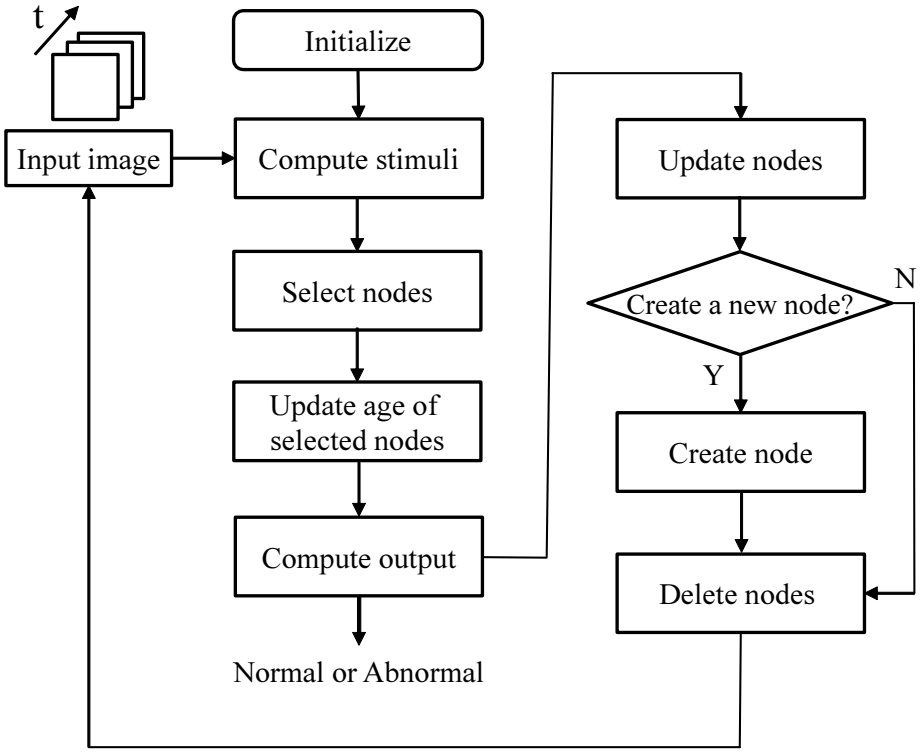


図 5.2: 提案手法の処理の流れ

（b）各ノードの重みベクトルをあらかじめ算出した入力刺激からランダムに選択する．

（c）各ノードの馴化係数 $h_i(0)$ を 1.0 にする．

（d）各ノードの年齢 $age_i(0)$ を 0 にする．

2 全てのモデルについて，入力動画像の 1 フレーム毎に以下の処理（a）から（h）を繰り返す．

（a）モデルが配置された格子領域から入力刺激を算出し，モデルに入力する．入力刺激は格子領域内の画素値から算出した平均，中央値などの画像統計量を用いる．全ての入力刺激は $[0.0, 1.0]$ に正規化される．

（b）モデル内のすべてのノードについて，図 5.3 に示すように，入力刺激との類似度が大きい上位 S 個のノードを選択する．ノード i における類似度 D_i の算出には式（5.1）を用いる．

$$D_i = \exp(-\|\boldsymbol{x} - \boldsymbol{v}_i\|) \tag{5.1}$$

ここで，$\boldsymbol{x}$ は入力刺激，$\boldsymbol{v}_i$ はノード i がもつ重みベクトル，$\|.\|$ はノルムである．式（5.1）は，類似度 D_i が大きいノードほど入力刺激と類似した重みベクトルをもっていることを示す．

（c）選択された S 個のノードの年齢 $age_i(t)$ を 0 にする．式（5.2）における i は選択された S 個のノードのインデックスを表す．

$$age_i(t) = 0 \tag{5.2}$$

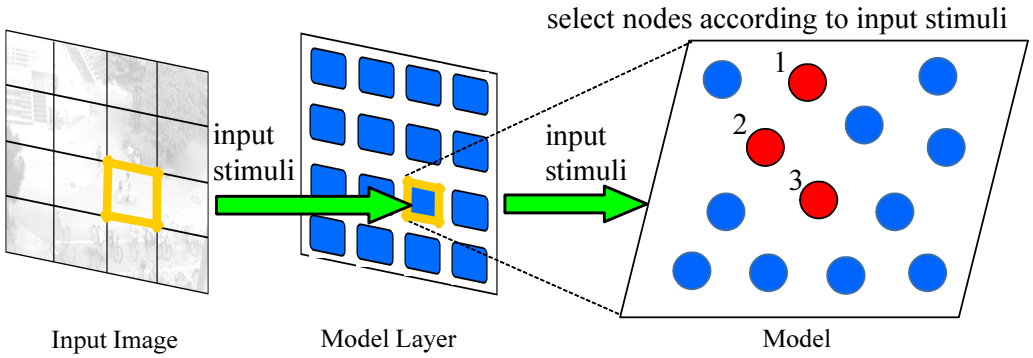


図 5.3: ノード選択の例（$S = 3$ の場合）．ノードの左上の数字は入力刺激に対する類似度の大きさの順位である．

（d）各格子領域に対するモデル l の出力値 O^l は選択された S 個のノードの類似度を用いて，式（5.3）によって算出される．

$$O^l = \begin{cases} D_{c(1)} & (S = 1) \\ \sum_{i=1}^{S-1} D_{c(i)} D_{c(i+1)} & (otherwise) \end{cases} \tag{5.3}$$

$c(i)$ は現フレームで選択された上位 S 個のノードのインデックスのうち，i 番目に類似度が大きいノードのインデックスである．本論文では映像内における出現頻度が低い刺激や特徴が大きく異なる刺激を異常として定義しているため，出力値 O^l がしきい値 O_{thr} より小さい場合，その格子領域を異常であると判定する．

（e）モデル内で最大の類似度 $D_{c(1)}$ であるノードの馴化係数 $h_{c(1)}(t)$ を式（5.4）によって更新する．

$$h_{c(1)}(t+1) = h_{c(1)}(t) - \gamma D_{c(1)} \tag{5.4}$$

馴化係数の更新式（5.4）には入力刺激との類似度を用いており，環境からの入力刺激と類似した重みベクトルをもつノードの馴化係数の値は減少していく．なお，馴化係数の値が 0 を下回った場合は，馴化係数の値は 0 とする．

（f）現フレームで選択された上位 S 個以外のノードの age_i を式（5.5）によって更新する．

$$age_i(t+1) = age_i(t) + 1 \tag{5.5}$$

（g）式（5.6）を満たす場合，新たなノード k をモデルに追加する．

$$\begin{cases} D_{c(1)} < D_{\text{thr}} \\ h_{c(1)} < H_{\text{thr}} \end{cases} \tag{5.6}$$

このとき，ノード k の重みベクトルは入力刺激 $\boldsymbol{x}$ とノード $c(1)$ の重みベクトル $\boldsymbol{v}_{c(1)}$ との平均値とし，馴化係数の初期値 $h_k(0)$ は 1.0，年齢の初期値 $age_k(0)$ は 0 とする．モデルが式（5.6）を満たすことは，ある程度モデルの更新が行われているにも関わらず環境からの入力刺激を表現するノードがモデル内に存在しないことを示している．このことから，環境の刺激をより正確に表現するために，環境の入力刺激を利用して新たなノードをモデルに追加する操作を行っている．

（h）age_i の値がしきい値 age_{thr} より大きい場合，ノード i をモデルから削除する．

以上の提案手法の処理についてまとめたものを Algorithm 1 に示す．

Algorithm 1: Anomaly detection algorithm

Initialize: initialize node set $\{\{N_i^k\}_{i=1}^I\}_{k=1}^L$, weight vector $\boldsymbol{v}_i$ are randomly chosen from input stimuli, habituation coefficient $h_i(0) \leftarrow 1.0$, $age_i(0) \leftarrow 0$.

Input: a set of frames $\{t_i\}_{i=1}^T$, input stimuli $\{\{\boldsymbol{x}_i^k\}_{k=1}^L\}_{i=1}^T$.

