

Title	カプセル内視鏡画像における周辺領域を考慮した病変追跡に関する研究
Author(s)	柳川, 由紀子
Citation	大阪大学, 2019, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/72606
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

カプセル内視鏡画像における周辺領域を 考慮した病変追跡に関する研究

提出先 大阪大学大学院情報科学研究科

提出年月 2019 年 1 月

柳川 由紀子

目 次

第 1 章	序論	5
1.1	本研究の背景	5
1.2	本研究の概要	6
第 2 章	関連研究	13
2.1	カプセル内視鏡画像の分析	13
2.2	時系列画像中の物体追跡	14
2.2.1	特徴点ベースの物体追跡	15
2.2.2	Convolutional Neural Networks を用いた物体追跡	16
第 3 章	追跡領域の周辺特徴点と三角形の幾何拘束を用いた病変領域追跡	19
3.1	三角形の幾何拘束を用いた周辺特徴点による病変追跡手法	20
3.1.1	ステージ 1 : 三角形の幾何拘束を用いたサポーターのマッチング	22
3.1.2	ステージ 2 : 三角形の幾何拘束条件に基づいた病変位置の投票	31
3.1.3	ステージ 3 : 色特徴量を用いた高精度位置推定	33
3.2	評価実験	34
3.2.1	評価用画像	34
3.2.2	三角形の幾何拘束条件による違い比較	36
3.2.3	従来手法との比較	39

3.3	まとめ	40
第4章 Convolutional Neural Networks を用いたイントラフレーム学習		
	による病変領域追跡	43
4.1	病変と周辺領域のイントラフレーム学習による病変領域追跡	44
4.1.1	イントラフレーム学習のための学習データの選び方と正解値 ラベルの付与	44
4.1.2	CNN のネットワーク構造	46
4.1.3	病変追跡処理における追跡位置の決定方法	52
4.1.4	オンライン学習	52
4.2	評価実験	53
4.2.1	提案手法の評価	53
4.2.2	従来手法との比較	58
4.2.3	矩形サイズおよび初期位置依存における評価	62
4.3	まとめ	64
第5章 特徴点ベースの追跡手法と病変と CNN を用いた追跡手法の比較		
5.1	病変追跡性能の比較	71
5.2	消失フレームの判定	75
5.3	特徴点ベースの追跡手法と病変と CNN を用いた追跡手法を組み合 わせた手法の提案	76
第6章 結論		
		79

目 次

1.1	Examples of gastrointestinal abnormalities in small bowels. Green circles drawn on images show abnormal areas.	11
1.2	Capsule endoscopy	12
1.3	Example sequences of successive images at low frame rate consists of dissimilar images.	12
3.1	Feature matching using a triangular constraint.	21
3.2	Voting-based affine transformation.	22
3.3	Endoscopy images processed using the Speeded-Up Robust Features (SURF) detector. The location and scale of the blue circles identify the abnormalities detected.	23
3.4	Description of supporters.	26
3.5	Point matching results using (a) no constraint (the conventional method), (b) an angular constraint, and (c) an affine constraint.	30
3.6	Movement between successive frames caused by (a) a parallel shift, (b) and (c) rotational movements of the camera. The red rectangles identify abnormal regions. The green rectangles show the attempt to track these abnormalities by assuming only a parallel shift (the conventional method) whereas the blue rectangles show the attempt to track these abnormalities using affine voting (the proposed method).	32
3.7	Decision method of target area.	35
3.8	An example of an image sequence and regions of ground truth for evaluating.	35

3.9	Precision and success plots per sequence. Here the performance of the proposed method is investigated by changing parameters.	37
3.10	Precision and success plots per sequence. Here the performance of the proposed method is compared with that of the conventional method and angular-constraint method.	38
3.11	Examples of the accurate tracking of abnormalities using the affine constraint. The green rectangles in each image are the ground truth region, while the yellow rectangles are our tracking result.	39
3.12	Examples of erroneous tracking using the affine constraint. Green rectangles in each image are the ground truth region, while the yellow rectangles show our tracking result.	40
3.13	Precision and success plots per sequence. Here the performance of the proposed method is compared with that of the mean-shifted method [1] and that of the MDNet method [2] trained using VOT2015 data or capsule endoscopy images.	41
3.14	Precision and success plots per frame, comparing the tracking of our proposed method with that of other state-of-the-art methods. .	42
4.1	A method of creating data sets and training process of the network.	47
4.2	Architecture of CNN.	49
4.3	An order of inputting color channels both query and probe images.	50
4.4	Tracking process on each frames.	51
4.5	Comparison between online training on and off on relationship between precision and mean value of tracking errors on image sequences.	54
4.6	Comparison between online training on and off on relationship between precision and tracking errors on each frame.	55

4.7	Comparison of several abnormalities with online training on relationship between precision and mean value of tracking errors on image sequences.	56
4.8	Comparison of several abnormalities with online training on relationship between precision and mean value of tracking errors on image sequences.	57
4.9	An example of successful tracking (with online training).	58
4.10	An example of unsuccessful tracking (with online training).	59
4.11	Comparison of the conventional technologies and the proposed method on relationship between precision and mean value of tracking errors on image sequences.	60
4.12	Comparison between MDNet[2] , C-COT[3], Staple[4] and the proposed method on relationship between precision and tracking errors on each frame.	61
4.13	Focusing partial range of Fig. 4.12.	62
4.14	Comparison between MDNet[2] and proposed method on output rectangle.	63
4.15	Comparison about sizes of the tracking rectangles on relationship between precision and mean value of tracking errors on image sequences.	64
4.16	Comparison about sizes of the tracking rectangles on relationship between precision and tracking errors on each frame.	65
4.17	Evaluation for misaligned position in a starting frame on relationship between precision and mean value of tracking errors on image sequences.	66

4.18	Evaluation for misaligned position in a starting frame on relationship between precision and mean value of tracking errors on image sequences.	67
4.19	Evaluation for misaligned position in a starting frame on relationship between precision and tracking errors on each frame.	68
4.20	Evaluation for misaligned position in a starting frame on relationship between precision and tracking errors on each frame.	69
5.1	Comparison between feature points matching based method and CNN based method on relationship between precision and tracking errors on each frame.	72
5.2	Comparison between feature points matching based method and CNN based method on output rectangle : Feature points matching method is better then CNN based method in those cases.	73
5.3	Comparison between feature points matching based method and CNN based method on output rectangle : CNN based method is better then feature points matching based method in those cases.	74
5.4	Comparison between feature points matching based method and CNN based method on output rectangle : In those cases , both method are correct.	74
5.5	An example of "stay" as a classified result of CNN based method because of disappeared abnormality.	76

表 目 次

3.1	Detail of lesions in evaluation database.	34
-----	---	----

4.1	Parameter of CNN.	53
5.1	The results of experiments related to detecting the disappearance of the abnormality.	75

業績一覧

博士論文を構成する業績

論文誌

- [1] 柳川 由紀子, 越後 富夫, 宮崎 祐太, 武村 紀子, 八木 康史
“イントラフレーム学習による Convolutional Neural Networks を用いたカプセル内視鏡画像における病変領域の追跡 ”
人工知能学会論文誌, 第 33 巻 6 号, J-STAGE, November, 2018.

- [2] Yukiko Yanagawa, Tomio Echigo, Hai Vu, Hirotochi Okazaki, Yasuhiro Fujiwara, Tetsuo Arakawa, Yasushi Yagi
“ Abnormality tracking during video capsule endoscopy using an affine triangular constraint based on surrounding features ”
IPSJ Transactions on Computer Vision and Applications, vol.9, no.1, pp.1, March, 2017.

査読付き国際学会

- [1] Yukiko Yanagawa, Tomio Echigo, Hai Vu, Hirotochi Okazaki, Yasuhiro Fujiwara, Tetsuo Arakawa, Yasushi Yagi

“Tracking abnormalities in video capsule endoscopy using surrounding features with a triangular constraint.”

In proceeding of the 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp.578-581, Barcelona, Spain, May, 2012.

関連するその他の業績

査読付き国際学会

- [1] Hai Vu, Yukiko Yanagawa, Tomio Echigo, Masatsugu Shiba, Hirotochi Okazaki, Yasuhiro Fujiwara, Tetsuo Arakawa, Yasushi Yagi

“A Web-Based Education System for Reading Video Capsule Endoscopy.”

In proceeding of the 22nd International Conference on Pattern Recognition (ICPR), pp.2721-2726, Stockholm, Sweden, August, 2014.

- [2] Hai Vu, Tomio Echigo, Yuma Imura, Yukiko Yanagawa, Yasushi Yagi

“Segmenting Reddish Lesions in Capsule Endoscopy Images Using a Gastrointestinal Color Space.”

In proceeding of the 22nd International Conference on Pattern Recognition (ICPR), pp.2721-2726, Stockholm, Sweden, August, 2014.

国内学会・研究会

- [1] 大内義成, 越後富夫, 柳川由紀子, 八木康史

“カプセル内視鏡映像における特徴追跡と画像間類似度を用いた連続フレーム検出”

電子情報通信学会総合大会 学生ポスターセッション, 電子情報通信学会, no.ISS-P-263, March, 2013.

- [2] 井村祐満, 柳川由紀子, 越後富夫, 八木康史

“カプセル内視鏡映像の病変局所構造における共起情報を用いた病変検出”

電子情報通信学会 医用画像3月研究会, 電子情報通信学会, no.MI2014-109, pp.271-276, March, 2015.

- [3] 宮崎祐太, 白神康平, 柳川由紀子, 越後富夫, 八木康史

“カプセル内視鏡画像における小腸病変検出のための特徴解析”

研究報告コンピュータビジョンとイメージメディア (CVIM), 電子情報通信学会, pp.1-8, May, 2016.

第1章 序論

1.1 本研究の背景

カプセル内視鏡は 2001 年に米国で認可された新たな小腸の疾患検査のための装置である．日本国内においては 2007 年に Given Imaging 社（現 Medtronic 社）の PillCamTMSB [5] が認可されて以来，活用が進んでいる．従来，消化器官内の異常を見つける検査は，細長い管の先端にカメラを装着した内視鏡が用いられてきた．内視鏡検査はこの管状の内視鏡を胃の検査であれば患者の口から，大腸の検査であれば肛門から医師が直接挿入し，医師がカメラ画像をリアルタイムで確認しながら異常がないかの検査を行う．しかし，小腸は口からも肛門からも最も遠い位置にあり，さらに小腸の伸縮性が高いという特性から従来の内視鏡が進まないため検査が困難であった [6],[7]．小腸の検査を行うために小腸の伸縮を活用しながら内視鏡を進めるように，管の先端に風船のついたダブルバルーン内視鏡 [8] があるが，コントローラで風船を収縮させ，内視鏡を小腸の深部に進めていくという高度な技術が必要とされる．さらに，4－5m ある小腸の中のどの位置に疾患があるかをあらかじめ医師が推定して口か肛門のどちらかに近い方からバルーン内視鏡を挿入して検査する必要があるが，小腸の全域検査には向かない．また，静脈麻酔が必要であるとともに，検査時間が長い入院が必要であり，患者への負担が大きい．一方でカプセル内視鏡を用いる検査では，患者に負担をかけずに小腸内部全域の診察が可能である [9],[10]．

カプセル内視鏡検査では，カプセル内視鏡によって撮影された小腸内部の画像

中に疾患がないかを医師が探索する，読影と呼ばれる作業がある．読影には，従来の内視鏡検査とは異なった診断技術が必要となる．カプセル内視鏡は腸管の蠕動運動により動くため，撮影位置や角度などを調節することはできない．病変の発見のために医師は，撮影された画像すべてを見続け，診断する必要があり負担が大きい．そのため，医師の診断を助けるためにさまざまな診断補助アプリケーションや診断医師の教育システムなどの研究・開発が進められている．

図 1.1 にカプセル内視鏡検査で見つけるべき病変の例を示す．カプセル内視鏡検査画像中の病変は図 1.1 の例からわかるように明確なエッジを持っておらず，背景となる正常領域との間に大きな差がない．病変を自動で検出する研究 [11],[12],[13] も進んでいるが，すべての病変をカバーするものではなく，医師による目視診断に頼っているのが現状である．

1.2 本研究の概要

日本において 2007 年に使用が認可されたカプセル内視鏡である Given Imaging 社（現 Medtronic 社）の PillCam SBTM[5] は，図 1.2 のような CMOS カメラと光源を内蔵した幅 11mm，長さ 26mm のカプセル型カメラ内蔵センサである．

現在，日本国内では Given Imaging 社（現 Medtronic 社）とオリンパス メディカルシステム社のカプセル内視鏡の使用が認可がされている．カプセル内視鏡で撮影された画像は無線で腰の位置に装着した受信端末に送信される．カプセル内視鏡の内部には 8 時間以上駆動が可能なバッテリーを有しており，カメラセンサと光源に電源を供給している．カプセル内視鏡は自己推進力を持たず，腸の蠕動運動によって移動する．そのため，得られる映像の動きは非常に不規則なものである．撮影される腸内の様子はカプセルの向きに大きく依存し，向きや動きによっては病変の発見は困難となる．さらに現在，使用されているカプセル内視鏡のフレー

ムレートは 2 – 6 frame per second(fps) であり，約 8 時間の間に撮影されたの腸管画像の総フレーム数は 2fps でも 57,000 フレームという膨大な量となる．6fps で撮影するカプセル内視鏡を用いるとさらに見なければならない画像が増えることになる．医師 は不規則な動きをする 膨大な量の画像の中から，周辺領域との特徴差があまりない小さな病変を発見する必要がある．その上，カプセル内視鏡の動きによっては病変がわずか数フレームにしか写っていないことも多い．そのため，各フレームを注意深く見る必要があり，長時間の集中を要することになる．本研究では医療画像のなかでも特に撮影時間が長く，診断の際の医師への負担が大きく，手助けするアプリケーション開発が不可欠なカプセル内視鏡の病変の追跡に注目する．

前述のとおり，カプセル内視鏡は，従来の内視鏡を用いた検査では難しかった小腸の病変の有無を検査でき，患者への負担も少ないことから，広く使われることが期待されている．患者が飲み込んだカプセル内視鏡は消化管の蠕動運動によって推進し，胃腸管を通過する約 8 時間の間に低フレームレートで画像を撮影し続ける．医師は長時間撮影された膨大な量の画像の中から些細な変化を見逃さずに病気の有無を診断する必要がある，カプセル内視鏡診断は，従来の内視鏡検査と違ったスキルを必要とする．そのため，コンピュータ支援診断，病変領域を特定するために見やすい表示を行うシステム [14] や医師の技術向上のための病変領域を見つける訓練システム [15] の重要性が高まっている．これらのアプリケーションの研究・開発を進める上で重要な技術の一つに時系列で得られた画像中のある特定領域を追跡する領域追跡技術がある．例えば，コンピュータ支援診断において，病変等の異常部位を検出した後，時系列に得られている画像列の前後フレームの病変位置を追跡し，医師に提示する際に病変を含む画像の中から選択して提示することにより，重複して提示することを防ぎ，より病変の診断に適した画像を提示することが可能になる．また，カプセル内視鏡診断の訓練システムの正解

データ作成にも病変の追跡は有効である。カプセル内視鏡診断の訓練システムには病変の出現フレームから消失フレームまでの病変位置の正解値付きデータが大量に必要である。現在、訓練システムに使われるデータは手作業で作られており、訓練のためのデータの作成に膨大な作業量が必要となっている。一方でカプセル内視鏡を使った病変診断は進んでおり、日々病変を含んだ画像データが蓄積されている。医師が病変の存在するシーケンス中の任意フレームの病変を指定するだけで、その病変の前後フレームを追跡して出現・消失フレームを特定できれば訓練データ作成にかかる作業量を大きく削減できる。そこで、本研究では、カプセル内視鏡画像の病変領域の追跡に焦点を当てる。

従来の画像における物体追跡手法は追跡対象物体の動きを予測し、時事刻々と変化する対象物体の見えの変化にロバストな特徴量を用いることが重要である。しかし、カプセル内視鏡画像の撮影対象である小腸壁は大きく変形するとともに、消化管の蠕動運動によって推進するカプセルに搭載されたカメラから低フレームレートで撮影されるため、追跡対象である病変の動作モデルを予測することが困難である。図 1.3 は 2fps のカプセル内視鏡で撮影された連続するフレームを示している。この図からも蠕動運動により小腸壁が変形し、画像中の見えが大きく変わることが分かる。また、カプセル内視鏡における腸管映像は、血管異常 (phlebectasia, angiodysplasia)、潰瘍 (ulcer) のような病変領域においても病変の特徴が不鮮明であり、近傍の正常領域との変化が小さく、追跡に有効な特徴を定義することも困難である。このように、カプセル内視鏡の病変追跡は動きの予測、追跡に有効な特徴量の定義の両方とも困難であるという課題がある。

一方で、医師は、病変のモニタリングをするときに、病変領域の特徴だけではなく周辺領域の色や模様も手がかりにする。また、人が画像中の不鮮明な対象を追跡する場合においては、対象のみを観察するのではなく、周辺の状況も勘案している。そこで、本研究では病変の追跡において、周辺情報を補助的に用いるこ

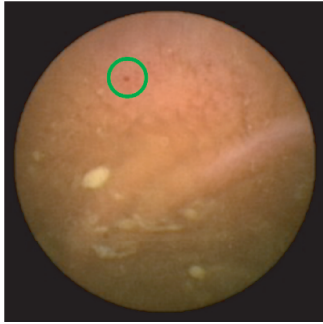
とが有効であると考え、周辺状態を考慮した領域追跡手法として以下の二手法を提案する．

- 病変の周辺の特徴を手がかりとした、周辺の特徴点と病変の幾何拘束に基づく領域追跡（特徴点ベースの追跡手法）[16],[17]
- 病変とその周辺領域を用いたイントラフレーム学習による領域追跡（Convolutional Neural Networks（CNN）を用いた追跡手法）[18]

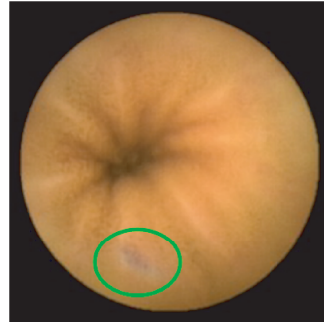
前者では、病変の周辺の特徴を手がかりとした、周辺の特徴点と病変の幾何拘束に基づく領域追跡の手法を提案する．提案手法は特徴点ベースの追跡手法を用いる．小腸における病変は小腸壁に存在するため、周辺の壁と連動する．つまり、局所的には剛体であり、周辺の壁と同じ平面上に存在しており、近接フレームでは病変と周辺領域では幾何学的位置関係が保持される．と仮定できる．そこで、病変の追跡に病変の特徴のみを用いるのではなく、周辺の特徴点を利用し、周辺の特徴点の追跡結果から病変の位置を推定する．また、特徴点の追跡においても、病変追跡と同様に有効な特徴量を定義することが難しいと考えられるため、全ての特徴点は同じ平面上に存在すると仮定し、特徴点のマッチングに三角形の幾何拘束を適用することを提案する．三角形の幾何拘束条件として、三つの特徴点のなす角度による拘束と三つの特徴点によるアフィン行列係数による拘束の二つのアプローチを考察し、評価実験では実際の病変を含むカプセル内視鏡画像列を用い、幾何拘束条件がない場合と二つの三角形の幾何拘束条件を適用した場合についての比較実験ならびに従来研究との比較実験を通して、特徴点ベースの領域追跡において三角形の幾何拘束を用いることの有効性評価を行う．

後者では、CNNを用いて病変とその周辺領域を用いたイントラフレーム学習による領域追跡の手法を提案する．提案手法は検出ベースの追跡手法を用いる．一

般に，カプセル内視鏡画像では病変部と周辺の特徴の差が明確でないため，追跡に有効な特徴をあらかじめ定義するのは困難である．そこで，初期フレームの病変部と周辺の画像を用いて追跡用いる特徴量を CNN を用いてイントラフレーム学習を行うことで，周辺との関係において追跡にロバストな特徴量選択を可能にする．評価実験では実際の病変を含むカプセル内視鏡画像列を用い，従来研究との比較実験を行い，これらの評価実験を通して，イントラフレーム学習による領域追跡手法の有効性評価を行う．



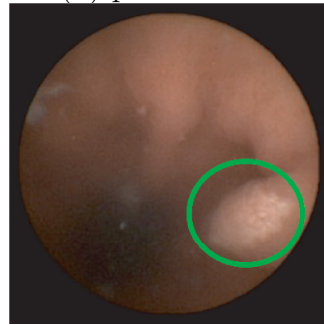
(a) red spot



(b) phlebectasia



(c) angiodysplasia



(d) lymphangiectasia



(e) erosion



(f) erythematous



(g) ulcer



(h) white villi

Fig 1.1: Examples of gastrointestinal abnormalities in small bowels. Green circles drawn on images show abnormal areas.

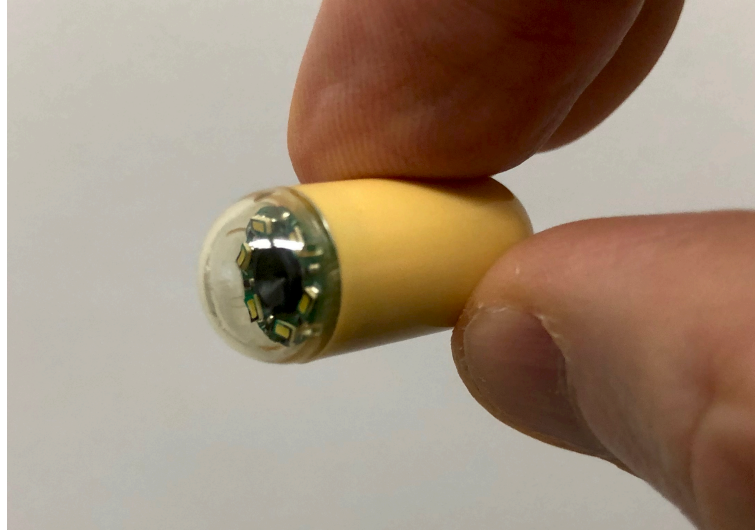
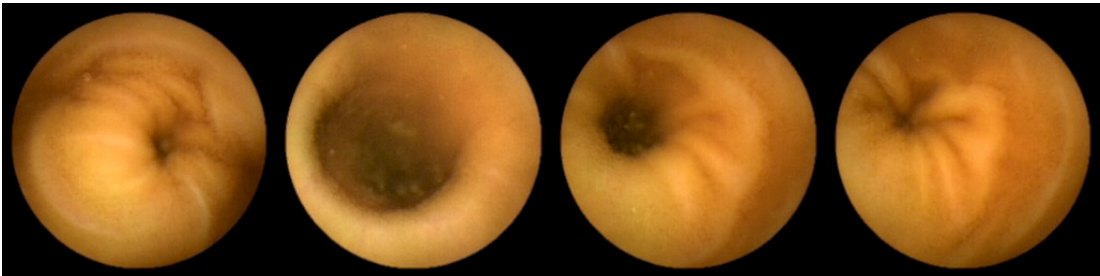
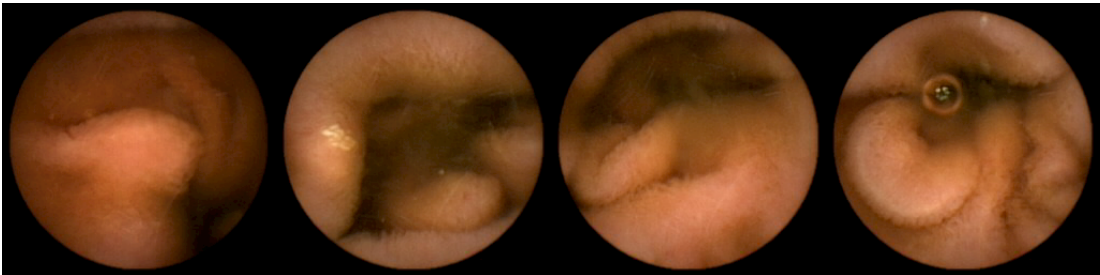


图 1.2: Capsule endoscopy



(a) Example 1



(b) Example 2

图 1.3: Example sequences of successive images at low frame rate consists of dissimilar images.

第2章 関連研究

2.1 カプセル内視鏡画像の分析

内視鏡検査は消化器系の疾患を見つけるのに広く使われている。その中でも、小腸の疾患を見つけるために使われるのがカプセル内視鏡である。前章で述べた通り、カプセル内視鏡は患者が飲み込んでから排出するまで消化器系器管を通る間の全ての時間で画像を撮影し保存し続ける。医師は、撮り溜められた膨大な量の画像から疾患有無を判断する作業をしなければならない。そこで、医師の負担を減らすために、カプセル内視鏡を用いた診断を助ける研究が進んでいる。

カプセル内視鏡を用いた医師の診断を助けるための研究の一番大きなトピックスは病変部の自動検出である。小腸で起こる疾患は図 1.1 で示した red spot, phlebectasia, angiodysplasia, lymphangiectasia, erosion, erythematous, ulcer, white-tipped villi の 8 つをはじめとして多くの種類が存在する。現状の研究ではこれらの疾患の一部を自動で検出する研究が数多くされている。

潰瘍 (ulcer) を検出する研究として、特徴量にテクスチャーを用いる手法 [19], Saliency map を用いる手法 [20], 特徴量に YIQ 色空間のヒストグラムを用い識別器に support vector machine (SVM) を用いる手法 [21], などがある。出血部を検出する研究として、特徴量に Normalized Gray Level Co-occurrence Matrix (NGLCM) of the magnitude spectrum を用い、識別器に SVM を用いる手法 [22], 特徴量に YCbCr 色空間のヒストグラムを用い、識別器に SVM を用いる手法 [12], 特徴量に superpixel-color histogram (SPCH) を用い、識別器に部分空間法を用いる手法

[23], CNN を用いる手法 [24] などが提案されている. びらん (erosions) を検出する研究として, CNN を用いる手法 [25] が提案されている. 腫瘍を検出する研究として, 特徴量に GLCM と DWT を検出器に PNN を用いる手法 [26], 色, エッジ, テクスチャを特徴とし, 識別器に SVM を用いてクローン病を検出する研究 [27] なども報告されている. また, 色とテクスチャを用いたセグメンテーションにより複数の疾患を同時に検出する研究も提案され始めている [28]. これらの研究は小腸で起こりうる疾患の一部を自動で検出するもので, 医師の見落としをなくすことにはつながるが, 医師が診断をしなくてもよいということにはつながらない.

また, 自動で検出できなくても, 医師による診断の手助けになるように, さまざまなシステムの研究がされている. 一つは病変部を強調して表示されるような表示方法の研究 [29] で医師が画像を凝視する必要がなく, 負担を軽減することにつながる. その他にも, 動きの少ない部分は早送りするように表示時間を変更するシステム [30],[31] などがある.

2.2 時系列画像中の物体追跡

時系列画像中における対象物体の追跡技術はコンピュータビジョン分野における重要な技術の一つである. 物体追跡は追跡対象自体や背景に動きや見えの変化があるなかで, 追跡対象をロバストに追跡することが求められている. 画像中の追跡対象物体をロバストに追跡するためには, 追跡対象の動きを正確に予測し, 時事刻々と変化する対象物体の見えの変化にロバストな特徴量を用いることが重要である. 動きの予測については過去の追跡対象の動きから現在の位置を予測するカルマンフィルタ [32], 追跡対象の存在しうる領域を尤度をもとにサンプリングするパーティクルフィルタ [33],[34],[35], 類似度マップの勾配を利用する Mean shift[1] などがある. 追跡対象の見えは, 照明の変化, 向きの変化, 追跡対象自体の形状変化,

周辺物体による対象の遮蔽などによって変化する．これらの変化に対してロバストに対応するために様々な特徴量が提案されている．単一の特徴量として，色ヒストグラムを用いる手法 [36],[37]，輪郭エッジを用いる手法 [38]，テンプレートを用いる手法 [39]，複数の特徴量を組み合わせる手法として，色ヒストグラムをモデル化する手法 [40] や色ヒストグラムと輪郭を組み合わせる手法 [41] などがある．これらの特徴量をより追跡対象と背景を区別するものにするために P-N Learning[42]，online Boosting[43]，などの学習方法も提案されている．また，特定の問題に対応するために提案されている手法もある．追跡対象の隠れに対応するために周辺領域を利用する方法 [44]，対象の変形や隠れに対応するために対象物内の局所特徴点を追跡する特徴点ベースの方法 [45] などが挙げられる．一方，動きモデルを用いない検出ベースの手法 [46] もある．これは，初期フレームを用いて学習した識別モデルを用いて物体検出を毎フレーム行うもので，精度の高い物体追跡を実現している．

近年高い精度を示している手法には相関フィルタに基づく手法 [4],[3],[47] と Convolutional Neural Networks を用いたものがあり，物体追跡のコンペション [48],[49],[50],[51] で両手法を用いたものが上位を占めている．

2.2.1 特徴点ベースの物体追跡

特徴点ベースの物体追跡は対象領域を特徴点の集合とみなし，各特徴点を追跡した結果を用いて領域の追跡を行う．Kanade-Lucas-Tomasi (KLT) 法 [52] は Kanade, Lucas らによって提案された特徴点の検出・追跡を用いた物体追跡手法である．Zhou は Scale-Invariant Feature Transform (SIFT) 特徴量を用いた物体追跡手法 [53] を，Wei らは Speeded-Up Robust Features (SURF) 特徴量を用いた物体追跡手法 [45] を提案している．Grabner らは追跡対象の特徴点だけでなく，追跡対象の周囲の特徴点も用いて物体追跡を行う手法 [44] を提案している．しかし，これらの手法は

特徴点同士のマッチングに点の特徴量を用いており、特徴量が弱い場合に対応していない。Kanazawa と Uemura らはアフィン不変の 2 枚の画像間において 3 つの特徴量の関係を用いて弱い特徴量をマッチングする手法 [54] を提案している。しかし、この方法は局所的に変化する画像に対応していない。

本研究ではこの課題を解決するため、病変の周辺の特徴点を利用したカプセル内視鏡の追跡方法を提案する。三角形の制約を適用することにより、連続する画像フレームに強い特徴を持たない病変を追跡することができる。また、特徴点の追跡によって得られた病変候補位置に対して、さらに色特徴量を用いて詳細な追跡を行うことによって、画像中の局所的な変化にも対応する。

2.2.2 Convolutional Neural Networks を用いた物体追跡

Convolutional Neural Networks (CNN) は、2012 年の ImageNet Large Scale Visual Recognition Challenge (ILSVRC2012), [55] にて AlexNet[56] が人の設計した特徴量を用いた手法に対し、性能面で圧倒的な差を示して以来、画像認識 [57],[58],[56],[59], セマンティックセグメンテーション [60],[61],[62],[63], 物体検出 [64],[65], などコンピュータビジョンにおける様々な要素技術に広く展開されている。

物体追跡技術についても Visual Object Tracking 2015(VOT2015)[48] にて優勝した MDNet[2] は CNN を用いたものであり、その後も様々な CNN の手法が提案されている [66],[67],[68],[69],[70]。VOT2015 では 3 つの CNN を用いた手法が提案され、そのうち事前学習をすでに公開されている追跡用のデータベースを用いて行い、追跡時には全結合層を 1 つ追加し、与えられた画像と追跡対象を用いて追加した全結合層の重みのみを学習する MDNet[2] が優勝し、SRDCF[71] の特徴量を CNN で学習させたものを用いた Deep SRDCF[67] が 2 位という結果であった。CNN を用いた追跡手法は追跡対象を人・車などに絞って事前学習にそれらの画像を用いるもの [68], 物体の動作などから一般物体認識用に学習されたモデルを追