Output: anomaly labels for each location of each frame $\{\{\boldsymbol{y}_i^k\}_{k=1}^L\}_{i=1}^T$.

foreach *frame* t_i **do**
  foreach *model* l **do**
    compute input stimuli $\boldsymbol{x}_{t_i}^l$;
    foreach *node* i^l **do**
      compute similarity D_i^l using (5.1);
    end
    select the top S nodes in D_i^l and update age of them using (5.2);
    compute normal value O^l using (5.3);
    if $O^l < O_{\text{thr}}$ **then**
      $\{\boldsymbol{y}_{t_i}^l\} \leftarrow$ *anomaly*;
    else
      $\{\boldsymbol{y}_{t_i}^l\} \leftarrow$ *normal*;
    end
    update habituation coefficient $h_{c(1)}^l(t_i)$ using (5.4) and age of nodes using (5.5);
    if $D_{c(1)}^l < D_{\text{thr}}$ *and* $h_{c(1)}^l < H_{\text{thr}}$ **then**
      insert a new node k with weight vector $\boldsymbol{v}_k^l$: $\boldsymbol{v}_k^l = \frac{1}{2}(\boldsymbol{v}_{c(1)}^l + \boldsymbol{x}_{t_i}^l)$;
    end
    foreach *node* i^l **do**
      if $age_i^l(t_i) > age_{\text{thr}}$ **then**
        delete node i^l;
      end
    end
  end
end

5.3 固定カメラと旋回カメラによる監視映像からの侵入物体検知実験

5.3.1 概要

提案する自己組織化モデルの有効性を検証するため，屋外監視映像内に現れる歩行者と車両を検出対象とした侵入物体検知問題に提案手法を適用し，先行研究との比較を行う．対象とした屋外監視映像は，固定カメラから撮影された監視映像と水平方向に等速旋回する旋回カメラから撮影された 2 種類の監視映像である．比較手法は，文献 [67] で提案された GWR ネットワーク（以下，GWR）と，Stauffer らの手法 [76] に対して直近の入力フレームを重視した更新を行う混合ガウス分布を用いた背景モデルの手法 [79]（以下，MOG）を用いた．GWR は [67] の筆者の Web ページ[1]をもとに筆者らが作成したものを用い，MOG は OpenCV に実装されている関数[2]を使用した．

各実験に共通して用いた提案手法と比較手法のパラメータをそれぞれ表 5.1，表 5.2 に示す．これらのパラメータは次節 5.3.2 で説明する F 値が最大となるように，事前に行った実験によって決定した．各パラメータの詳細は文献 [67, 79] を参照されたい．提案手法と GWR の入力刺激には，格子領域内の画素値の平均，最大値，最小値，レンジ，中央値，第一四分位数，第三四分位数の 7 種類の統計量を RGB カラー画像，RGB エッジ画像のそれぞれから算出した計 42 次元の特徴量を用いた．レンジは格子領域内の最大画素値と最小値画素値の差分値である．本論文では入力刺激が 42 次元ベクトルであるため，提案モデルの各ノードは 42 次元空間の 1 点を表す．

5.3.2 評価方法

まず，異常検知性能の定量評価を行うために，監視映像内に現れる歩行者と車両を対象に正解画像を作成した．正解画像の例を図 5.4（b），図 5.4（e）に示す．さらに，提案手法と GWR では画像内の格子領域単位の出力であるため，作成した正解画像から正解格子画像を作成した．正解格子画像は，正解画像を提案手法および GWR と同様のサイズの格子領域に分割し，格子領域内に占める正解画素の割合が 0.1 以上の格子領域を正解格子領域とすることで作成した．1 つの格子領域サイズが 10×10 画素とした場合の正解格子画像の例を図 5.4（c），図 5.4（f）に示す．また，画素単位で出力する MOG と評価方法を揃えるために，MOG による検出画像を同じサイズの格子領域で分割し，格子領域内に占める検出画素の割合が同じく 0.1 以上の格子領域を MOG による検出格子領域と定義する．

本論文では作成した正解格子画像をもとに，以下の式で示される再現率（R），適合率（P），F 値（F）の指標を用いて定量評価を行う．なお，過剰な過検出結果をノイズとして除外するために，物体の侵入によって生じた過検出格子領域は検出格子領域から除外した．

$$R = \frac{C}{C_{\mathrm{a}}} \tag{5.7}$$

$$P = \frac{C}{A} \tag{5.8}$$

$$F = \frac{2 \cdot R \cdot P}{R + P} \tag{5.9}$$

C は正しく検出した格子領域数，C_a は正解格子領域数，A は検出した格子領域数である．

[1] https://seat.massey.ac.nz/personal/s.r.marsland/gwr.html

[2] http://docs.opencv.org/2.4.9/modules/video/doc/motion_analysis_and_object_tracking.html

表 5.1: 提案手法に関するパラメータ

パラメータ	固定	旋回
初期化時のノード数 I	20	
ノードの選択数 S	1	
馴化係数の更新係数 γ	0.1	
ノード追加の馴化係数しきい値 H_{thr}	0.1	
最大年齢 age_{thr}	800	
ノード追加の類似度しきい値（モデル適用時）D_{thr}	0.1	
ノード追加の類似度しきい値（モデル構築時）D_{thr}	0.6	0.9
出力値のしきい値 O_{thr}	0.76	0.87

表 5.2: 比較手法に関するパラメータ

比較手法	パラメータ	固定	旋回
GWR	発火係数のしきい値 h_T	0.1	
	発火係数の更新係数 α_b, α_n	1.05	
	重みベクトルの更新係数 ϵ_b	0.05	0.01
	重みベクトルの更新係数 ϵ_n	0.01	0.001
	活性度のしきい値 a_T	0.9	0.8
	発火係数の更新係数 τ_b	0.1	0.4
	発火係数の更新係数 τ_n	0.05	0.4
MOG	分布数	5	
	背景を決定するしきい値	0.6	0.9
	更新に利用するフレーム数	300	200
	分布一致を決定するしきい値	6	30

5.3.3 固定カメラからの監視映像による侵入物体検知

実験設定

固定カメラからの監視映像には，PETS2001 データセット中の DATASET2 TESTING CAMERA1 を使用した．用いた画像は 320×240 画素のカラー画像である．提案手法と GWR では 10×10 画素ごとにに 1 つのモデルを配置したため，画像サイズからモデル数は 768（$= \frac{320}{10} \times \frac{240}{10}$）である．この動画像では手前の樹木が風によって揺れており，また背景にちらつきなどのノイズがみられる．固定カメラ映像の例を図 5.4（a）に示す．

歩行者と車両が現れない開始から 300 フレームまでを提案手法と比較手法のモデル構築期間とし，301 フレームから 1800 フレームまでをモデル適用期間とする．なお，正解格子画像は 301 フレームから 1800 フレームまで 10 フレームおきに作成し，計 150 フレーム分を用意した．また，提案手法のノード追加の類似度しきい値 D_{thr} は，表 5.1 に示すようにモデル構築時と適用時では異なる値を用いた．

(a) 入力画像　(b) 正解画像　(c) 正解格子画像

(d) 入力画像　(e) 正解画像　(f) 正解格子画像

図 5.4: 入力画像と正解画像例．上段は固定カメラから撮影された映像例．下段は旋回カメラから撮影された映像例．

表 5.3: 検出結果に対する定量評価（固定カメラ映像）

	Proposal	GWR［67］	MOG［79］
再現率	0.838	0.656	0.897
適合率	0.868	0.758	0.874
F 値	0.853	0.703	0.885

実験結果

図 5.5 に各手法による検出結果例を示す．なお，図 5.5 の各画像は原画像を一部拡大した画像を示している．また，MOG は格子領域単位に変換した結果を示している．図 5.5 の上図の結果から，提案手法では手前の揺れている樹木領域への過検出を抑制しつつ，歩行者を検出できていることがわかる．MOG では樹木領域で生じる多様な入力パターンを適切にモデル化できなかったため，樹木領域に対して過検出してしまっている．GWR は提案手法と同様に樹木領域に対して過検出を抑制できているが，図 5.5（i）に示すように，歩行者に対して検出漏れをしてしまう場合が多くみられた．

再現率，適合率，F 値による定量評価結果を表 5.3 に示す．なお，正解格子領域数は正解画像 150 フレーム分に対して 1732 である．表 5.3 から提案手法は GWR と比べて再現率，適合率において優れており，MOG と比べると適合率で同等程度の性能を示していることがわかる．しかし，再現率において提案手法は MOG より低い値となっている．これは図 5.5（f）内の赤枠で示すように，背景と類似した色情報をもつ歩行者領域に対して検出漏れをしてしまっていることが原因だと考え

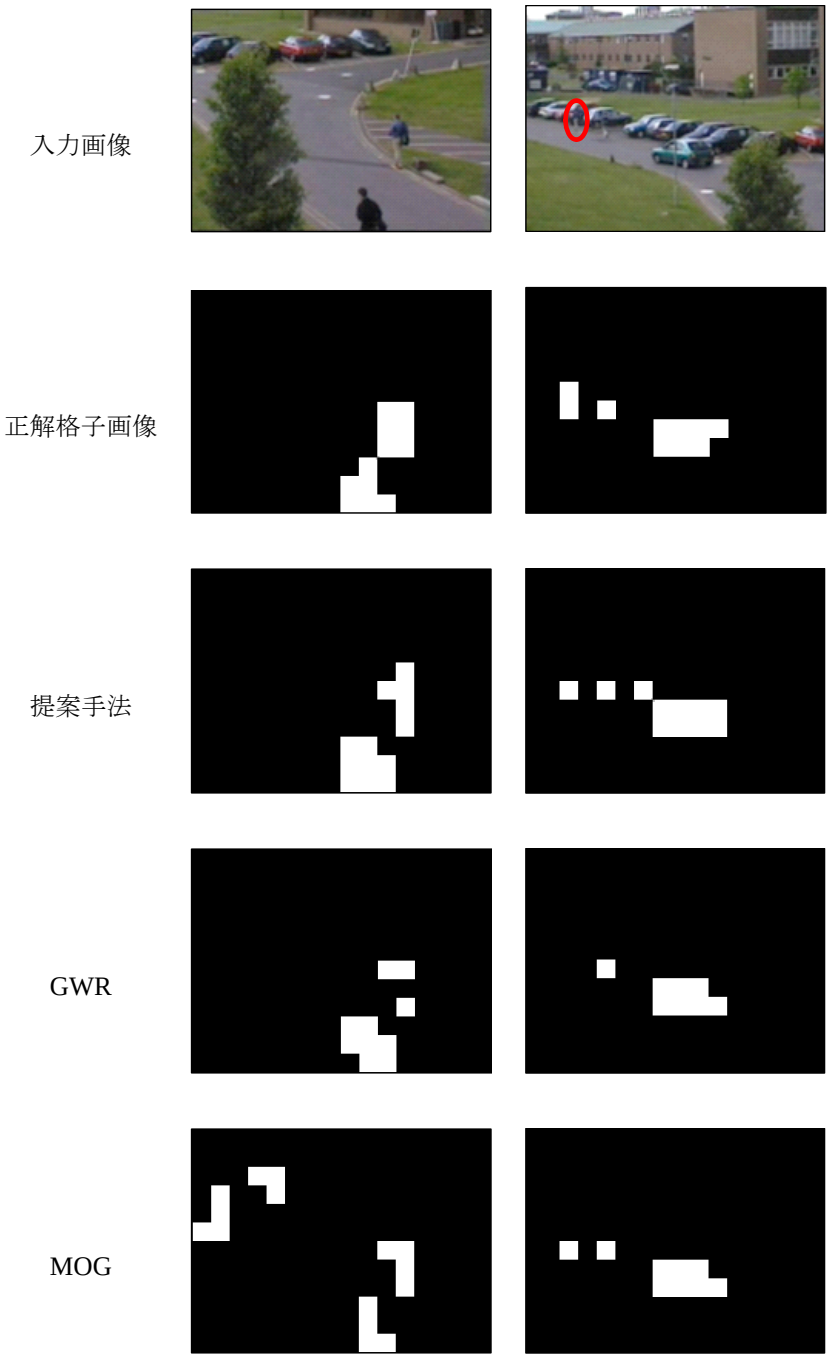


図 5.5: 各手法の検出結果例（固定カメラ映像）

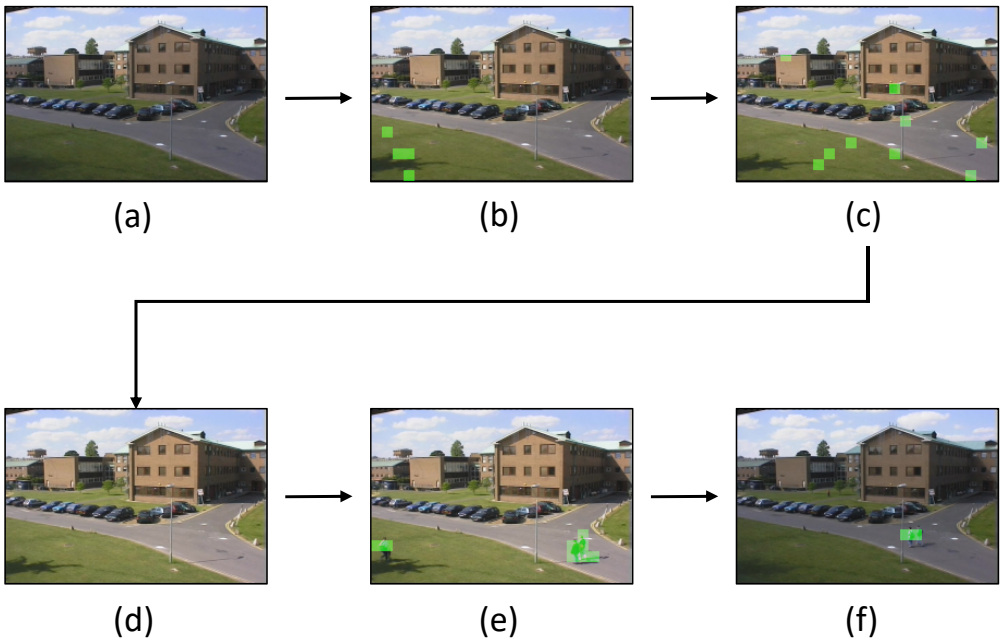


図 5.6: 照明変化に対する結果の例

られる．提案手法では映像内の色情報をもとに映像内の正常性を記述するため，背景の色情報と類似している侵入物体に対しては検出が抑制されてしまう．同様の理由で GWR の再現率の値も低い結果となっている．これは今後，入力刺激に時空間特徴を使用し，入力刺激の時間的な変化も考慮することで解決が可能であると考える．

照明変化のある環境における実験結果

提案手法を照明変化のある環境に適用した際の結果例を図 5.6 に示す．この例では，時間が経過するとともに図 5.6(a) から (f) に示すように，天候による照明変化が生じている．図 5.6(a) では，この状態が正常シーンであると提案手法では定義されている．そのため，照明変化が生じ始めた図 5.6(b) や (c) において，輝度値の明るい領域や樹木の影に対して反応を示していることがわかる．これは図 5.6(a) の環境ではみられなかった特徴が照明変化によって生じたためである．しかし，照明変化が生じた状態が続くと，今度はその状態（図 5.6(c)）が正常シーンとなるように提案手法の更新が行われるため，図 5.6(d) に示すように照明変化に対して反応を示さなくなる．このように提案手法では映像を観測しながら，環境からの入力との類似度を考慮して環境の正常性を表現するため，照明変化にも対応可能な手法となっている．

また，図 5.6(e) や (f) に示すように，照明変化が生じる環境においても侵入物体に対して適切に反応を示せていることがわかる．このことから，本手法は照明変化の生じる環境においても侵入物体検知が行えることが確認できた．

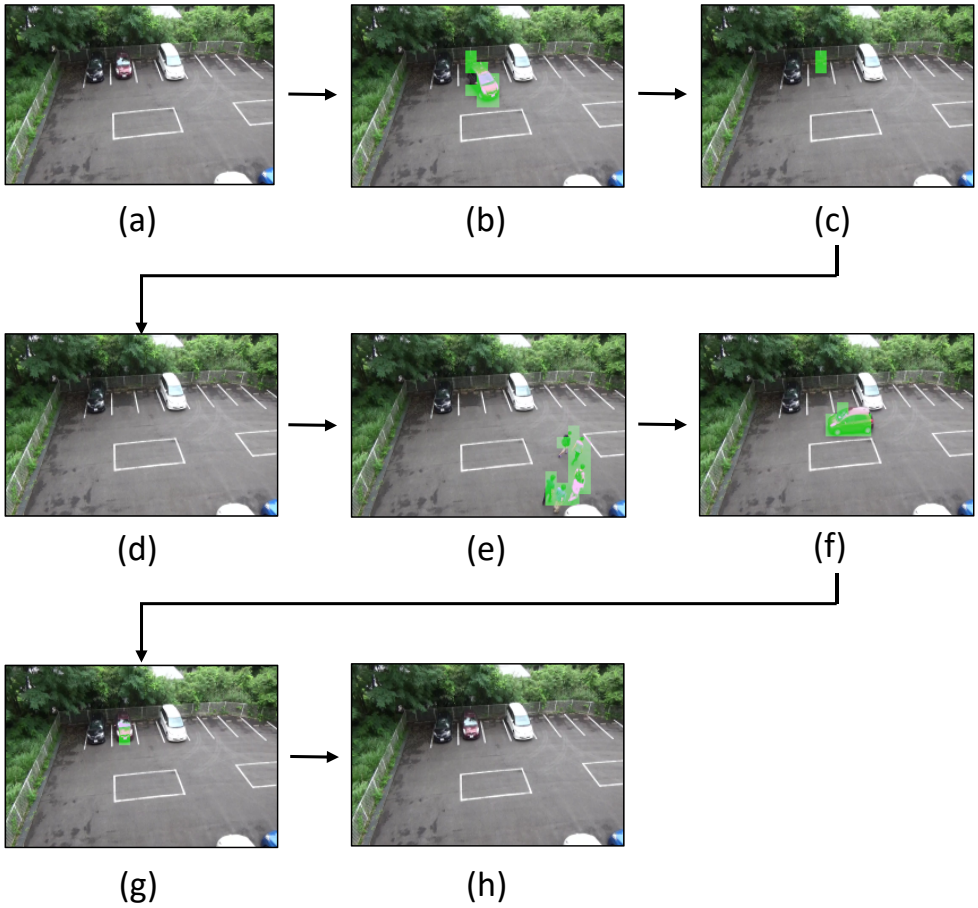


図 5.7: 自動車の出入りに対する結果の例

自動車の出入りが生じる環境における実験結果

提案手法を自動車の出入りが生じる環境に適用した際の結果例を図 5.7 に示す．この例では，時間が経過するとともに図 5.7(a) から (h) に示すように，自動車による出入りが生じている．図 5.7(a) では，この状態（中央奥に自動車が 3 台存在する状態）が正常シーンであると提案手法では判断されている．そのため，中央の赤色自動車の出車が行われている図 5.7(b) において，赤色自動車の領域に対して提案手法が反応を示していることがわかる．また，図 5.7(c) に示すように，今まで自動車が存在していた領域に対しても同様に提案手法が反応を示していることがわかる．これは，図 5.7(a) では存在していた赤色の自動車が，図 5.7(c) では無いため，正常状態と異なるためである．しかし，赤色自動車が存在しない状態が続くと，今度はその状態が正常シーンとなるように提案手法の更新が行われるため，図 5.7(d) に示すように次第に反応を示さなくなる．また，図 5.7(f) や (g) に示すように，再び自動車が登場した場合は自動車領域に反応を示す．これは現在の正常状態が図 5.7(d) のようなシーンに基づいているからである．これまでと同様に，赤色の自動車が戻ってきた状態（5.7(g)）が続くと，提案手法は変化が生じた領域に対して次第に反応を示さなくなる（図 5.7(h)）．