跡時に追跡対象に応じたものに変換するもの [69],[70] などが提案されている．フレームレートが低く，動作モデルの予測が難しいカプセル内視鏡画像の病変の追跡には有効な特徴量を定義することが重要であり，追跡対象ごとにロバストな特徴量を学習から自動的に定義できる CNN を用いることが有効と考えられる．しかし，CNN を用いた物体追跡はあらかじめ大量の類似画像による学習データが必要となるという大きな課題がある．本論文ではこの課題を解決するために，開始フレームと追跡領域の初期位置のみから CNN の学習に必要な大量の学習用データ作成し学習することで，事前に大量の類似画像を学習用データとして集めて学習する必要としないイントラフレーム学習による追跡手法を提案する．

第3章 追跡領域の周辺特徴点と三角形の幾何拘束を用いた病変領域追跡

カプセル内視鏡画像中の小腸の病変の正確な追跡は，病変診断の訓練システムのデータ作成に有用である．しかしながら，撮影対象の小腸壁は予測ができないほど変形する(図1.3)とともに，消化管の蠕動運動によって推進するカプセルに搭載されたカメラから低フレームレートで撮影されるため，動作モデルを予測することが困難である．図1.3は2fpsのカプセル内視鏡で撮影された連続するフレームを示している．この図からも蠕動運動により小腸壁が変形し，画像中の見えが大きく変わることが分かる．また，消化器内の血管異常(phlebectasia, angiodysplasia)，潰瘍(ulcer)のような病変領域においても近傍の正常領域との変化が小さく，追跡に有効な特徴量を定義することも困難である．

本章では，この問題を解決するために，病変の周辺の特徴点を利用したカプセル内視鏡の追跡方法を提案する．小腸における病変は小腸壁に存在するため，周辺の壁と連動する．つまり，局所的には剛体であり，周辺の壁と同じ平面上に存在していると仮定できる．そこで，病変の追跡に周辺の特徴点を利用し，周辺の特徴点の追跡結果から病変の位置を推定する．また，特徴点の追跡においても，病変追跡と同様に有効な特徴量を定義することが難しいと考えられるため，全ての特徴点は同じ平面上に存在すると仮定し，特徴点の追跡に三角形の幾何拘束を適

用する．これにより，連続する画像フレーム間で強い特徴を持たない特徴点を追跡することができる．特徴点の追跡後，病変領域の特定には候補領域に対して色特徴を用いて最適化させる．本論文では本提案手法のことを特徴点ベースの追跡手法と呼ぶ．

3.1 三角形の幾何拘束を用いた周辺特徴点による病変追跡手法

追跡のターゲットとなるカプセル内視鏡画像内の病変領域は医師や検査技師による手動指定，またはコンピューターによる自動検出によって得られるとする．本研究ではそのような方法で得られたカプセル内視鏡画像列の病変の存在するフレームから前後フレームへとそれぞれ病変の追跡を行う．

本章では病変領域の追跡を病変の特徴のみならず，病変周辺の特徴点とその特徴点同士の位置関係，また，病変領域と周辺特徴点の位置関係を用いて推定する手法を提案する．

病変周辺の特徴点のことをここではサポーターと呼ぶ．カプセル内視鏡画像は追跡に適した特徴が得難いため，連続するフレーム間でサポーターを個別に追跡することも難しい．カプセル内視鏡に映る小腸壁は局所的に見れば剛体であり，全てのサポーターは同じ平面に存在すると仮定できる．同じ平面に存在する3点はカメラの撮影位置が変わっても幾何学的に位置関係が保持される．そこで，連続するフレーム間のサポーターの追跡に3つのサポーターの位置関係による幾何拘束を適用することを提案する．本研究では3つのサポーターのなす三角形の角度による幾何拘束条件と3つのサポーターのアフィン行列のパラメータによる幾何拘束条件の二つを提案する．また，連続するフレーム間ではサポーターと病変領域が同じ平面上にあると仮定し，サポーターと病変の位置関係が保持されること

を利用して病変の位置推定を行う。提案手法は以下の3つのステージで行う。

最初のステージでは、追跡ターゲット（病変領域）の周辺部領域において特徴点（サポーター）を検出し、そのサポーターを連続するフレーム間で追跡（マッチング）する。特徴量が明確に得られないカプセル内視鏡の画像ではサポーターのマッチングを特徴量のマッチングのみで一意に決定するのが困難なため、図3.1の左の図のように、一つのサポーターに対して複数のサポーターが対応付けされることが起こり得る。そのため、サポーターを完全に独立させて扱うと正確に追跡することが難しい。この課題に対応するために本研究では図3.1の右の図のように、連続するフレーム間でのサポーターの相対位置関係に基づく三角形の拘束を導入する。

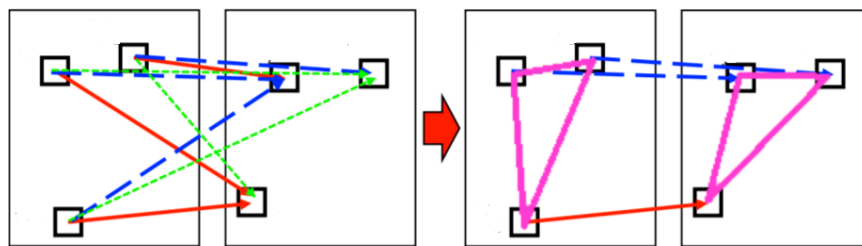


図 3.1: Feature matching using a triangular constraint.

次のステージでは病変の追跡位置を病変周辺の3つのサポーターペアから算出されるアフィン行列に基づいた投票によって大まかに推定する(図3.2)。これは、病変と周辺のサポーターの位置関係が局所的に同じ平面上で維持されているという仮定に基づいている。

最後のステージでは病変固有の特徴を用いて正確な位置の推定を行う。前の二つのステージではカプセル内視鏡画像列における前後フレームでは小腸壁に変形がないという、剛体仮定を用いて病変位置の推定を行ってきた。しかし、実際的小腸壁は非剛体であり、カプセル内視鏡画像列の局所的な前後フレームであっても変形を伴うことがある。とはいえ、カプセル内視鏡画像列の前後フレームでは

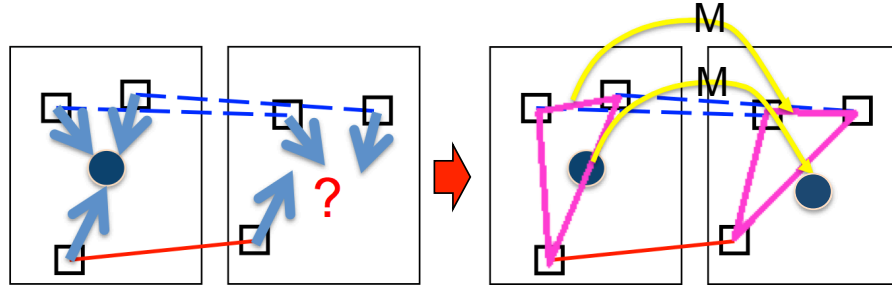


図 3.2: Voting-based affine transformation.

時間変化が小さいため，剛体仮定によって前二つのステージで推定した位置から大きく外れることはない．そこで，最後のステージでは二つ目のステージで推定した病変位置の近傍を病変そのものの色特徴を用いて探索することで，正確な位置の推定を行う．

3.1.1 ステージ 1：三角形の幾何拘束を用いたサポーターのマッチング

ステージ 1 ではカプセル内視鏡画像列の連続フレーム t と $t-1$ の間で一致するサポーターペア集合 \mathbf{P}^t を作成する．

特徴点の検出手法は古くより SIFT[72]，SURF[73] をはじめとしてさまざまな方法が提案されている．本研究ではフレーム間で同じ位置に検出される特徴点が必要な数を得られればよいとため，特に特徴点の検出手法は問わない．本論文では実験的に試した結果，一番性能のよかった KLT 法 [52] を特徴点の検出手法として用いた．小腸内の小領域では明確なテクスチャを有さないため，小領域での領域ベースの特徴量を元に特徴点の位置を決定する SIFT，SURF は安定しなかったと考えられる．それに比べて KLT 法は通常， 3×3 ピクセルより小さな領域で定義されるため，色の変化にロバストであったと推定される．図 3.3 に SURF を用いてカプ

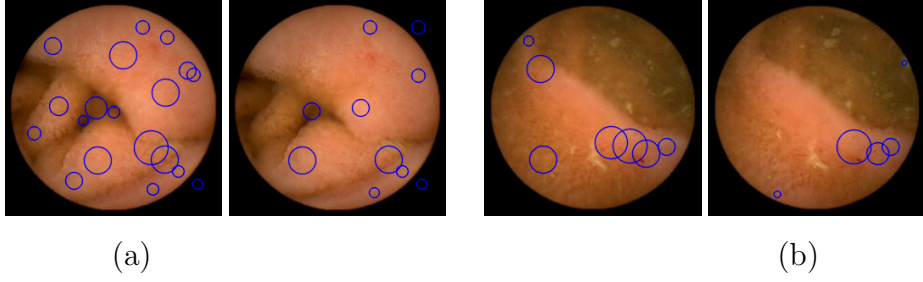


図 3.3: Endoscopy images processed using the Speeded-Up Robust Features (SURF) detector. The location and scale of the blue circles identify the abnormalities detected.

セル内視鏡画像の連続するフレーム間で検出した特徴点を示す．連続するフレーム間で特徴点の位置，数が安定しないのが見て取れる．本論文では KLT 法は特徴点の検出にのみ使い，KLT 法で使われている特徴量を特徴点マッチングには用いていない．

サポーターペア集合 \mathbf{P}^t を作成する手法を説明する．全てのサポーターペア集合 $\mathbf{P}_{\text{all}}^t$ は式 3.1 で表される．

$$\mathbf{P}_{\text{all}}^t = \{\mathbf{p}_{(i,l)}^t \mid i \in N_p^t, l \in N_p^{t-1}\}, \quad (3.1)$$

$\mathbf{P}_{\text{all}}^t$ は t と $t-1$ フレーム目で検出された全てのサポーターの組み合わせである $N_p^t \times N_p^{t-1}$ 個のサポーターペア集合となる．ここで， $\mathbf{p}_{(i,l)}^t$ はサポーターペア， N_p^t と N_p^{t-1} はそれぞれ t と $t-1$ フレームにおける特徴点の数と定義する．

サポーターペア

$$\mathbf{p}_{(i,l)}^t = (\mathbf{F}_i^t, \mathbf{F}_l^{t-1}), \quad (3.2)$$

は要素

$$\mathbf{F}_i^t = (\mathbf{x}_i^t, \mathbf{f}_i^t), \quad (3.3)$$

$$\mathbf{f}_i^t = \{\mathbf{h}_{i1}^t, \mathbf{h}_{i2}^t, \dots, \mathbf{h}_{iH}^t\}, \quad (3.4)$$

を持つ。

ここで、 \mathbf{F}_i^t は t フレーム目における i 番目の特徴点（サポーター）で、 \mathbf{x}_i^t は t フレーム目における i 番目の特徴点の座標位置、 \mathbf{f}_i^t は KLT 法により抽出された特徴点周りの特徴量を表している。本研究では、サポーターのマッチングに使う特徴量を KLT 法で提案されたものではなく、Vu Hai らによって提案された GI color histogram [74] を用いた。GI color histogram はカプセル内視鏡の画像における色特徴を記述する効果的な手段として提案された特徴量である。GI color histogram は、さまざまな患者から得られたカプセル内視鏡検査画像シーケンスの主成分分析結果の 3 番目のコンポーネントを使用し、画像を過度に強調することなく、病変と疑わしい領域を正常領域からよりよく区別する特徴量である。 H は、ヒストグラムの要素数、 \mathbf{h}_{ia}^t は、ヒストグラムの a 番目の要素の値を示す。 \mathbf{f}_i^t は正規化されたヒストグラムである。

サポーターの検出には、画像を縦横半分ずつの 4 分割にし、分割したそれぞれのエリアで検出されるサポーターの数が同じになるようにした。

サポーターペアを決めるためのマッチングスコアには特徴量 \mathbf{f}_i^t と \mathbf{f}_l^{t-1} のマッチングスコア $\mathbf{Bc}_f(\mathbf{p}_{(i,l)}^t)$ に三角形拘束による重み $\mathbf{Mw}(\mathbf{p}_{(i,l)}^t)$ を加味する。 \mathbf{f}_i^t と \mathbf{f}_l^{t-1} の特徴量のマッチングスコアにはバチャタリア距離 $\mathbf{Bc}_f(\mathbf{p}_{(i,l)}^t)$ を用いた（式 3.5）。

$$\mathbf{Bc}_f(\mathbf{p}_{(i,l)}^t) = \sum_{a=1}^H \sqrt{h_{ia}^t h_l^{t-1}{}_a}. \quad (3.5)$$

さらに、特徴量のマッチングスコアフラグ $\mathbf{J}_f(\mathbf{p}_{(i,l)}^t)$ を式 3.6 のように定義する。ここで、 \mathbf{Th}_B は特徴量のマッチングスコアの閾値である。

$$\mathbf{J}_f(\mathbf{p}_{(i,l)}^t) = \begin{cases} 1, & \text{if } \mathbf{Bc}_f(\mathbf{p}_{(i,l)}^t) \geq \mathbf{Th}_B, \\ 0, & \text{otherwise,} \end{cases} \quad (3.6)$$

カプセル内視鏡画像は局所的には剛体であり，連続フレームでのサポーター同士は同じ平面に存在し，同じ動きをしていると仮定できる．そこで，本研究では連続するフレーム間で3つの特徴点の位置関係が維持するように三角形による幾何拘束の重み $\mathbf{Mw}(\mathbf{p}_{(i,l)}^t)$ を特徴点マッチングのスコアに加味する．

三角形拘束の重み $\mathbf{Mw}(\mathbf{p}_{(i,l)}^t)$ を加味した，サポーターペア集合 $\mathbf{P}_{\text{all}}^t$ に格納されたサポーターペア $\mathbf{p}_{(i,l)}^t$ のマッチングスコアは $\mathbf{Zc}_{(i,l)}^t$ のように定義する (式 3.7)．

$$\mathbf{Zc}_{(i,l)}^t = \mathbf{Mw}(\mathbf{p}_{(i,l)}^t) \mathbf{Bc}_f(\mathbf{p}_{(i,l)}^t) \quad (3.7)$$

このスコアを用いて連続する画像間でマッチングするサポーターペア集合 \mathbf{P}^t を作る．サポーターペアの集合 $\mathbf{P}_{\text{all}}^t$ からマッチングするサポーターペア集合 \mathbf{P}^t を作るには3つのステップがある．

まず，前処理として全てのサポーターペア \mathbf{p}_x^t をサポーターペア集合 \mathbf{S}^t に格納する．

ステップ1：サポーターペアの集合 \mathbf{S}^t に格納されている，全てのサポーターペア \mathbf{p}_x^t に対してマッチングスコア $\mathbf{Zc}_{(i,l)}^t$ を計算する．

ステップ2：サポーターペア集合 \mathbf{S}^t の中で最も高いマッチングスコア $\mathbf{Zc}_{(i,l)}^t$ をもつサポーターペア \mathbf{p}_x^t をサポーターペア集合 \mathbf{P}^t に格納する．

ステップ3：サポーター \mathbf{F}_i^t と \mathbf{F}_l^{t+1} を含むサポーターペアを，サポーターペア集合 \mathbf{S}^t から削除する．

ステップ2に戻る．

ステップ 2-3 のプロセスを，すべてのサポーターペアがサポーターペア集合 \mathbf{S}^t からなくなるまで繰り返す．

本研究はで三角形拘束の重み拘束の条件として角度による拘束とアフィン行列のパラメータによる拘束の二つを提案する．

角度による三角形拘束

カプセル内視鏡画像は，連続するフレーム間で局所的には剛体であり同じ動きをすると仮定できる．そこで，連続するフレーム間に三角形の形状を維持するサポーターと一致するように，三角形の角度による幾何拘束条件を提案する．

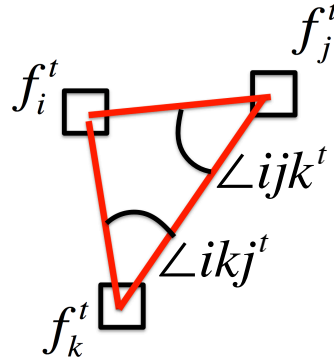


図 3.4: Description of supporters.