また，図 5.7(e) に示すように，自動車の出入りが生じる環境においても侵入物体に対して適切に反応を示せていることがわかる．このことから，本手法は自動車の出入りなどのような，映像内の

物体の有無が変化する環境においても良好に異常検知が行えることが確認できた．

5.3.4 旋回カメラからの監視映像による侵入物体検知

実験設定

本論文では，水平方向に 45° の範囲で等速旋回する旋回カメラから撮影された監視映像を対象に侵入物体検知を行う．11 ~ 13 秒で監視範囲をカバー可能なように，背景画像が 1 フレームで約 0.4 ~ 0.6 画素の速度で動く旋回カメラとなっている．動画像のフレームレートは 30[fps] である．図 5.4（d）に旋回カメラ映像の例を示す．対象とした環境では映像内の左右上下から歩行者が映像内に出現する．動画像は 240×160 画素のカラー画像を使用し，5.3.3 の実験と同様に提案手法と GWR では 10×10 画素に 1 つのモデルを配置したため，画像サイズからモデル数は 384 である．

歩行者が出現しない開始から 3000 フレームまでを各手法のモデル構築期間とし，歩行者が出現する 3001 フレームから 5000 フレームまでをモデル適用期間とする．正解格子画像は 3001 フレームから 5000 フレームまで 10 フレームおきに作成し，計 200 フレーム分を用意した．提案手法のパラメータ D_{thr} は，5.3.3 の実験と同様にモデル構築時と適用時ではそれぞれ表 5.1 に示す値を用いた．

旋回カメラ映像への対処法として，旋回による移動量をあらかじめ算出しておき，その移動量を用いて入力画像列を基準となるフレームの画像に補正し，固定カメラを想定した手法を適用する方法が考えられる．本章では，固定カメラ映像での異常検知精度が高い MOG をこの補正画像列に適用し，比較手法として用いる．まず，映像内の白線などのエッジ強度が高い特徴点に対してフレーム間で画像の対応付けを行うことで，それら特徴点の 1 フレームにおける移動量を算出する．そして，これら移動量の平均値を旋回による 1 フレームでの移動量とする．次に，事前に定めた基準となるフレームから現在のフレームまでの経過時間と先に求めた旋回による移動量を用いて，現在の画像を移動させることで基準となるフレームの画像に変換を行う．全ての入力画像に対してこの補正処理を行うことで，旋回カメラによる画像列から擬似的な固定カメラ画像列を生成する．そして，生成された擬似的な固定カメラ画像列に対して MOG を適用することで評価を行う．なお，旋回による動き補正を行った際に，補正後の画像内に画像情報が欠落してしまう領域が存在するが，情報が欠落している領域に対しては MOG の更新および異常検知判定は行わないようにした．補正処理を行う MOG のパラメータは分布数を 5，背景を決定するしきい値を 0.9，更新に利用するフレーム数を 200，分布一致を決定するしきい値を 27.5 とした．

表 5.4: 検出結果に対する定量評価（旋回カメラ映像）

	Proposal	GWR［67］	MOG［79］	MOG(correction)
再現率	0.753	0.478	0.340	0.725
適合率	0.845	0.831	0.081	0.512
F 値	0.797	0.607	0.131	0.600

実験結果

再現率，適合率，F 値による定量評価結果を表 5.4 に示す．補正処理を行った MOG の結果は MOG(correction) として示している．また，正解格子領域数は正解画像 200 フレーム分に対して 247 である．表 5.4 から提案手法は再現率，適合率，F 値において他の 3 手法と比べて高い値を示していることがわかる．F 値は比較手法と比べて約 19%以上優れている．補正処理を行う前の MOG では，変動する背景領域の多様な状態を適切にモデル化できなかったため，再現率，適合率ともに低い数値となっているが，補正処理を行うことで再現率，適合率ともに大幅な改善がみられた．

図 5.8 に各手法における検出結果例を示す．図 5.8 の各画像は原画像を一部拡大した画像を示している．なお，図 5.8 には補正処理を行った MOG の結果を示している．図 5.8 の結果から，提案手法では変動する背景領域への過検出を抑制しつつ，侵入物体である歩行者を検出できていることがわかる．図 5.8（h）の歩行者領域の一部に対して検出漏れがみられるが，前後の評価対象外のフレームをみると，この検出漏れの領域に対して正しく検出できていることが確認されている．GWR は提案手法と同様に変動する背景領域への過検出を抑制できているが，歩行者領域に対して検出漏れをしてしまう場合が多くみられた．補正処理を行った MOG では歩行者領域に対して良好に検出できているが，変動する背景領域に対して過検出する場合がみられた．これは旋回装置の旋回精度や振動，補正時の補正誤差が原因で適切に背景領域のモデル化が行えなかったためであると考えられる．

5.3.5 考察

各手法で異常検知のしきい値を変えたときの異常検知率と誤検知率を図 5.9 の ROC 曲線に示す．図 5.9（a）は固定カメラ映像に対する結果，図 5.9（b）は旋回カメラ映像に対する結果を示している．なお，図 5.9（b）には補正処理を行った MOG の結果を示している．提案手法では出力値のしきい値，GWR では出力値のしきい値と活性度のしきい値，MOG では背景を決定するしきい値と分布一致を決定するしきい値を変えて異常検知率と誤検知率を求めた．図 5.9 は横軸に誤検知率，縦軸に異常検知率をとり，グラフが左上にあるほど性能が優れていることを表している．

図 5.9（a）の固定カメラ映像の結果から，提案手法と MOG が GWR と比べて優れた結果を示しており，MOG が提案手法より良好な結果となっている．しかし，提案手法では図 5.5 の上図に示したように風によって揺れている樹木領域への過検出を抑制しつつ，侵入物体を検知できている．このことを定量的に評価するため，図 5.10 内の赤枠で示した樹木領域における過検出数を 301 フレームから 1800 フレームの 1500 フレーム分について調べた．その結果，過検出数が提案手法では 179，GWR では 301，MOG では 402 となった．このことから，提案手法は他の 2 手法と比べて変動が生じる領域への過検出を抑制できていることがわかる．

次に，図 5.9（b）の旋回カメラ映像の結果から，提案手法は比較手法と比べて優れた結果を示していることがわかる．固定カメラ映像における樹木領域への過検出の抑制結果と旋回カメラ映像における評価結果から，提案手法は比較手法と比べてより環境の変化に頑健に異常検知を行うことができることが確認された．

このように提案手法が環境の変化に対して頑健に異常検知を行えるのは，環境からの入力刺激との類似度を考慮して環境の正常性を表現している点にある．まず，提案手法では入力刺激と類似しているノードがモデル内に存在しない場合，入力刺激と類似したノードをモデル内に追加することで環境の変化に適応する．例えば，図 5.11 内の赤枠で示した位置のモデルでは，旋回カメラの旋回によって移動した図 5.11 右のフレームで白線を観測したときにノード追加が行われる．これは