図 3.4 に示す通り，サポーター \mathbf{F}_i^t を含む 3 点のサポーター \mathbf{F}_i^t ， \mathbf{F}_j^t ， \mathbf{F}_k^t による三角形の作る角のうち，サポーター \mathbf{F}_j^t を中心とするものを $\angle ijk^t$ ， \mathbf{F}_k^t を中心とするものを $\angle ikj^t$ とする．サポーター \mathbf{F}_i^t と他のサポーター間の関係を示す \mathbf{G}_i^t は式 3.8 と式 3.9 で示すことができる． t フレームにおける，1 つのサポーター \mathbf{F}_i^t を含む三角形は $_{N_p^t-1}C_2$ 個存在することとなる．

$$\mathbf{G}_i^t = (\mathbf{g}_{i,1,2}^t, \dots, \mathbf{g}_{i,n-1,n}^t) \quad (3.8)$$

$$\mathbf{g}_{i,j,k}^t = (\angle \mathbf{ijk}^t, \angle \mathbf{ikj}^t) \quad (3.9)$$

サポーターペア $\mathbf{p}_{(i,l)}^t$ の角度による三角形拘束の重み $\mathbf{Mw}(\mathbf{p}_{(i,l)}^t)$ は、式 3.10 のように表される。

$$\mathbf{Mw}(\mathbf{p}_{(i,l)}^t) = \mathbf{J}_f(\mathbf{p}_{(i,l)}^t) \mathbf{T}(\mathbf{G}_i^t, \mathbf{G}_l^{t+1}) \quad (3.10)$$

$$\mathbf{T}(\mathbf{G}_i^t, \mathbf{G}_l^{t+1}) = \frac{\sum t}{(N_p^t - 1)(N_p^t - 2)} \quad (3.11)$$

$$t = \begin{cases} 1 & \text{if } \angle ijk \simeq \angle lmn \cap \angle ikj \simeq \angle lnm \\ 0 & \text{otherwise} \end{cases} \quad (3.12)$$

ここで、 $\sum t$ は連続する各フレームにおいてサポーター \mathbf{F}_i^t および、サポーター \mathbf{F}_l^{t+1} を含む三角形が大きく形状を変えなかった数を数えたものである。もし、サポーター \mathbf{F}_i^t および、サポーター \mathbf{F}_l^{t+1} を含む三角形のなす角の 1 つが 180 度に近ければその三角形はカウントしないものとする。

アフィン行列のパラメータによる拘束

前項で三角形の角度によるサポーターのマッチングの幾何拘束条件を説明した。角度による幾何拘束条件は、多少の物体の変形を許容した弱い特徴量の追跡を対象に設計されている。しかし、角度による幾何拘束条件はペアの一致に高い柔軟性をもたらすので、拘束条件としては不十分であった。そこで本項では、3 つのサポーターペアから計算されるアフィン行列のパラメータによる幾何拘束の方法を提案する。

サポーターペア $\mathbf{p}_{(i,l)}^t$ のアフィン行列による三角形拘束の重みは、

$$\mathbf{Mw}(\mathbf{p}_{(i,l)}^t) = \mathbf{J}_{\mathbf{f}}(\mathbf{p}_{(i,l)}^t) \frac{\sum_j \sum_m \sum_k \sum_n \mathbf{J}_{\mathbf{f}}(\mathbf{p}_{(j,m)}^t) \mathbf{J}_{\mathbf{f}}(\mathbf{p}_{(k,n)}^t) \mathbf{DM}(\mathbf{p}_{(i,l)}^t, \mathbf{p}_{(j,m)}^t, \mathbf{p}_{(k,n)}^t)}{\sum_j \sum_m \sum_k \sum_n \mathbf{J}_{\mathbf{f}}(\mathbf{p}_{(j,m)}^t) \mathbf{J}_{\mathbf{f}}(\mathbf{p}_{(k,n)}^t)}, \quad (3.13)$$

$$j = [1, 2, \dots, N_p^t - 1, j \neq i], \quad (3.14)$$

$$k = [j + 1, j + 2, \dots, N_p^t, k \neq i], \quad (3.15)$$

$$m = [1, 2, \dots, N_p^{t-1} - 1, m \neq l], \quad (3.16)$$

$$n = [m + 1, m + 2, \dots, N_p^{t-1}, n \neq l], \quad (3.17)$$

ここで

$$\mathbf{DM}(\mathbf{p}_{(i,l)}^t, \mathbf{p}_{(j,m)}^t, \mathbf{p}_{(k,n)}^t) = \begin{cases} 0, \\ \text{if Either the triplet supporters } (\mathbf{F}_i^t, \mathbf{F}_j^t, \mathbf{F}_k^t), \text{ or } (\mathbf{F}_l^{t-1}, \mathbf{F}_m^{t-1}, \mathbf{F}_n^{t-1}) \text{ is collinear.} \\ 1, \\ \text{if all parameters of the affine matrix } \mathbf{A}_{(i,l),(j,m),(k,n)}^t \text{ are within acceptable value.} \\ 0, \text{ otherwise} \end{cases} \quad (3.18)$$

$$\mathbf{p}_{(j,m)}^t = (\mathbf{F}_j^t, \mathbf{F}_m^{t-1}), \quad (3.19)$$

$$\mathbf{p}_{(k,n)}^t = (\mathbf{F}_k^t, \mathbf{F}_n^{t-1}). \quad (3.20)$$

アフィン行列の計算に使われるサポーターペアを $\mathbf{p}_{(i,l)}^t, \mathbf{p}_{(j,m)}^t, \mathbf{p}_{(k,n)}^t$ とすると、ア

フィン行列は次の式で表される.

$$\begin{aligned}\mathbf{A}_{(i,l),(j,m),(k,n)}^t &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \\ &= \mathbf{A}_{\text{Scale}} * \mathbf{A}_{\text{Rotation}} * \mathbf{A}_{\text{Shear}} * \mathbf{A}_{\text{Translation}},\end{aligned}\quad (3.21)$$

ここで,

$$\mathbf{A}_{\text{Scale}} = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \end{bmatrix}, \quad (3.22)$$

$$\mathbf{A}_{\text{Rotation}} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \end{bmatrix}, \quad (3.23)$$

$$\mathbf{A}_{\text{Shear}} = \begin{bmatrix} 1 & \tan \alpha_y & 0 \\ \tan \alpha_x & 1 & 0 \end{bmatrix}, \quad (3.24)$$

$$\mathbf{A}_{\text{Translation}} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix}. \quad (3.25)$$

と定義する.

アフィン行列 $\mathbf{A}_{x,y,z}^t$ はスケール変換パラメータ (S_x, S_y) , 角度歪みパラメータ (α_x, α_y) , 回転角パラメータ θ , および平行移動パラメータ (t_x, t_y) によって定義される. ここで, 要素 a_{13} と a_{23} は, 平行移動パラメータ (t_x, t_y) によって定義される. 残りの要素 a_{11} , a_{12} , a_{21} , a_{22} はスケール変換パラメータ (S_x, S_y) , 角度歪みパラメータ (α_x, α_y) , 回転角パラメータ θ によって得られた結果である. 式 3.21 - 3.25 本研究では, これらのパラメータ値の許容範囲を次のとおりに設定した,

- $-20^\circ < \text{rotational angle } \theta < 20^\circ$,

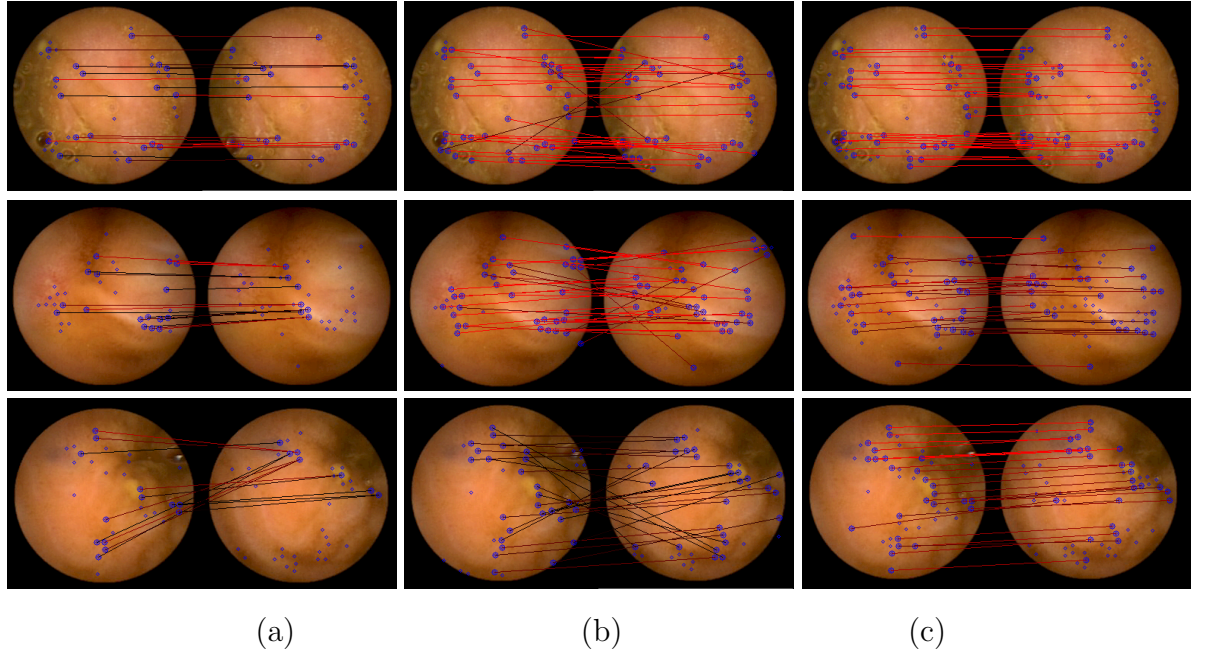


図 3.5: Point matching results using (a) no constraint (the conventional method), (b) an angular constraint, and (c) an affine constraint.

- $-10^\circ < \text{shearing angle} < 10^\circ$,
- $0.25 < \text{scale change} < 3.0$,
- maximum translational range: $1/2$ image size.

これらのパラメータ許容値は，実験的に決定した．

図 3.5 は，拘束を用いずに特徴点のマッチングスコアのみを用いた場合 (no constraint (the conventional method)) と，三角形の角度による拘束 (an angular constraint) と三角形のアフィン行列のパラメータによる拘束 (an affine constraint) のそれぞれのサポーターのマッチング結果を示している．青色の点は検出された点 (サポーター) を表し，赤色の線がマッチング結果である．線の彩度はマッチングスコアによって変わっており，高い彩度が高いマッチングスコアを表す．

3.1.2 ステージ 2 : 三角形の幾何拘束条件に基づいた病変位置の投票

このステージでは、病変の大まかな位置を推定する．連続フレーム間でサポーターペア集合 \mathbf{P}^t を作成した後、病変位置を推定する投票マップを作成する．病変位置は、3 点のサポーターペアから計算されるアフィン行列を使用して、投票マップ上に投票される．従来は、カメラが平行移動すると仮定して、追跡対象の位置の投票にサポーターペア 1 つに対して 1 つの投票を行う手法が用いられてきた（図 3.2 左）．しかし、従来手法では、カメラの回転運動や小腸壁に対して斜めに移動する運動に対処することはできない．この課題を克服するために、本研究では、追跡対象の位置を投票するとき、3 つのサポーターペアと病変位置との間のアフィン変換の関係を使用する．図 3.6 は、従来の方法 (a parallel shift (the conventional method)) と本提案手法 (affine voting (the proposed method)) の病変位置の投票結果を示している．青色の矩形は病変領域を示し、赤色の矩形は従来の方法で推定した病変領域を示し、緑色の矩形は提案手法を用いて病変領域を推定した結果を示している．図 3.6(a) は画像がカメラが並行移動した場合の連続 2 フレームで、図 3.6(b) と (c) はカメラが回転移動した場合の連続 2 フレームの例本提案手法が従来の手法よりも優れていることがわかる．提案手法による投票マップの作成方法を、以下で詳しく説明する．

投票マップにおける各位置 (x, y) のスコアは式 3.26 と定義する．

$$\mathbf{C}(x, y) = \sum_{\mathbf{p}_{(i,l)}^t}^{N_s^t-2} \sum_{\mathbf{p}_{(j,m)}^t}^{N_s^t-1} \sum_{\mathbf{p}_{(k,n)}^t}^{N_s^t} \mathbf{f}(x, y) \mathbf{B}(\mathbf{p}_{(i,l)}^t) \mathbf{B}(\mathbf{p}_{(j,m)}^t) \mathbf{B}(\mathbf{p}_{(k,n)}^t), \quad (3.26)$$

ここで、

$$\mathbf{f}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2}\right)\right) \quad (3.27)$$

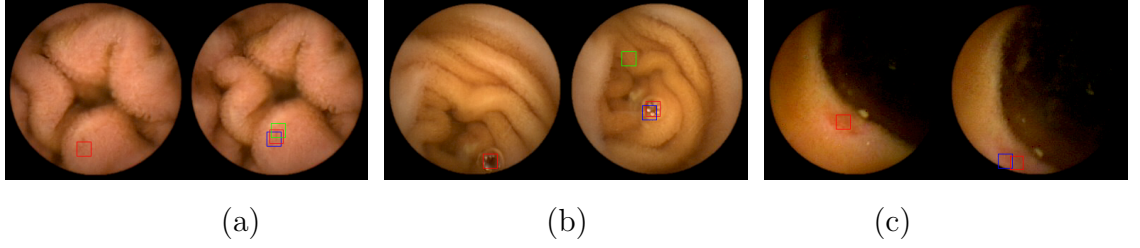


図 3.6: Movement between successive frames caused by (a) a parallel shift , (b) and (c) rotational movements of the camera. The red rectangles identify abnormal regions. The green rectangles show the attempt to track these abnormalities by assuming only a parallel shift (the conventional method) whereas the blue rectangles show the attempt to track these abnormalities using affine voting (the proposed method).

$$\begin{bmatrix} \mu_x \\ \mu_y \\ 1 \end{bmatrix} = \mathbf{A}_{s,t,u} \begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix}. \quad (3.28)$$

は，アフィン行列 $\mathbf{A}_{s,t,u}^t$ を用いて生成された病変の推定位置までの距離の重みである．追跡領域は， x 軸と y 軸に平行な辺を持つ長方形として定義する． (σ_x, σ_y) は，追跡対象の大きさ， $\mathbf{Bc}_f(\cdot)$ は特徴量のマッチングスコアである．投票スコア $\mathbf{C}(x, y)$ は，3 つの特徴量マッチングスコア $\mathbf{Bc}_f(\cdot)$ と重み関数 $\mathbf{f}(x, y)$ を乗算して計算される．重み関数 $\mathbf{f}(x, y)$ は三つのサポーターペアから推定された病変の位置からの距離によって定義される．

前フレームの病変位置 (t_x, t_y) とアフィン変換行列を用いて，現在のフレームにおける病変の推定位置 (μ_x, μ_y) は式 3.28 のように計算する．

3.1.3 ステージ3：色特徴量を用いた高精度位置推定

次に，正確な病変領域をどのように追跡するかについて説明する．ステージ2の投票結果は，小腸が局所的に変形するため実際の病変領域の位置と誤差がある．そこで，ターゲットと周辺領域が変形可能である状況において，より信頼性の高いターゲットの色情報を使用して詳細な位置推定を行う．最初のフレームで，小腸内の病変領域を区別するために開発された GI color histogram を用いて病変領域の特徴量を定義する．

次に，ステージ2で作成した投票マップのスコアが閾値よりも大きいすべての位置について，式 3.29 を用いてスコアを計算する．

$$\mathbf{V}(x, y) = \mathbf{C}(x, y) \mathbf{Bc}_a(\mathbf{Q}_{initial}, \mathbf{K}_{(x, y)}) \quad (3.29)$$

$$\mathbf{Bc}_a(\mathbf{Q}_{initial}, \mathbf{K}_{(x, y)}) = \sum_{a=1}^H \sqrt{q_{ia}^t k_l^{t-1}{}_a}, \quad (3.30)$$

ここで (x, y) は位置であり， $\mathbf{C}(x, y)$ はステージ2で計算された投票スコアである． $\mathbf{Bc}_a(\mathbf{Q}_{initial}, \mathbf{K}_{(x, y)})$ は初期フレームの病変の GI color histogram 特徴と現在のフレーム内の位置 (x, y) における GI color histogram 特徴の Bhattacharyya の距離である．このスコア $\mathbf{V}(x, y)$ が最も高い値をとる位置を病変領域とする． $\mathbf{V}(x, y)$ が最も高い値をとる位置の $\mathbf{Bc}_a(\mathbf{Q}_{initial}, \mathbf{K}_{(x, y)})$ があらかじめの実験により設定した閾値より低く，画像領域内の投票スコアが閾値以下である場合，現在のフレームは病変を含まないと判断する．

3.2 評価実験

3.2.1 評価用画像

評価用画像として 1.1 で示した病変のうち, red spot, phlebectasia, angiodysplasia, erosion, erythematous, ulcer, whitevilli の 7 種類の小腸病変を含んだ 120 シーケンスのカプセル内視鏡画像を用いた. 120 シーケンスの病変毎の内訳と各病変の特徴を表 3.1 に示す.