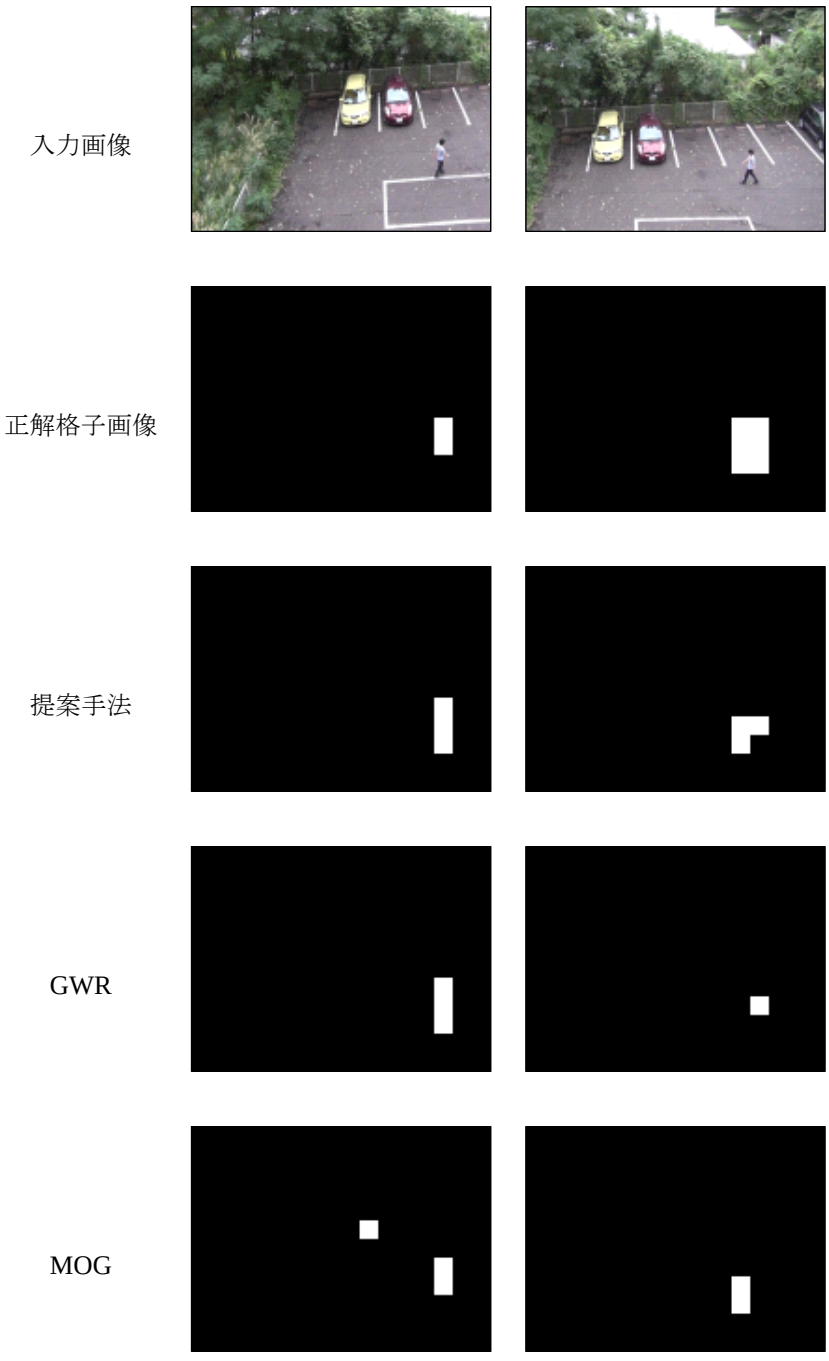
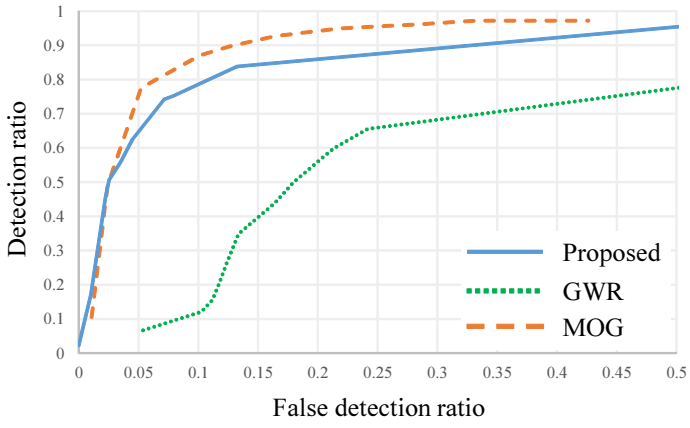
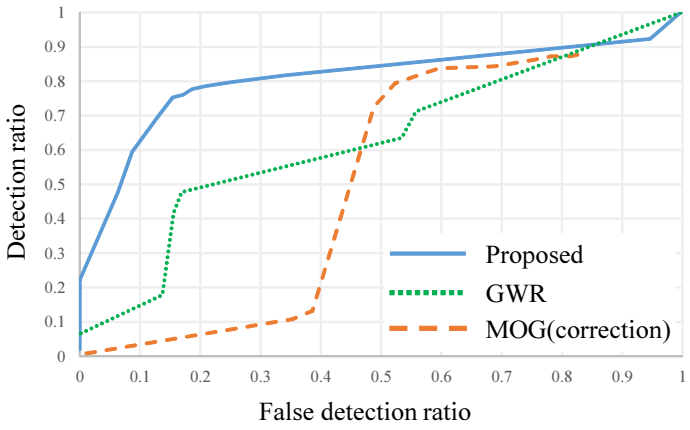


図 5.8: 各手法の検出結果例（旋回カメラ映像）



(a) 固定カメラ映像



(b) 旋回カメラ映像

図 5.9: 各手法の ROC 曲線

現在までのフレームで白線を表現するノードがモデル内に存在せず，式（5.6）を満たしたためである．また，ノードの削除も行われており，環境内で出現頻度が低い入力刺激を表現するノードはモデルから削除される．固定カメラ映像における樹木領域や，旋回カメラ映像における背景領域のように様々な入力刺激が観測される環境では，このノード追加や削除の操作によって環境の状態を表現するモデルを構築していく．次に，提案手法ではこの構築されたモデル内のノードと入力刺激との類似度を用いて出力値を算出するため，環境内で特徴の異なる入力刺激に対して出力値の正常度を抑制することができる．GWR では入力刺激との類似度ではなく，出現頻度を考慮して異常判定を行うため，モデルの入力刺激に対する表現力が弱いといえる．また MOG については，変動領域で観測される様々な入力パターンを背景として適切にモデル化できなかったことが過検出結果の原因だと考えられる．

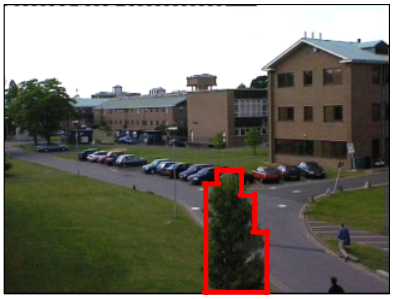


図 5.10: 樹木領域

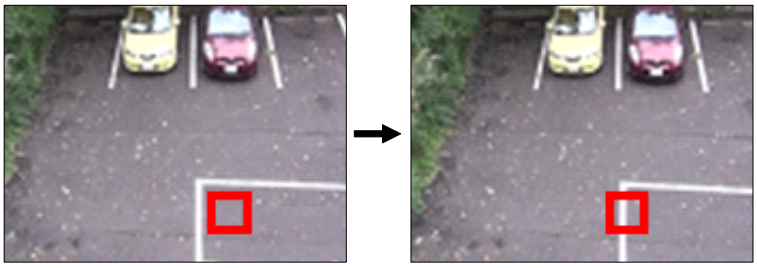


図 5.11: ノード追加シーンの例

5.3.6　パラメータによる影響

提案手法の各パラメータが出力に及ぼす影響について考察する．ここでは，特に出力に影響を及ぼすと考えられる，ノード選択数 S，モデル構築時のノード追加の類似度しきい値 D_{thr}，モデル適用時のノード追加の類似度しきい値 D_{thr} の 3 つのパラメータについて検討を行う．出力値のしきい値 O_{thr} については，図 5.9 に示した通りである．これらの 3 つのパラメータを一つずつ独立に変化させながら，5.3.4 と同様の評価実験を行った結果を表 5.5，表 5.6，表 5.7 に示す．なお，変更していないパラメータについては表 5.1 と同様の値である．

まず選択ノード数 S について，提案手法では選択したノード数分の類似度を使用して出力値を算出することができる．表 5.5 の結果から選択するノード数 S の数は少ない方がより優れた性能を示していることがわかる．旋回カメラ映像では背景が大きく変動するため，それぞれ異なる重みベクトルの値をもったノードが多く存在する．そのため選択数 S を増加させると出力値が抑制されてしまい，結果として再現率，適合率が下がる結果となった．

次に，モデル構築時のノード追加の類似度しきい値 D_{thr} について，表 5.6 の結果から D_{thr} の値が高くなるとより優れた性能を示すことがわかる．これはモデル構築時に D_{thr} の値を低くしてノード追加の制限を厳しくすると，変動する背景領域を表現するノードが十分にモデルに追加されず，モデル適用時に背景領域に対して過検出を多く引き起こしてしまうためである．そのため結果として，D_{thr} の値が低いと適合率が大幅に減少してしまったと考えられる．反対にしきい値 D_{thr} を高くすると，環境内の背景領域を表現するノードが追加されやすくなるため，過検出が抑制され適合率が上昇する結果となった．

最後に，モデル適用時のノード追加の類似度しきい値 D_{thr} に対する結果を表 5.7 に示す．類似度しきい値は値が高くなると前述したように，環境内の刺激を表現するノードがモデルに追加されや

すくなる．そのため，適用時の類似度しきい値を高く設定すると，背景領域に加えて歩行者領域に対する検出も抑制されるようになる．表 5.7 から，しきい値が高くなると歩行者領域に対する再現率が低下していることがわかる．また，しきい値 D_{thr} の値が 0.1，0.3，0.5，0.7，0.8 のときの背景領域に対する過検出数はそれぞれ 34，34，34，30，21 であり，しきい値が高くなると過検出数が抑制されることが確認された．したがって，適用時の類似度しきい値 D_{thr} を高く設定すると，再現率は低下するが，過検出数は抑制される結果となった．反対にモデル適用時の D_{thr} を低く設定すると，歩行者領域を表現するノードがモデルに追加されにくくなり，結果として再現率が上昇する結果を示した．