表 3.1: Detail of lesions in evaluation database.

病変名	シーケンス数	見た目の特徴
erythematous	34	正常領域よりも赤く, 範囲が広い.
erosion	24	病変の中央領域が白く, その周囲が少し赤みを帯びることもある.
phlebectasia	6	静脈が浮き出ているため, 青く線状に見られる.
red spot	10	赤い小さな斑点状の病変. 赤みが強く, 領域が小さい.
ulcer	12	正常領域よりも色が白い病変.
white villi	4	正常領域よりも色が白い病変. ulcer に比べて白色がはっきりしており, 白い斑点が集まっているような見た目.
angiodysplasia	30	毛細血管の拡張症で正常領域よりも赤く, red spot よりも広範囲に広がっている.
総数	120	

評価用画像を撮影したカプセル内視鏡は Given Imaging 社の PillCamTMSB を用いている. 撮影画像サイズは 256 × 256 ピクセル, フレームレートは 2fps である. 患者の数は 36 名. 病変の初期位置を小腸病変検査の専門の医師が手動で指定し (3.7 参照), 指定されたフレームから前後 100 フレームを切り出した.

追跡フレームにおける病変の正解位置は医師が設定した初期位置をもとに専門

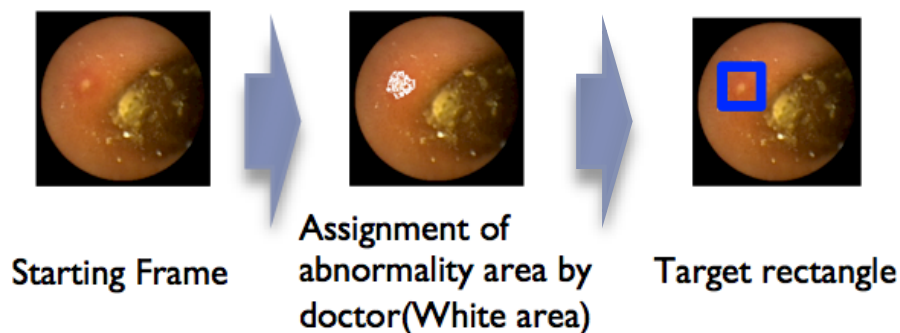


図 3.7: Decision method of target area.

知識を持たない人が手動で矩形を設定した．評価には病変正解位置の入ったフレームのみを用いた．病変正解位置の入ったフレーム数は 2~100 フレームでシーケンスによって異なる．3.8 に評価に用いたカプセル内視鏡画像のサンプルシーケンスを示す．青枠領域が追跡対象の正解病変位置である．

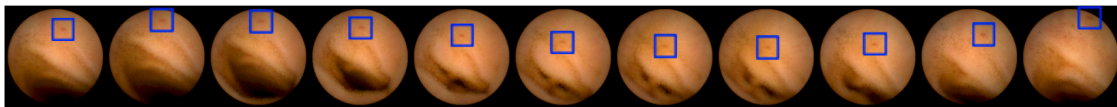


図 3.8: An example of an image sequence and regions of ground truth for evaluating.

実験に用いたカプセル内視鏡は自力で動くことができず，小腸内を蠕動運動によって移動する．蠕動運動がない場合にその場にとどまって揺らぐことはあるが，小腸の直径は小さいため，移動量は小さい．カプセル内視鏡の移動が小さい場合，小腸内は狭く，病変とカプセル内視鏡の距離は大きく変わらないため，画像の見えが変わっても病変サイズは大きく変わらない．また，カプセル内視鏡が大きく移動する場合，画像に写る病変サイズが変化することもあると考えられるが，この場合は病変に対する照明不足や画像からはみ出すことによる追跡不能になることが多

い。大きな病変は蠕動運動によりサイズが変化することも考えられるが、本論文では図 1.1 に示す通り、小さな病変を対象としたため、評価用シーケンスでは蠕動運動による病変の大きさの変化は見られなかった。

3.2.2 三角形の幾何拘束条件による違い比較

病変サイズが小さい場合、ステージ 3 による特徴量のマッチングに十分な情報を得られないため、1 辺 30 ピクセルより小さいものは 30 ピクセルに設定した。追跡エラーは式 3.31 のように定義した。

$$\mathbf{E}^t = \sqrt{(x_{et} - x_{at})^2 + (y_{et} - y_{at})^2}, \quad (3.31)$$

ここで、 (x_{et}, y_{et}) はフレーム t における推定した追跡病変領域の中心位置であり、 (x_{at}, y_{at}) はフレーム t における病変の正解領域の中心位置である。

まずはじめに、アフィン行列のパラメータによる拘束において、幾何拘束条件であるパラメータの設定を変化させた場合の評価を行った。図 3.9 に評価結果を示す。パラメータを変化させてもあまり大きな評価結果の差がないことが見て取れる。

図 3.10 に拘束条件をつけなかった場合と、本論文での提案手法である、角度による拘束、アフィン行列のパラメータによる拘束をつけたものの結果を示す。拘束条件をつけなかった場合とつけた場合ではかなり大きな性能差が出ているのが見て取れる。また、角度による拘束とアフィン行列のパラメータによる拘束では多少ではあるがアフィン行列のパラメータによる拘束の方が精度が良いことが見て取れる。

図 3.11 にアフィン行列のパラメータによる拘束において追跡に成功したシーケンス例を示す。緑色の矩形は正解追跡領域を示し、黄色の矩形は追跡結果を示す。

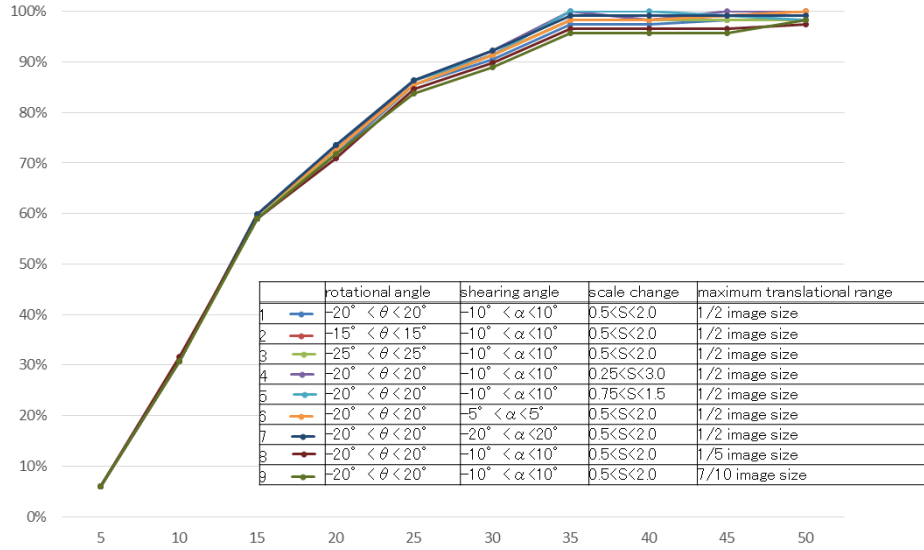


図 3.9: Precision and success plots per sequence. Here the performance of the proposed method is investigated by changing parameters.

これらのシーケンスは不規則な動きがあったにもかかわらず，提案された方法は正確に病変を追跡している．背景変化が大きく，人の目でも追跡がむずかしいシーンにおいても追跡ができているのが分かる．

図 3.12 は，アフィン行列のパラメータによる拘束における追跡失敗例を示す．上段の例は特徴点検出時に各フレームで検出された特徴点の位置が連続するフレーム間で一致するものが少なかったため，サポーターのペア作成時に正確なペアを作ることができず，追跡を失敗した例を示している．下段の例は病変の領域が小さすぎたため，病変自身の特徴が周辺と区別がつかず，ステージ 3 の病変自身の特徴を使った高精度マッチングステージで徐々に離れていったケースを示している．

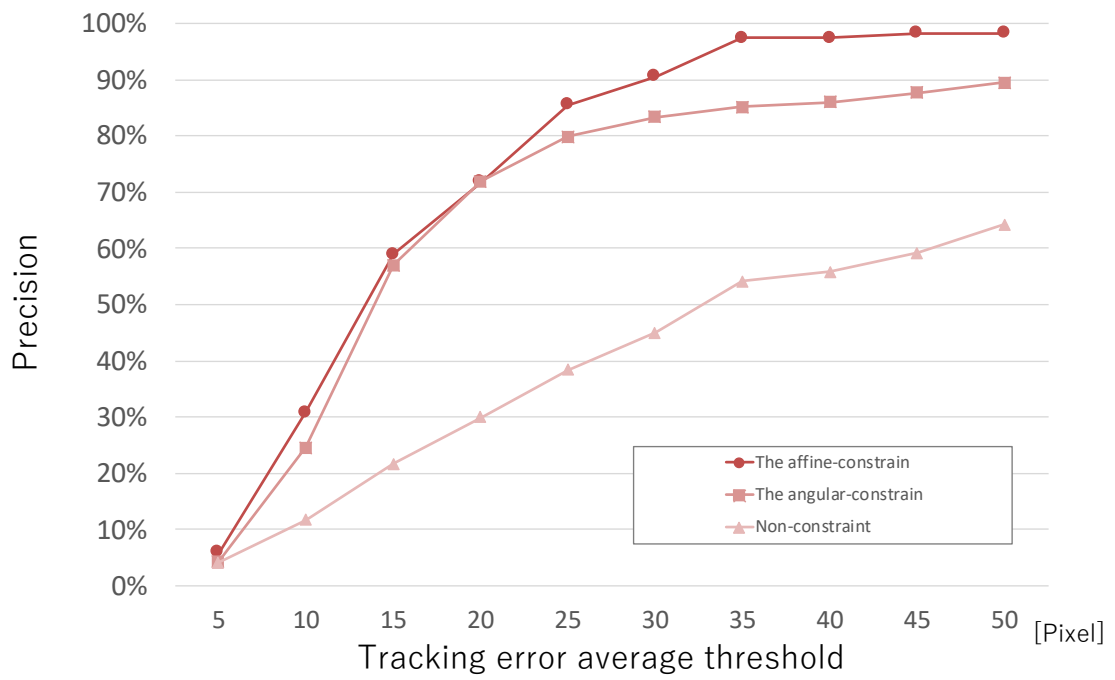


Fig. 3.10: Precision and success plots per sequence. Here the performance of the proposed method is compared with that of the conventional method and angular-constraint method.

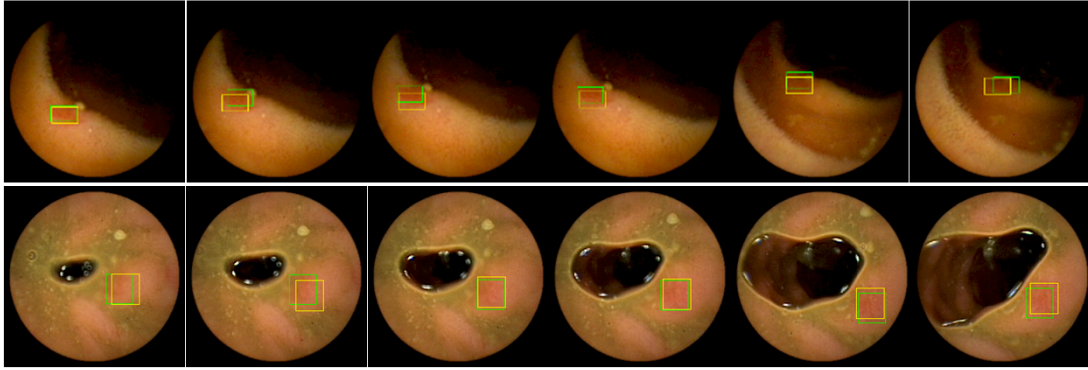


図 3.11: Examples of the accurate tracking of abnormalities using the affine constraint. The green rectangles in each image are the ground truth region, while the yellow rectangles are our tracking result.

3.2.3 従来手法との比較

次に従来研究との比較を行う．図 3.13 は，アフィン行列のパラメータによる拘束と従来手法である mean-shift 法 [1] と 2015 年に Visual Tracking Challenge (VOT2015) で優勝した Multi-Domain Network (MDNet) を用いた方法 [48] についてシーケンス毎の追跡誤差の平均値に対し，閾値以下の場合を正解とする正解率をプロットしたグラフである．MDNet を用いた手法は CNN を用いた手法のため，事前トレーニングを必要とする．そこで，2つのトレーニングモデルを用意した．1つは VOT2015 と同じトレーニングデータを使用し，もう1つはカプセル内視鏡画像を使用した．アフィン行列のパラメータによる拘束を用いた提案手法は，mean-shift 法と VOT2015 と同じトレーニングモデルを用いた MDNet 法よりも優れていた．

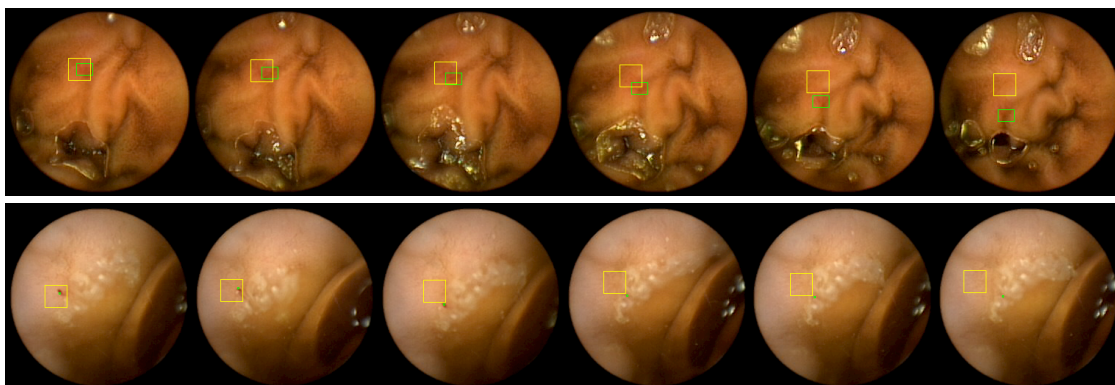


図 3.12: Examples of erroneous tracking using the affine constraint. Green rectangles in each image are the ground truth region, while the yellow rectangles show our tracking result.