表 5.5: ノード選択数 S に対する提案手法の性能の変化

S	1	2	3	4	5
再現率	0.753	0.680	0.628	0.607	0.425
適合率	0.845	0.824	0.752	0.630	0.587
F 値	0.797	0.745	0.684	0.619	0.493

表 5.6: モデル構築時の類似度しきい値 D_{thr} に対する提案手法の性能の変化

D_{thr}	0.825	0.850	0.875	0.900	0.925	0.950
再現率	0.773	0.765	0.761	0.753	0.709	0.700
適合率	0.136	0.376	0.793	0.845	0.862	0.869
F 値	0.231	0.505	0.777	0.797	0.778	0.776

表 5.7: モデル適用時の類似度しきい値 D_{thr} に対する提案手法の性能の変化

D_{thr}	0.1	0.3	0.5	0.7	0.8
再現率	0.753	0.749	0.737	0.518	0.340
適合率	0.845	0.845	0.843	0.810	0.800
F 値	0.797	0.794	0.786	0.632	0.477

5.4 まとめ

本章では，検出対象である異常（負例）と正常（正例）のラベルデータを用いないイベント検出として，監視映像からの異常検知を行った．本研究では，環境から入力される刺激にもとづいて環境の正常性を表現し，異常検知を行う自己組織化モデルを提案した．このモデルでは，環境からの入力刺激との類似度に応じてノードの追加，削除，状態の更新，また出力値を算出することで環境の正常性を表現する．提案手法を固定カメラと旋回カメラから撮影された 2 種類の監視映像からの侵入物体検知問題に適用し，先行研究と比較を行った結果，風による樹木の揺れや，旋回による変動がある領域への過検出を抑制しつつ，異常である侵入物体を検知できることが確認された．

今後の課題として，長時間の監視映像への適用実験と，その実験を通じての精度向上およびパラメータ設定に関する明確な基準を求めることが挙げられる．また，本手法では画像内の領域ごとに異なる正常性を表現しているが，近傍の領域や画像全体で正常性を共有することで，より効率的な正常性の表現を行う必要があると考えている．さらに，提案手法の拡張として人物の異常動作検知や，混雑シーンにおける適用などを検討したい．

第6章　結論

6.1　本論文で得られた成果および課題

本論文では，動画像から顕著性のあるイベントを自動検出する手法を提案し，先行研究と比較することでその有効性を検証した．顕著性のあるイベント（負例）と顕著性のないイベント（正例）のラベルデータの有無による各条件において，各章でイベント検出手法を提案した．各章で得られた成果は以下の通りである．

- **覚醒下脳腫瘍摘出術における電気刺激位置（イベント）の自動検出**
  正例および負例のラベルデータが使用可能な例として，覚醒下脳腫瘍摘出術における電気刺激位置の自動検出する手法を提案し，その有効性を検証した．提案手法は電極先端位置の検出と電気刺激終了タイミング検出の 2 段階から構成される．電極先端位置の検出では，電極全体の形状と色特徴に着目した検出，電極先端の形状と色特徴に着目した検出，電極先端位置の追跡，の 3 手法による検出結果を統合することで高精度な電極先端位置検出を実現した．また電気刺激終了タイミング検出では，検出した電極先端周辺のオプティカルフローの分布を特徴量とし，SVM へ入力することで電気刺激終了タイミングの検出を行った．これら電極先端位置検出と電気刺激終了タイミング検出を組み合わせることで電気刺激位置の検出を行い，F 値 0.6806 の識別率を示した．

  今後の課題として，より高精度な電気刺激終了タイミングの検出手法の提案が挙げられる．特に，本論文における電気刺激終了タイミングの検出では，電極先端周辺のオプティカルフローしか特徴量として用いていないため，電極全体のオプティカルフローの分布を用いるなど，他の特徴量の追加の必要があると考えている．さらに今後は，症例数を増やして実験を行うことで，より信頼性のある検証を行いたい．

- **混雑シーンにおけるイベント検出**
  正例のラベルデータのみが使用可能な例として，混雑シーンにおける異常検知を行う手法を提案し，その有効性を検証した．提案手法は Convolutional Autoencoder（CAE）の再構築誤差を利用して，異常検知を行う．先行研究では，異常検知に重要な形状特徴や速度特徴をあらかじめ人手によって与えている手法が多いが，本手法ではこれらの特徴を与えずに異常検知を行う，より柔軟な異常検知手法の提案を行った．UCSD pedestrian dataset（Ped1，Ped2）に提案手法を適用した結果，Ped2 のデータに対してはすべての先行研究よりも優れた性能を示した．このことから，あらかじめ特徴を与えなくても良好に異常検知が行えることが確認できた．

  一方，Ped1 のような縦構造の映像に対しては検出漏れが目立つ結果となり，このような縦構造の映像に対する異常検知性能の向上が今後の課題として挙げられる．また，Ped1 および Ped2 で共通する課題として，速度特徴が類似している異常対象を検出漏れしてしまう点が挙げられる．そのため，今後は形状特徴もより適切に学習できるような枠組みを検討する必要がある．

- **固定および旋回監視映像からのイベント検出**

  正例および負例のラベルデータを使用せずにイベント検出を行うタスクとして，固定および旋回監視映像からの侵入物体検出を扱った．本章では，環境から入力される特徴にもとづいて環境の正常性を表現し，侵入物体検知を行う自己組織化モデルを提案した．提案手法では，環境からの入力特徴との類似度に応じてノードの追加，削除，状態の更新，また出力値を算出することで環境の正常性を表現する．提案手法を固定カメラと旋回カメラから撮影された 2 種類の監視映像からの侵入物体検知問題に適用し，先行研究と比較を行った結果，風による樹木の揺れや，旋回による変動がある領域への過検出を抑制しつつ，異常である侵入物体を検知できることが確認された．

  今後の課題として，長時間の監視映像への適用実験と，その実験を通じての精度向上およびパラメータ設定に関する明確な基準を求めることが挙げられる．また，本手法では画像内の領域ごとに異なる正常性を表現しているが，近傍の領域や画像全体で正常性を共有することで，より効率的な正常性の表現を行う必要があると考えている．さらに，提案手法の拡張として人物の異常動作検知や，混雑シーンにおける適用などを検討したい．

以上の成果から，本研究では正例と負例のラベルデータの有無における各条件において，動画像から顕著性のあるイベントの検出を実現することができた．

謝辞

博士課程後期進学の後押しをして頂き，本研究を進めるにあたり終始多大なるご指導ご助言，豊かな研究環境を賜りました長尾智晴先生に深く感謝致します．また，本論文を執筆するにあたり，ストーリーの方向性に関する貴重なご指導，ご助言を頂きました田村直良先生，森辰則先生，富井尚志先生，白川真一先生に感謝致します．

また，日本学術振興会には特別研究員 DC1 として採用していただき，研究生活へ多くのご支援をいただきました．ありがとうございます．

長尾研究室の皆様には研究をはじめ多くの事柄に関して貴重なご意見を頂きました．特に，先輩方から教わりました文章の書き方や研究に対する考え方は，今でも自分の礎となっております．

最後に，家族や友人をはじめ，温かく見守ってくれた方々に感謝します．ありがとうございました．

参考文献

［1］F. Sakabe, M. Murakawa, T. Kobayashi, T. Higuchi, and T. Otsu. Chapter mark addition based on anomalousness for surgery videos using chlac features. *International Journal of Advanced Science and Technology*, Vol. 13, pp. 1–14, 2009.

［2］D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe. Learning deep representations of appearance and motion for anomalous event detection. *British Machine Vision Conference*, 2015.

［3］M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe. Plug-and-play cnn for crowd motion analysis: An application in abnormal event detection. *arXiv preprint arXiv:1610.00307*, 2016.

［4］武田真人, 矢田紀子, 長尾智晴. 映像監視のための環境に適応的な異常検知ネットワーク. 電子情報通信学会論文誌, Vol. J94-D, No. 10, pp. 1631–1639, 2011.

［5］村垣善浩, 丸山隆志, 伊関洋, 高倉公朋, 堀智勝. 覚醒下機能マッピングとモニタリングを用いた手術. 脳神経外科ジャーナル, Vol. 17, No. 1, pp. 38–47, 2008.

［6］F. Lalys, L. Riffaud, X. Morandi, and P. Jannin. Surgical phases detection from microscope videos by combining svm and hmm. *Medical computer vision. Recognition techniques and applications in medical imaging*, pp. 54–62, 2011.

［7］F. Lalys, L. Riffaud, D. Bouget, and P. Jannin. An application-dependent framework for the recognition of high-level surgical tasks in the or. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 331–338, 2011.

［8］A. James, D. Vieira, B. Lo, A. Darzi, and G. Yang. Eye-gaze driven surgical workflow segmentation. *Medical image computing and computer-assisted intervention (MICCAI)*, pp. 110–117, 2007.