図 3.14 は，VOT2015 の評価で用いられている評価指標である各フレーム毎の追跡位置と正解位置の誤差と正解率をプロットしたグラフを示す．こちらのグラフでは比較した 4 つの手法のうち提案手法が最も優れていることがわかる．

3.3 まとめ

本章では，カプセル内視鏡における小腸画像は病変と周辺領域それぞれにおいて追跡において有効な特徴が得られないため，病変とその周辺領域の位置関係を利用した特徴点ベースの追跡手法を提案した．

追跡ターゲットとなる病変領域の大まかな位置は病変周辺の特徴点 (サポーター) の三角形による幾何拘束を用いたマッチングと，アフィン変換による投票によって推定された．三角形による幾何拘束は角度によるものと，アフィン行列のパラメータによるものの 2 つを提案した．

評価実験により，提案手法は，動きの変位が大きく，肉眼では検出が困難な場合でも，病変を追跡可能であることを示した．

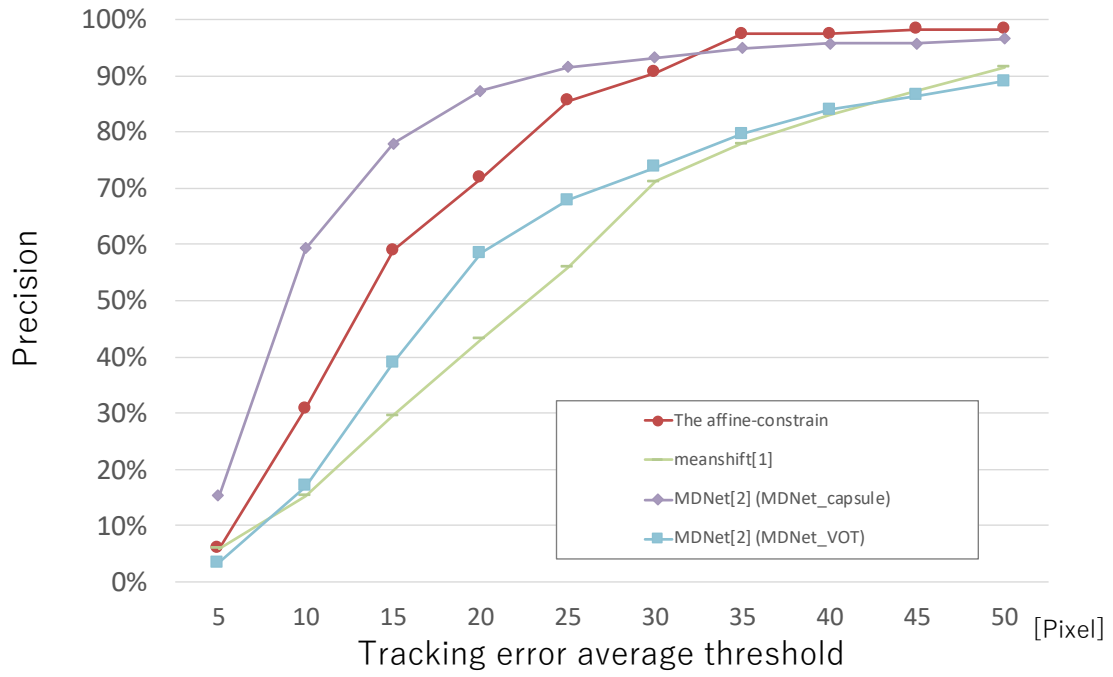


Fig. 3.13: Precision and success plots per sequence. Here the performance of the proposed method is compared with that of the mean-shifted method [1] and that of the MDNet method [2] trained using VOT2015 data or capsule endoscopy images.

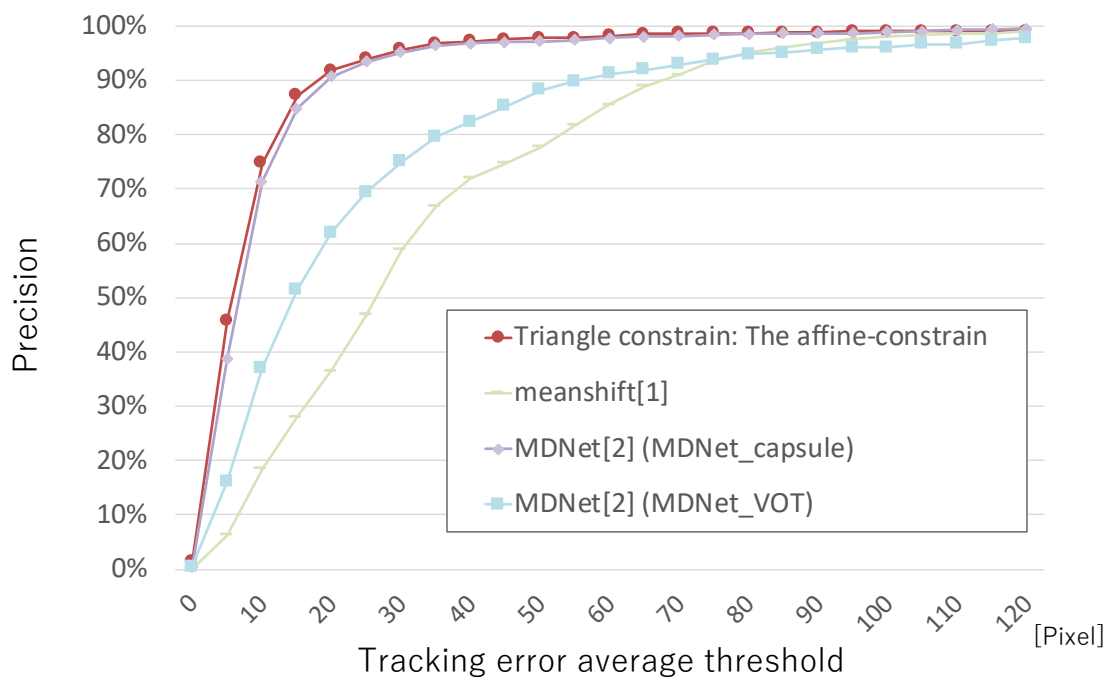


图 3.14: Precision and success plots per frame, comparing the tracking of our proposed method with that of other state-of-the-art methods.

第4章 Convolutional Neural Networksを用いたイントラフレーム学習による病変領域追跡

本章では Convolutional Neural Networks(CNN) を用いた病変領域の検出による追跡手法を提案する。前章までで述べた通り，フレームレートが低く，動作モデルの予測が難しいカプセル内視鏡画像の病変の追跡には有効な特徴量を定義すること困難である。そのため，追跡対象ごとにロバストな特徴量を学習から自動的に定義できる CNN を用いることが有効と考えられる。しかし，従来の CNN を用いた物体追跡手法はあらかじめ大量の学習データが必要であるという大きな課題がある。これまで提案されている CNN を用いた追跡手法の多くは大量の正解ラベルが付与された対象画像と類似したデータを用いた事前学習が前提になっている。本章では類似データを用いた事前学習を行わず，追跡開始時に与えられる開始フレームと追跡ターゲットである病変の初期位置のみを用いて CNN の学習に必要なデータを構築することで，事前に大量の学習データを必要としないイントラフレーム学習による追跡手法を提案する。本論文では本章で提案する追跡手法を CNN を用いた追跡手法と呼ぶ。

4.1 病変と周辺領域のイントラフレーム学習による病変領域追跡

本章で新たに提案する内容は以下の通りである．

- モデルの学習に用いるデータは追跡する病変の開始フレームと初期位置のみ
- 病変画像と同サイズに切り出した 矩形画像のペアを入力データとし，一致度を出力する CNN を構築
- 学習時に用いる一致度の正解ラベルは開始フレームにおける切り出し矩形領域と病変領域の矩形の重なり率 ($Score$)
- 追跡結果をもとにしたオンライン学習

まずはじめに，学習時は開始フレーム (f -th frame) の病変の初期位置と開始フレームを切り出した矩形画像 i を入力データとして，重なり率 ($Score_i$) を出力するネットワークの重みを学習する．次に，追跡時には学習したネットワークの重みを用い開始フレーム (f -th frame) の病変画像 (query 画像) と追跡フレーム ($(f+t)$ -th frame) の切り出し矩形 l を入力とし $Score_l$ を算出する．各追跡フレーム中で $Score_l$ が最大となる位置が検出位置となり，検出位置と追跡フレームを用いて追加学習を行う．各ステップの詳細を次節以降にて順に説明する．

4.1.1 イントラフレーム学習のための学習データの選び方と正解値ラベルの付与

本節では学習データとして用いる入力データと正解値ラベルの $Score_i$ について説明する．本論文では，学習データを追跡開始時に与えられた開始フレーム (f -th frame) の RGB 画像とその中に存在する病変領域の位置から生成する．図 4.1 に学

習データと正解ラベルの説明図を示す．まずはじめに，開始フレーム (f -th frame) の RGB 画像と病変領域が与えられると病変領域の外接矩形 (図 4.1(1) カプセル内視鏡画像中の青枠) を設定し，切り出す．これを query 画像とする．次に開始フレーム (f -th frame) の座標原点である左上からラスタスキャンし query 画像と同サイズで画像を順次切り出す (図 4.1(2))．これを probe 画像とする．query 画像と probe 画像はそれぞれ RGB の 3 チェンネルからなる．1 枚の開始フレームから query 画像は 1 枚，probe 画像は入力画像サイズに応じて複数枚得られる．例えば入力画像サイズ 256×256 ピクセル，病変領域矩形が 50×50 ピクセルで 1 ピクセルごとにラスタスキャンを行った場合，42436 枚 (206×206 枚) の probe 画像が得られることになる．つまり，query 画像は固定で probe 画像が変わることによって複数の画像ペアができ，これが入力データとなる．この入力データの画像ペアに対して正解ラベルを与えることになるが，従来研究ではこの正解ラベルに一致しない (0) もしくは一致する (1) の 2 値のラベルを用いていた．そのため，1 枚の開始フレームからは一致する (1) に相当するサンプル画像を十分に得ることができないため，それだけでは十分な学習ができず，多数の類似する画像を用いた事前学習が必要となっていた．本論文では正解ラベルとして式 4.1 に示す，query 画像の切り出し矩形領域と probe 画像の切り出し矩形領域 i の重なり率 $Score_i$ を用いる．

$$Score_i = (O_{Y_i} \times O_{X_i}) / (H_q \times W_q) \quad (4.1)$$

$$O_{Y_i} = \max(0, (\min(y_{q2}, y_{pi2}) - \max(y_{q1}, y_{pi1})))$$

$$O_{X_i} = \max(0, (\min(x_{q2}, x_{pi2}) - \max(x_{q1}, x_{pi1})))$$

$$H_q = y_{q2} - y_{q1}$$

$$W_q = x_{q2} - x_{q1}$$

座標原点は各フレーム画像の左上とし、 $x_{q1}, x_{q2}, y_{q1}, y_{q2}$ は query 画像を切り出す際の座標位置、 $x_{pi1}, x_{pi2}, y_{pi1}, y_{pi2}$ は probe 画像 i を切り出す際の座標位置とする (図 4.1 参照). O_{Y_i} は query 画像領域と probe 画像 i 領域との Y 軸方向での重なる距離. O_{X_i} は query 画像領域と probe 画像 i 領域との X 軸方向での重なる距離. H_q, W_q はそれぞれ Y 軸方向, X 軸方向の query 画像のサイズとなる. これにより, 1 枚の開始フレームと初期位置から異なるラベルをもった多くの学習データを得ることができる.

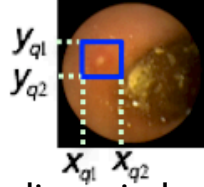
正解ラベルに一致しない (0) もしくは一致する (1) の 2 値のラベルを用いた場合と提案する重なり率を正解ラベルに用いる場合で開始フレームと初期位置のみから学習データを作成する場合に学習データの数と比較する. 例えば入力画像サイズ 256×256 ピクセル, 病変領域矩形が 50×50 ピクセルで 1 ピクセルごとにラスタスキャンを行った場合, 42436 組 (206×206 枚) の画像ペアが得られることになる. 2 値のラベルの場合は厳密に一致するものは一つであるため, 一致する (1) ラベルを付与されるものが 1 枚となるが, 周囲複数ピクセル分を一致する (1) ラベルを付与するのが一般的である. そこで, 周囲プラスマイナス 10 ピクセル分を一致するとラベルを付与するとした場合, 441 枚に一致する (1) ラベルが付与され, 41995 枚に一致しない (0) ラベルが付与されることとなり, ラベルによる枚数の差が大きくなり, 学習データとして成立しない, 一方, 提案手法は病変の位置によるが, 約 2 万枚が 0 より 1 の様々な数値をラベルとして付与され, 残りの約 2 万枚が 0 の値をラベルとして付与されることになり, バランスのよい学習データが得られる.

4.1.2 CNN のネットワーク構造

本論文で提案する CNN のネットワーク構造を図 4.2 に示す. 入力 is query 画像と probe 画像の 2 枚の RGB 画像を重ね合わせた 6 チャンネルのデータで, 畳み込

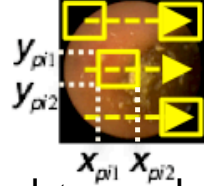
(1). Input starting frame and abnormality area position.

f-th frame

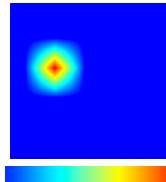
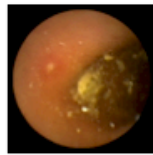


(2). Sampling windows

f-th frame



(3). Calculate overlap ratios (Eq. 1)



0.0 0.5 1.0
Score map

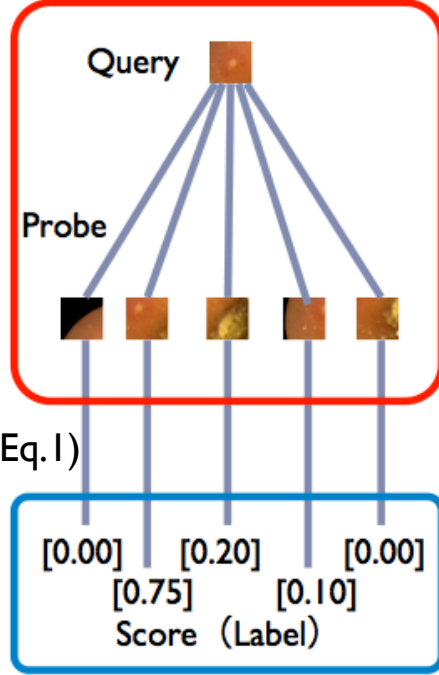


図 4.1: A method of creating data sets and training process of the network.

み層と subsampling 層のセットを 3 層，および全結合層 3 層からなる．出力値は 1 つで入力した 2 枚の画像の一致度に当たる $Score_i$ を出力する．入力に用いる 2 枚の画像は 2 枚を単純に重ね合わせた $R_q B_q G_q R_{pi} B_{pi} G_{pi}$ の 6 チャンネルの A．画像順並べ (In order of image) と色ごとに隣になるように重ね合わせた $R_q R_{pi} B_q B_{pi} G_q G_{pi}$ の 6 チャンネルの B．色相順並べ (In order of hue) の 2 パターンを検討した (図 4.3 参照)．query 画像と probe 画像を重ねて入力とすることで，それぞれ別々に畳み込み層を通してから重ねるよりも，直接 probe 画像と query 画像の違いを捉えることができる．また，ネットワークが単純になり，学習がしやすくなるとい

うメリットがある.