［9］L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286, 1989.

［10］N. Padoy, T. Blum, H. Feussner, M. Berger, and N. Navab. On-line recognition of surgical activity for monitoring in the operating room. In *Proceedings of the national conference of Innovative applications of artificial intelligence*, pp. 1718–1724, 2008.

［11］N. Padoy, T. Blum, S. Ahmadi, H. Feussner, M. Berger, and N. Navab. Statistical modeling and recognition of surgical workflow. *Medical image analysis*, Vol. 16, No. 3, pp. 632–641, 2012.

［12］T. Blum, H. Feußner, and N. Navab. Modeling and segmentation of surgical workflow from laparoscopic video. *Medical image computing and computer-assisted intervention (MICCAI)*, pp. 400–407, 2010.

［13］ B. Lo, A. Darzi, and G. Yang. Episode classification for the analysis of tissue/instrument interaction with multiple visual cues. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 230–237, 2003.

［14］ B. Bhatia, T. Oates, Y. Xiao, and P. Hu. Real-time identification of operating room state from video. In *Proceedings of the national conference on Innovative applications of artificial intelligence*, Vol. 2, pp. 1761–1766, 2007.

［15］ T. Kobayashi and N. Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 4, pp. 741–744, 2004.

［16］ T. Suzuki, K. Yoshimitsu, Y. Sakurai, K. Nambu, Y. Muragaki, H. Iseki. Automatic surgical phase estimation using multiple channel video data for post-operative incident analysis. *Medical Image Computing and Computer-Assisted Intervention (MICCAI) Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI)*, 2011.

［17］ R. Richa, M. Balicki, R. Sznitman, E. Meisner, R. Taylor, and G. Hager. Vision-based proximity detection in retinal surgery. *IEEE Transactions on biomedical engineering*, Vol. 59, No. 8, pp. 2291–2301, 2012.

［18］ S. Speidel, M. Delles, C. Gutt, and R. Dillmann. Tracking of instruments in minimally invasive surgery for surgical skill analysis. In *Medical Imaging and Augmented Reality, Lecture Notes in Computer Science*, pp. 148–155, 2006.

［19］ R. Sznitman, K. Ali, R. Richa, R. Taylor, G. Hager, and P. Fua. Data-driven visual tracking in retinal microsurgery. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 568–575, 2012.

［20］ J. Rosen, J. D. Brown, L. Chang, M. N. Sinanan, and B. Hannaford. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. *IEEE Transactions on Biomedical engineering*, Vol. 53, No. 3, pp. 399–413, 2006.

［21］ H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager. Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery*, Vol. 11, No. 5, pp. 220–230, 2006.

［22］ J. JH. Leong, M. Nicolaou, L. Atallah, G. P. Mylonas, A. W. Darzi, and G-Z. Yang. Hmm assessment of quality of movement trajectory in laparoscopic surgery. *Computer Aided Surgery*, Vol. 12, No. 6, pp. 335–346, 2007.

［23］ J. Rosen, B. Hannaford, C. G. Richards, and M. N. Sinanan. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 5, pp. 579–591, 2001.

［24］ C. E. Reiley and G. D. Hager. Decomposition of robotic surgical tasks: an analysis of subtasks and their correlation to skill. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI) Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI)*, 2009.

［25］A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 3, pp. 555–560, 2008.

［26］A. Basharat, A. Gritai, and M. Shah. Learning object motion patterns for anomaly detection and improved object detection. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

［27］H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 819–826, 2004.

［28］T. Xiang and S. Gong. Video behavior profiling for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 5, pp. 893–908, 2008.

［29］J. Kim and K. Grauman. Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2928, 2009.

［30］南里卓也, 大津展之. 複数人動画像からの異常動作検出. 情報処理学会論文誌, コンピュータビジョンとイメージメディア, Vol. 46, No. 15, pp. 43–50, 2005.

［31］N. Nomoto, Y. Shinohara, T. Shiraki, T. Kobayashi, and N. Otsu. A new scheme for image recognition using higher-order local autocorrelation and factor analysis. In *IAPR Conference on Machine Vision Applications*, pp. 265–268, 2005.

［32］Z. Fu, W. Hu, and T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *Proceedings of the IEEE International Conference on Image Processing*, 2005.

［33］C. Piciarelli, C. Micheloni, and G. L. Foresti. Trajectory-based anomalous event detection. *IEEE Transactions on Circuits and Systems for video Technology*, Vol. 18, No. 11, pp. 1544–1554, 2008.

［34］F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos. Anomalous video event detection using spatiotemporal context. *Computer Vision and Image Understanding*, Vol. 115, No. 3, pp. 323–333, 2011.

［35］S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2054–2060, 2010.

［36］C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 747–757, 2000.

［37］V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1975–1981, 2010.

［38］Y. Benezeth, P. M. Jodoin, V. Saligrama, and C. Rosenberger. Abnormal events detection based on spatio-temporal co-occurences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2458–2465, 2009.

［39］ T. Hospedales, S. Gong, and T. Xiang. Video behaviour mining using a dynamic topic model. *International journal of computer vision*, Vol. 98, No. 3, pp. 303–323, 2012.

［40］ Vikas Reddy, Conrad Sanderson, and Brian C Lovell. Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 55–61, 2011.

［41］ V. Saligrama and Z. Chen. Video anomaly detection based on local statistical aggregates. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2112–2119, 2012.

［42］ B. Antić and B. Ommer. Video parsing for abnormality detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2415–2422, 2011.

［43］ T. Xiao, C. Zhang, and H. Zha. Learning to detect anomalies in surveillance video. *IEEE Signal Processing Letters*, Vol. 22, No. 9, pp. 1477–1481, 2015.

［44］ Y. Cong, J. Yuan, and Y. Tang. Video anomaly search in crowded scenes via spatio-temporal motion context. *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 10, pp. 1590–1599, 2013.

［45］ M. J. Roshtkhari and M. D. Levine. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. *Computer vision and image understanding*, Vol. 117, No. 10, pp. 1436–1452, 2013.

［46］ B. Antić and B. Ommer. Spatio-temporal video parsing for abnormality detection. *arXiv preprint arXiv:1502.06235*, 2015.

［47］ A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 5, pp. 909–926, 2008.

［48］ W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 1, pp. 18–32, 2014.

［49］ D. Gao and N. Vasconcelos. Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics. *Neural computation*, Vol. 21, No. 1, pp. 239–271, 2009.

［50］ R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 935–942, 2009.

［51］ L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1446–1453, 2009.

［52］ Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3449–3456, 2011.

[53] C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2720–2727, 2013.

[54] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette. Fully convolutional neural network for fast anomaly detection in crowded scenes. *arXiv preprint arXiv:1609.00866*, 2016.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612, 2004.

[56] D. Brunet, E. R. Vrscay, and Z. Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, Vol. 21, No. 4, pp. 1488–1499, 2012.

[57] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, Vol. 13, No. 7, pp. 1443–1471, 2001.

[58] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis. Learning temporal regularity in video sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–742, 2016.

[59] N. Dalal, B. Triggs. Histogram of oriented gradients for human detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.

[60] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. *Proceedings of the European Conference on Computer Vision*, pp. 428–441, 2006.

[61] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[62] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette. Fully convolutional neural network for fast anomaly detection in crowded scenes. *arXiv preprint arXiv:1609.00866*, 2016.

[63] Y. Feng, Y. Yuan, and X. Lu. Learning deep event models for crowd anomaly detection. *Neurocomputing*, Vol. 219, pp. 548–556, 2017.

[64] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 3, pp. 555–560, 2008.

[65] J. Kim and K. Grauman. Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2921–2928, 2009.

[66] J. Feng, C. Zhang, and P. Hao. Online learning with self-organizing maps for anomaly detection in crowd scenes. In *Proceedings of International Conference on Pattern Recognition*, pp. 3599–3602, 2010.

[67] S. Marsland, J. Shapiro, and U. Nehmzow. A self-organising network that grows when required. *Neural Networks*, Vol. 15, No. 8, pp. 1041–1058, 2002.

［68］ S. Marsland, U. Nehmzow, and J. Shapiro. Detecting novel features of an environment using habituation. In *Proceedings of Simulation of Adaptive Behavior*, pp. 189–198, 2000.

［69］ S. Marsland, U. Nehmzow, and J. Shapiro. Environment-specific novelty detection. In *From Animals to Animats, the 7th International Conference on Simulation of Adaptive Behaviour*, pp. 36–45, 2002.

［70］ J. C. Stanley. Computer simulation of a model of habituation. *Nature*, Vol. 261, pp. 146–147, 1976.

［71］ D. Wang and M. A. Arbib. Modeling the dishabituation hierarchy: The role of the primordial hippocampus. *Biological Cybernetics*, Vol. 67, No. 6, pp. 535–544, 1992.

［72］ M. A. Arbib. *The handbook of brain theory and neural networks.* The MIT press, 1995.

［73］ HV. Neto and U. Nehmzow. Visual novelty detection for inspection tasks using mobile robots. In *Proceedings of the 8th Brazilian Symposium on Neural Networks*, 2004.

［74］ U. Nehmzow and HV. Neto. Novelty-based visual inspection using mobile robots. In *Proceedings of Towards Autonomous Robotic Systems*, 2004.

［75］ U. Nehmzow and HV. Neto. Visual attention and novelty detection: Experiments with automatic scale selection. In *Proceedings of Towards Autonomous Robotic Systems*, pp. 139–146, 2006.

［76］ C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, 1999.

［77］ W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 22–29, 1998.

［78］ C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, pp. 747–757, 2000.

［79］ P. KadewTraKuPong and R. Bowden. An improved adaptive background mixture model for realtime tracking with shadow detection. In *Proceedings of 2nd European Workshop on Advanced Video-Based Surveillance Systems*, 2001.

［80］ B. Han, D. Comaniciu, and L. Davis. Sequential kernel density approximation through mode propagation: applications to background modeling. In *Proceedings of Asian Conference on Computer Vision*, Vol. 4, pp. 818–823, 2004.

［81］ Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 2, pp. 28–31, 2004.

［82］ Z. Zivkovic and F. Van Der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, Vol. 27, No. 7, pp. 773–780, 2006.

［83］ B. Zhao, L. Fei-Fei, and E. P. Xing. Online detection of unusual events in videos via dynamic sparse coding. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3313–3320, 2011.

［84］ R. E Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, No. 37, pp. 297–336, 1999.

［85］ Y. Freund and R. E. Schapire. A decisiontheoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119–139, 1997.

［86］ D. Comaniciu and P. Meer. Mean shift analysis and applications. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, pp. 1197–1203, 1999.

［87］ D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 5, pp. 564–577, 2003.

［88］ V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning*, pp. 807–814, 2010.

［89］ S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pp. 448–456, 2015.

［90］ D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

［91］ J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, Vol. 12, No. Jul, pp. 2121–2159, 2011.

［92］ Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, Vol. 4, No. 2, pp. 26–31, 2012.

［93］ S. Tokui, K. Oono, S. Hido, and J. Clayton. Chainer: a next-generation open source framework for deep learning. In *Proceedings of Workshop on Machine Learning Systems in the twenty-ninth Annual Conference on Neural Information Processing Systems*, Vol. 5, 2015.

［94］ R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 935–942, 2009.

研究業績リスト

論文誌

1. 小林雅幸，菅沼雅徳，崎津実穂，長尾智晴：進化的条件判断ネットワークにおける画像分類過程の可視化，進化計算学会論文誌, Vol.7, No.3, pp.65-76, 2016
2. 菅沼雅徳，土屋大樹，白川真一，長尾智晴：遺伝的プログラミングを用いた階層的な特徴構築による画像分類，情報処理学会論文誌 数理モデル化と応用 (TOM), Vol.9, No.3, pp.44-53, 2016
3. 工藤理人，菅沼雅徳，長尾智晴 : ユニットの冗長化による耐故障性を考慮した進化型ニューラルネットワーク，情報処理学会論文誌 数理モデル化と応用 (TOM), Vol.9, No.3, pp.54-60, 2016
4. 菅沼雅徳，長尾智晴：異常検知のための自己組織化モデルとその監視映像への適用，情報処理学会論文誌 数理モデル化と応用 (TOM), Vol.9, No.1, pp.23-32, 2016
5. 崎津実穂，菅沼雅徳，土屋大樹，長尾智晴 : 決定木及び決定ネットワークによる画像分類過程の説明文の自動生成，情報処理学会論文誌 数理モデル化と応用 (TOM), Vol.9, No.1, pp.43-52, 2016
6. 菅沼雅徳，長尾智晴，田村学，村垣善浩，伊関洋 : 覚醒下脳腫瘍摘出術における皮質マッピング動画像記録の電気刺激位置の自動検出，Medical Imaging Technology, Vol.32, No.4, pp.272-281, 2014

国際会議発表

1. Masanori Suganuma, Shinichi Shirakawa, Tomoharu Nagao: A genetic programming approach to designing convolutional neural network architectures, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2017), Berlin, Germany, 15-19 July, 2017 (Accepted)
2. Masanori Suganuma, Daiki Tsuchiya, Shinichi Shirakawa, Tomoharu Nagao: Hierarchical feature construction for image classification using genetic programming, Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016), pp.1423-1428, Budapest, Hungary, 9-12 October, 2016
3. Masanori Suganuma, Toshihiko Nishimura, Tomoharu Nagao, Hiroshi Iseki, Yoshihiro Muragaki, Manabu Tamura, Shinji Minami: Automatic detection of electrical stimulation timing in operation videos of cortical mapping in awake brain surgery, International Conference on Medical Image Computation and Computer Assisted Intervention (MICCAI 2013) Workshop on Modeling

and Monitoring of Computer Assisted Interventions (M2CAI 2013), pp.37-46, Nagoya, Japan, 22 September, 2013

4. Toshihiko Nishimura, Masanori Suganuma, Tomoharu Nagao, Hiroshi Iseki, Yoshihiro Muragaki, Manabu Tamura, Shinji Minami: Intraoperative voice classification for analysis of cortical mapping during awake surgery, International Conference on Medical Image Computation and Computer Assisted Intervention (MICCAI 2013) Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI 2013), pp.27-36, Nagoya, Japan, 22 September, 2013

5. Masanori Suganuma, Tomoharu Nagao: Detection of electrical stimulation position in recorded surgery videos of cortical mapping in awake brain surgery, IEEE 6th International Workshop on Computational Intelligence and Applications (IWCIA 2013), pp.131-136, Hiroshima, Japan, 13 July, 2013

国内学会発表

1. 小林雅幸，菅沼雅徳，崎津実穂，長尾智晴 : 進化的条件判断ネットワークにおける画像分類過程の可視化，第 11 回進化計算学会研究会，2016

2. 菅沼雅徳，土屋大樹，白川真一，長尾智晴 : 遺伝的プログラミングを用いた階層的な特徴構築による画像分類，情報処理学会研究報告 第 108 回 数理モデル化と問題解決（MPS）研究会，Vol. 2016-MPS-108, No. 4, pp. 1-6, 2016

3. 工藤理人，菅沼雅徳，長尾智晴 : ユニットの冗長化による耐故障性を考慮した進化型ニューラルネットワーク，情報処理学会研究報告 第 108 回 数理モデル化と問題解決（MPS）研究会，2016

4. 畠崇人，菅沼雅徳，長尾智晴 : モジュール切替による未知環境に適応可能な探索エージェントの行動制御，情報処理学会研究報告 第 108 回 数理モデル化と問題解決（MPS）研究会，2016

5. 小林雅幸，菅沼雅徳，崎津実穂，長尾智晴 : 進化的条件判断ネットワークの画像分類過程の可視化，情報処理学会研究報告 第 108 回 数理モデル化と問題解決（MPS）研究会，Vol. 2016-MPS-108, No. 1, pp. 1-7, 2016

6. 前原良美，菅沼雅徳，長尾智晴 : 三次元空間把握のための音楽化手法，情報処理学会研究報告 第 13 回 デジタルコンテンツクリエーション (DCC) 研究会，Vol. 2016-DCC-13, No. 3, pp. 1-7, 2016

7. 菅沼雅徳，長尾智晴 : 環境に応じた侵入物体検知を行う監視カメラ，STARC フォーラム 2015, 2015

8. 菅沼雅徳，長尾智晴 : 異常検知のための自己組織化ネットワークとその監視映像への適用，情報処理学会研究報告 第 105 回 数理モデル化と問題解決（MPS）研究会，Vol. 2015-MPS-105, No. 5, pp. 1-6, 2015

9. 崎津実穂，菅沼雅徳，土屋大樹，長尾智晴 : 決定木及び決定ネットワークによる画像分類過程の説明文の自動生成，情報処理学会研究報告 第 105 回 数理モデル化と問題解決（MPS）研究会，Vol. 2015-MPS-105, No. 4, pp. 1-6, 2015

10. 菅沼雅徳，長尾智晴 : 脳の記憶構造に着目した監視カメラからの異常検知，STARC シンポジウム 2015, 2015
11. 菅沼雅徳，長尾智晴，田村学，村垣善浩，伊関洋：覚醒下脳腫瘍摘出術の動画像記録における電気刺激位置の自動検出，電子情報通信学会総合大会，D-7-11，2014
12. 西村俊彦，菅沼雅徳，長尾智晴：覚醒下脳腫瘍摘出術の動画像記録に対する自動解析，情報処理学会第 75 回全国大会，3ZG-5，2013

受賞

1. Masanori Suganuma, Shinichi Shirakawa, Tomoharu Nagao：A genetic programming approach to designing convolutional neural network architectures, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2017), Best paper nomination, 2017
2. 菅沼雅徳，長尾智晴 : 環境に応じて侵入物体検知を行う監視カメラ，STARC フォーラム 2015, 優秀ポスター賞，2015
3. 西村俊彦，菅沼雅徳，長尾智晴 : 覚醒下脳腫瘍摘出術の動画像記録に対する自動解析, 情報処理学会第 75 回全国大会, 学生奨励賞，2013