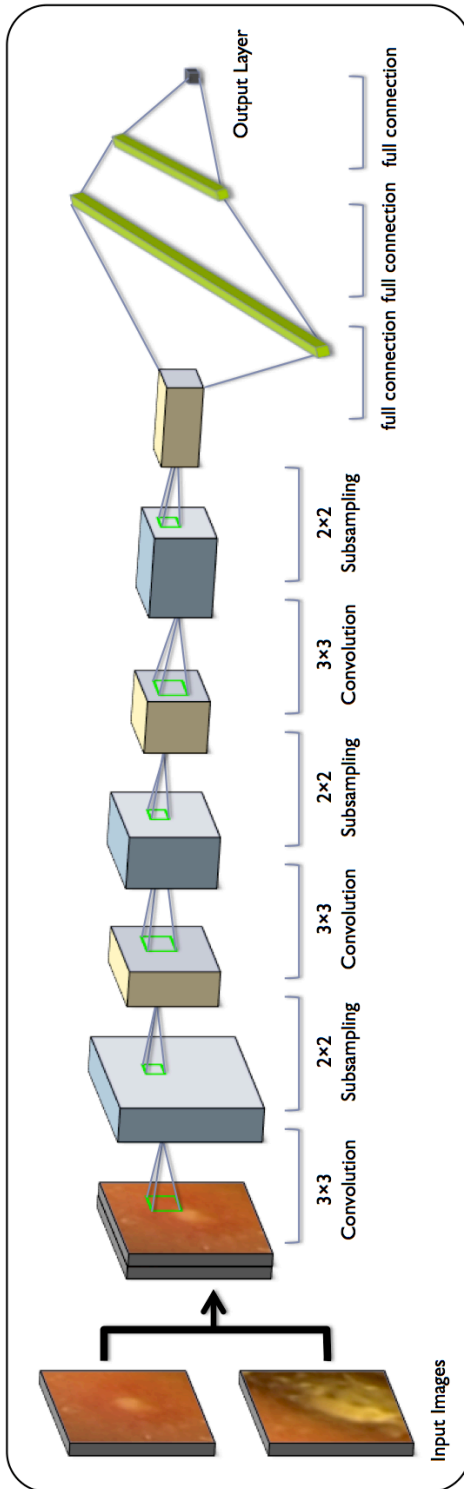


Fig 4.2: Architecture of CNN.

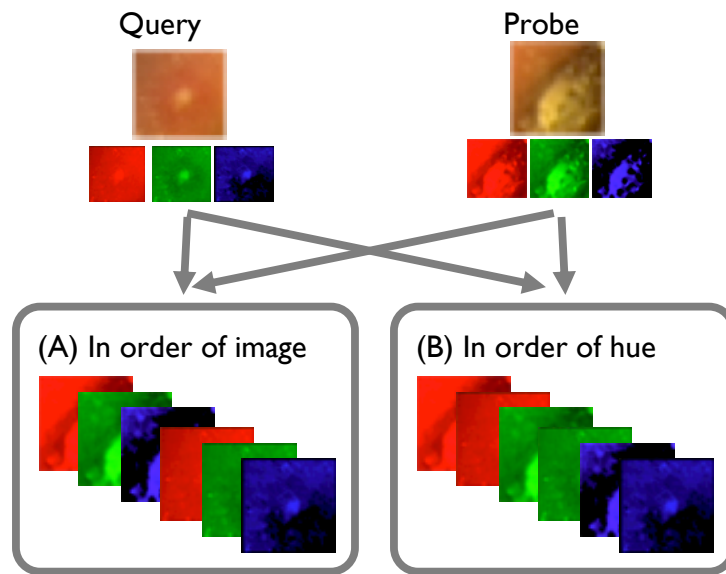
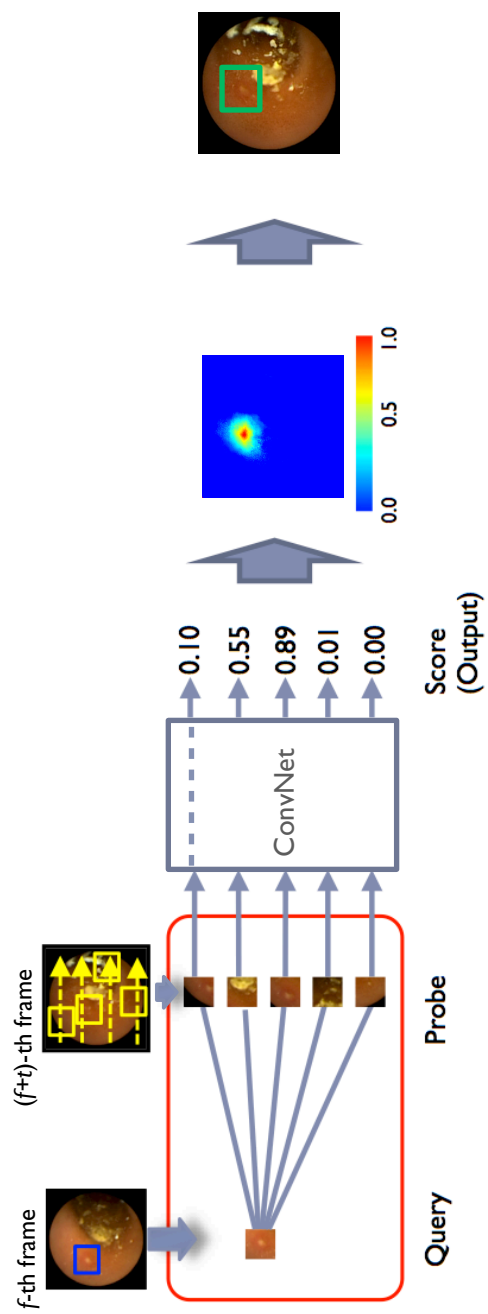


图 4.3: An order of inputting color channels both query and probe images.



(1). Scanning evaluation image (2). Create score map (3). Decision tracking position (Max Score Position)

Fig 4.4: Tracking process on each frames.

4.1.3 病変追跡処理における追跡位置の決定方法

次に追跡処理時の処理の流れを説明する．図 4.4 に追跡時の追跡位置の決定方法を説明した図を示す．追跡時には開始フレーム (f -th frame) の病変画像を query 画像とし，追跡フレーム $((f + t)$ -th frame) から query 画像と同サイズで切り出した領域を probe 画像 i として入力データを作成し，学習したネットワークを通し $Score_i$ を算出する．probe 画像を追跡フレーム $((f + t)$ -th frame) の左上から順番に切り出して同様の処理を行い，各切り出し位置 l での $Score_l$ を計算する．追跡フレーム中の全点の $Score$ を算出し，Score map を作成し，座標原点を左上としたときの最大値を取る画素位置 (x_t, y_t) を算出し追跡位置とする．座標原点はすべてのフレームで左上固定であり，検出結果によって変化することはない．小腸という閉鎖空間での撮影のため，カプセル内視鏡と病変の距離は大きく変わることがない．そのため，追跡病変のサイズは大きく変化しない．本提案手法では追跡物体のサイズは変化しないものとして，初期病変サイズをそのまま適用する．

4.1.4 オンライン学習

カプセル内視鏡画像は画像中心部が明るく，端に行くほど暗くなるという特徴がある．また，病変に対してカプセルの角度が変化することによって明るさが変化する．このような変化に対応するために，追跡結果を用いてオンライン学習を行う．オンライン学習は前節で説明した追跡結果領域を query 画像とし，追跡に用いたフレーム $((f + t)$ -th frame) を左上から順番に query 画像と同サイズで切り出したものを probe 画像として開始時と同様に学習データを生成する．正解ラベルとなる $Score_i$ は追跡フレーム $((f + t)$ -th frame) における probe 画像位置 (追跡結果) と query 画像の切り出し位置の重なり率 (式 4.1) を用いる．追跡に用いたモデルに対して追加学習を行い，ネットワークの重みを更新する．次フレーム $((f + t + 1)$ -th

frame) の追跡処理の時にはこの新たに学習した重みのネットワークを用いる．なお，追跡時の query 画像は初期の病変画像のままで更新しない．

4.2 評価実験

評価用画像として前章と同じものを用いた．

4.2.1 提案手法の評価

本節では提案手法の query 画像，probe 画像の並べ方，オンライン学習の有無の差についての評価を行った結果を報告する．CNN のネットワーク構造は図 4.2 の通りで，各層の構成を表 4.1 に示す．ロス関数は mean squared error を用いた．初期学習時の繰り返し回数は 50 回，オンライン学習時の追加学習の繰り返し回数は 2 回である．実装には Deep Learning のフレームワークである chainer[75] を用いた．入力の病変サイズは様々だが，今回は 50×50 ピクセルに正規化してから入力した．

表 4.1: Parameter of CNN.

ID	出力 Map size	出力 チャンネル数	window サイズ	正規化手法	活性化 関数	subsampling 手法	その他
Input	50×50	6	-	-	-	-	-
畳み込み層 1	48×48	32	3×3	バッチ正規化	relu	-	-
subsampling 層 1	24×24	32	2×2	-	-	max pooling	-
畳み込み層 2	24×24	64	3×3	バッチ正規化	relu	-	padding=1
subsampling 層 2	12×12	64	2×2	-	-	max pooling	-
畳み込み層 3	12×12	64	3×3	バッチ正規化	relu	-	padding=1
subsampling 層 3	6×6	64	2×2	-	-	max pooling	-
全結合層 1	1280	-	-	-	relu	-	-
全結合層 2	500	-	-	-	-	-	-
全結合層 3	1	-	-	-	-	-	-

追跡矩形の中心位置と正解領域の中心位置の位置ずれを追跡誤差と定義し，シーケンス毎の追跡誤差の平均値に対し，閾値以下の場合に正解とする正解率をプロッ

トしたグラフを図 4.5 に示す。正解率の母数は総シーケンス数 (120) で、横軸は追跡誤差のピクセル距離である。オンライン学習の有無の比較ではオンライン学習を行った方が精度が良く、query 画像, probe 画像の並べ方による差は、同じ色相を重ねる方 (B. 色相順並べ) が良いことが読み取れる。また、85% 以上のシーケンスで、追跡誤差が 20 ピクセル以内で追跡できている。また、閾値 20pixel の時のオンライン学習有りなしのフィッシャーの正確確率検定の生起確率は $2.3e-6$ となり、有意水準 5% とすると有意差があると言える。処理速度は Intel(R) Core(TM) i7-6850K CPU @ 3.60GHz 12Core NVIDIA GeForce GTX 1080 搭載マシンにおいて、初期学習に約 5 分、追跡時の 1 フレームあたり画像を読み込んでから追跡位置決定までに約 4 秒、オンライン学習に約 15 秒を要した。

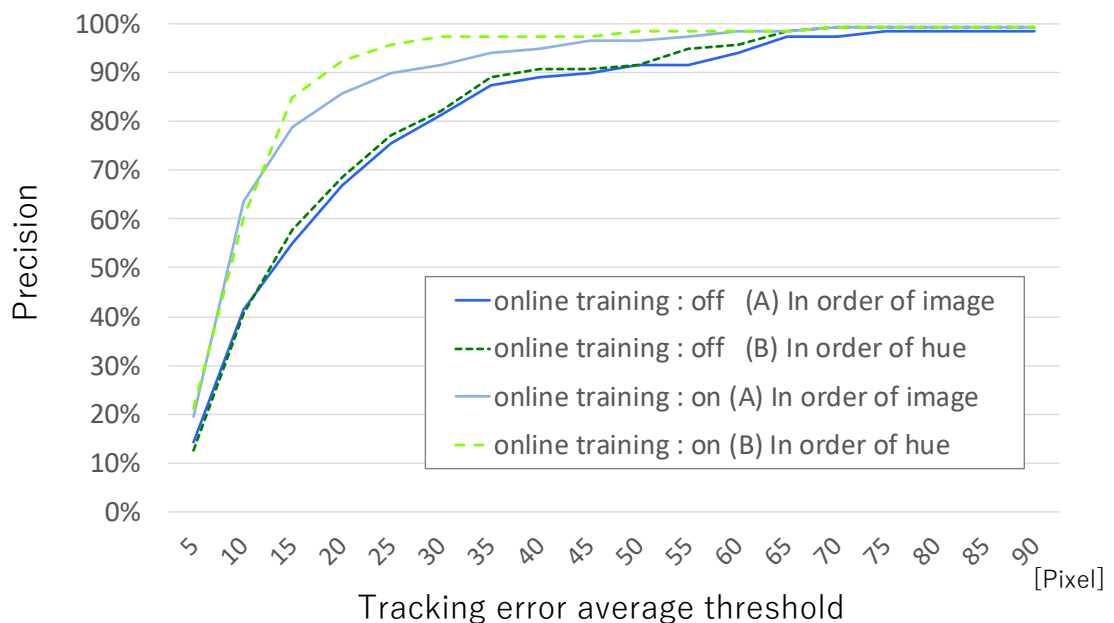


図 4.5: Comparison between online training on and off on relationship between precision and mean value of tracking errors on image sequences.

図 4.6 には VOT2015 の評価で用いられている評価指標である各フレームごとの

追跡位置と正解位置の誤差と正解率をプロットしたグラフを示す．こちらのグラフからもオンライン学習の有無ではオンライン学習有りの方が，query 画像，probe 画像の並べ方による差はこちらも同じ色相を重ねる方 (B. 色相順並べ) がよいことがわかる．query 画像，probe 画像の並べ方による優位は色相順並べの方が良かったため，この後の評価には色相順並べの方を用いることとする．

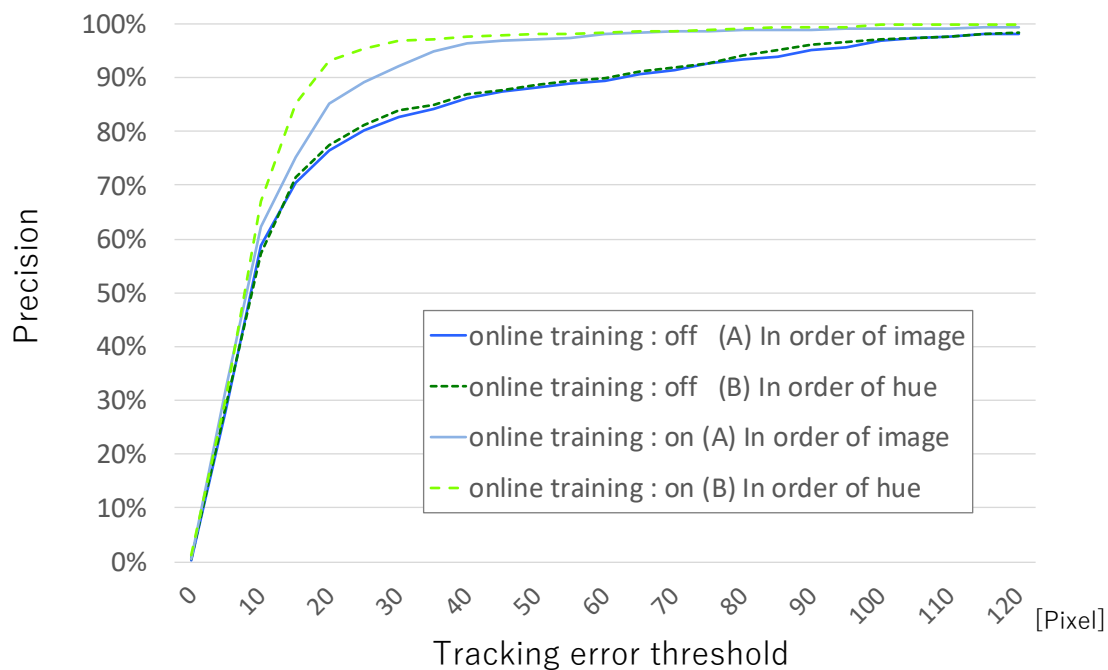


図 4.6: Comparison between online training on and off on relationship between precision and tracking errors on each frame.

図 4.7 にオンライン学習有り，病変ごとのシーケンス毎の追跡誤差に対する正解率をプロットしたグラフを示す．erythematous, erosion の精度が良く，red spot, phlebotasia の精度が悪かった．red spot の精度が悪いのは，病変領域が小さいことに加えて，小腸の壁の動きによって襞の中に病変が隠れ，見えが変わるためと考えられる．phlebotasia の精度が悪いのは青みを帯びた病変であり他の色の病変に比べて背景との差が明瞭でなかったと考えられる．図 4.8 にオンライン学習な

し，病変ごとのシーケンス毎の追跡誤差に対する正解率をプロットしたグラフを示す．phlebectasia 以外の病変に関してはオンライン学習有りの方の精度が良いのが読み取れる．phlebectasia に関しては表 3.1 で述べた通り静脈が浮き出た病変であり，少しのカメラ位置の変化で色味が大きく変わる．そのため低フレームレートの本評価環境ではオンライン学習の効果が出なかったと考えられる．

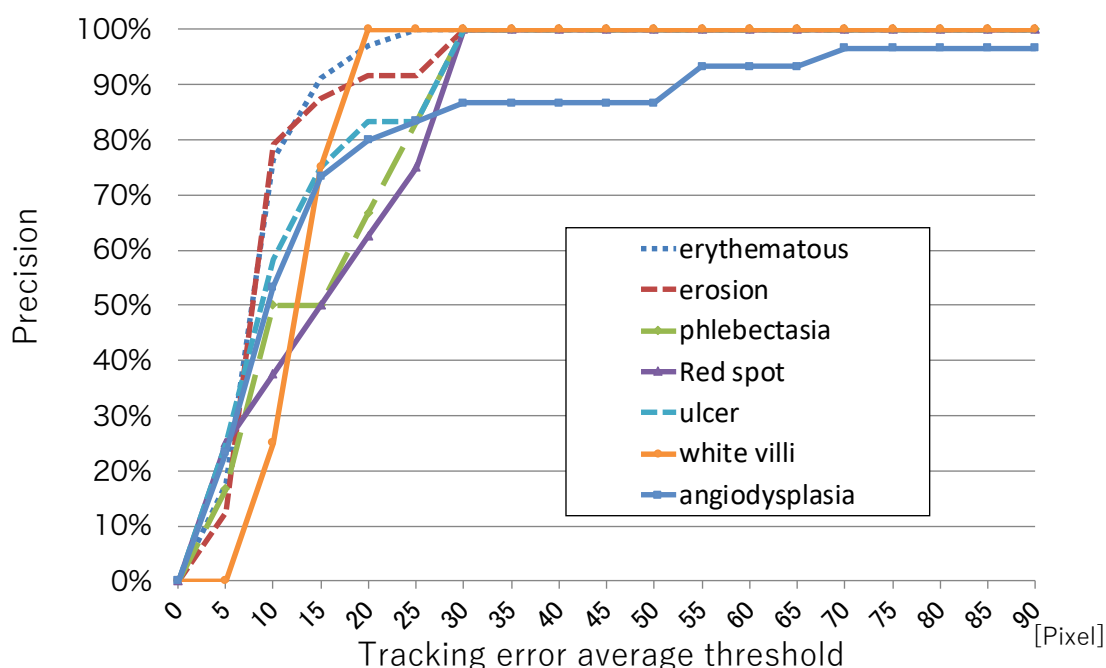


図 4.7: Comparison of several abnormalities with online training on relationship between precision and mean value of tracking errors on image sequences.

図 4.9 にオンライン学習有りでの追跡成功例を示す．青枠が正解病変位置，緑枠は提案手法が出力した追跡結果位置である．各カプセル内視鏡画像の下画像は出力された score map である．上段の結果から，背景が大きく変化しても病変位置の *Score* が高く，正しく追跡できていた．また，下段の結果からは正常領域との差が明確でない病変でも病変位置と背景にスコアの差が明確にあり，追跡できていた．上段，下段ともに病変は erythematous である．本論文では，正解矩形と

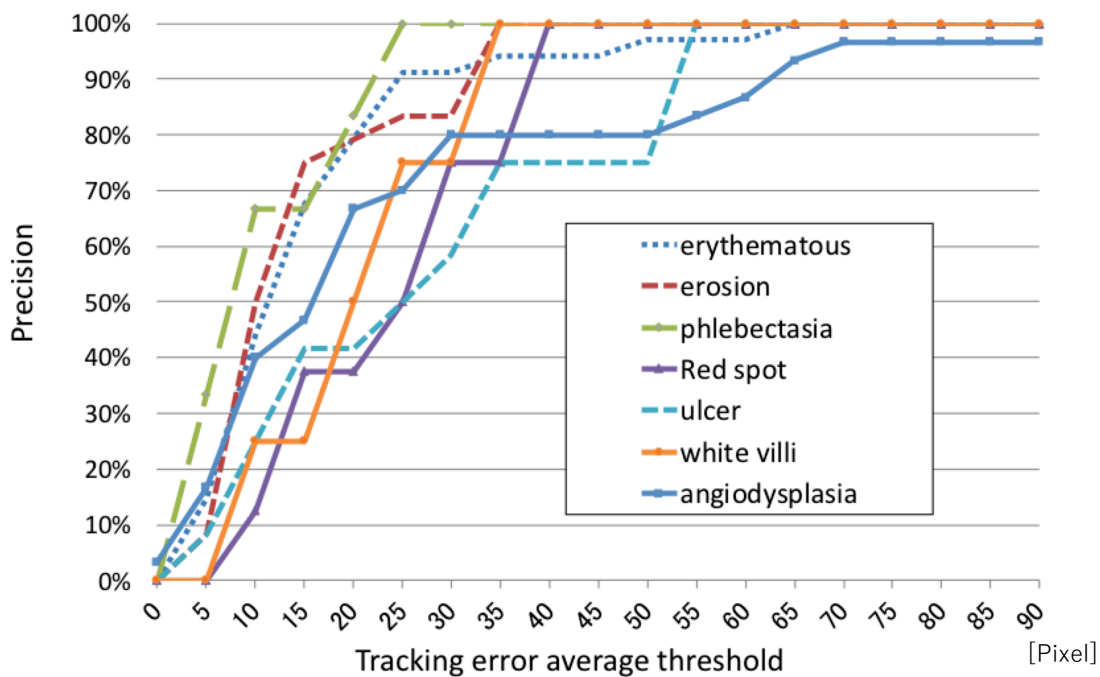


図 4.8: Comparison of several abnormalities with online training on relationship between precision and mean value of tracking errors on image sequences.

追跡矩形の中心誤差が 20 ピクセル以下の場合を成功と定義する。

図 4.10 にオンライン学習有りでの追跡失敗例を示す。上段は追跡領域の明るさが大きく変わってしまったために追跡を失敗した例である。3 フレーム目で一度ずれた追跡領域が 4 フレームでは戻っている。これは本論文では query 画像の更新を行っていないため持ち直した例である。しかし、追跡に失敗した 6 フレーム目から徐々に追跡領域がずれている。これはオンライン学習の結果、ずれた領域が正解とされる学習結果となってしまったためである。中段は背景変化が大きすぎたために失敗した例である。下段は最後のフレームで病変の色が大きく変わっており、追跡に失敗している。それぞれ、score map から追跡ができていないことがわかる。下段は病変名が phlebectasia のものであり、病変と正常領域の色差がはっきりしないため、追跡開始 1 フレーム目ではスコアが低くなっているが、オ

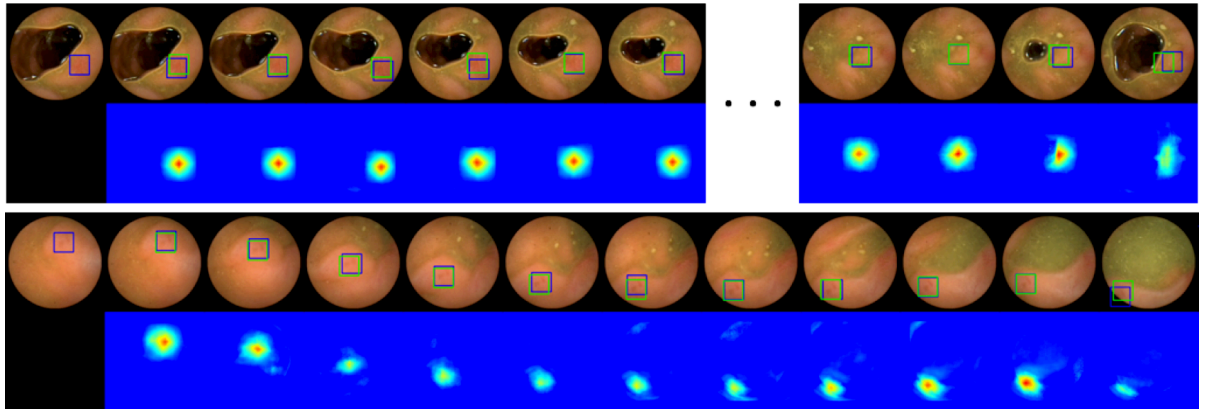


図 4.9: An example of successful tracking (with online training).

オンライン学習を行っているため，フレームが進むにつれて病変領域のスコアが高くなっていることがわかる．上段の病変は ulcer，中段は angiodyplasia，下段は phlebectasia である．

4.2.2 従来手法との比較

本節では提案手法と従来研究との比較実験を行った結果を示す．従来研究は一般的な追跡手法である mean shift 法 [1]，VOT2015 で優勝した CNN を用いたトラッキング手法である MDNet[2]，VOT2016 で相関フィルタを用いた手法で 1 位と 2 位である C-COT[3] と Staple[4] の 4 つを評価した．MDNet は事前学習が必要であることから，事前学習のモデルは VOT2015 で優勝した時の学習セットを用いたもので，著者らが [76] にて公開しているものを用いた評価 (MDNet_VOT) とそのモデルにカプセル内視鏡画像の病変画像を用いて追加学習したモデルを用いた評価 (MDNet_capsule) の二つの評価を行った．カプセル内視鏡画像を用いて追加学習を行った MDNet_capsule の評価は 120 シーケンスを 4 つのグループに分け，各グループ毎に他の 3 グループの評価画像を用いて追加学習を行った．追加学習

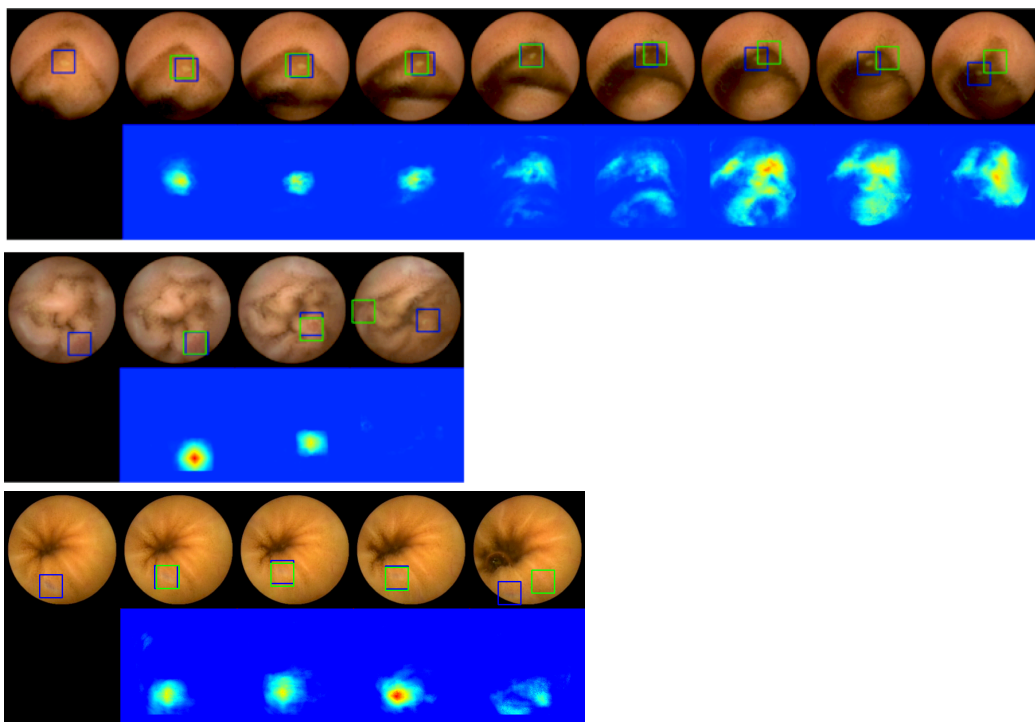


図 4.10: An example of unsuccessful tracking (with online training).

に用いたカプセル内視鏡画像はそれぞれ約 1000 枚である．C-COT はいくつか特徴量が提案されているが，今回は学習の必要ないもので一番性能が良いと報告されている，HOG 特徴量と色特徴量を組み合わせたものを用いた．mean shift 法は Open CV のコード，MDNet は Nan らが公開しているコード [76]，C-COT は著者の Danelljan が公開しているコード [77]，Staple は著者の Bertinetto が公開しているコード [78] を用いた．図 4.11 は提案手法のオンライン学習有りと従来手法 4 つのシーケンス毎の追跡誤差に対する正解率を表したグラフである．図 4.11 から提案手法が最も高い正解率であった．

図 4.12 に提案手法と MDNet，C-COT，Staple の各手法を VOT2015，VOT2016 の評価で用いられている評価指標である各フレーム毎の追跡誤差と正解率をプロットしたグラフを示す．図 4.13 は図 4.12 の一部を拡大したものである．こちらの評価

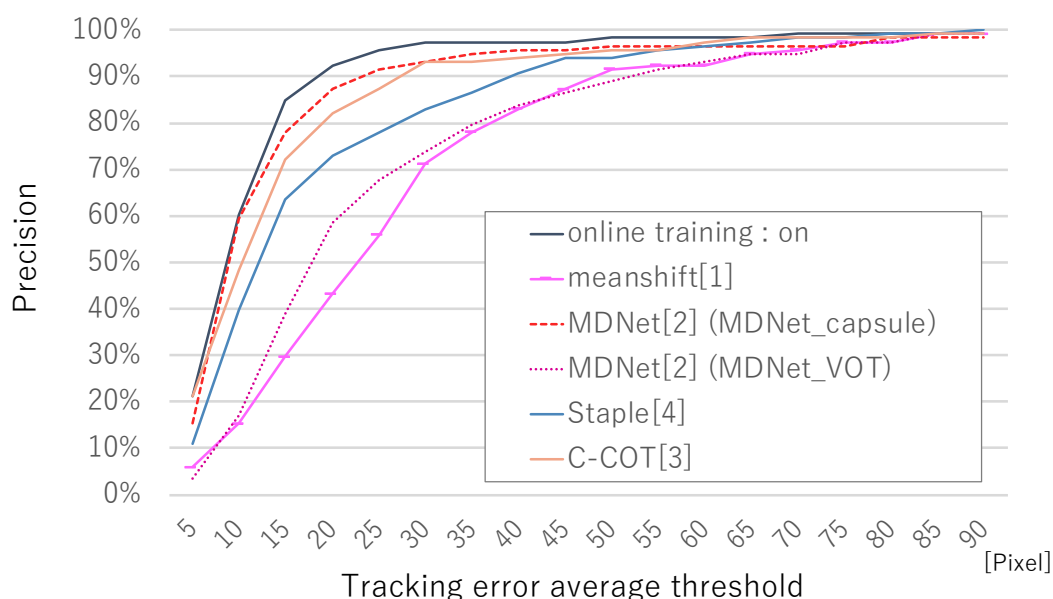


図 4.11: Comparison of the conventional technologies and the proposed method on relationship between precision and mean value of tracking errors on image sequences.

においても提案手法が最も正解率が高く，提案手法の有効性が示せた．また，MDNet はカプセル内視鏡画像の病変画像を用いて事前学習したもの (MDNet_capsule) と一般物体追跡の画像を用いて学習したもの (MDNet_VOT) とで比較を行っているが，カプセル内視鏡画像を用いた事前学習の方が正解率が高く，追跡に想定される画像を事前に学習することが効果的であったと考えられる．一方，本評価における MDNet のカプセル内視鏡画像の病変画像を用いた事前学習で，評価用シーケンス以外に学習画像の準備ができなかったため，学習に用いた画像が約 1000 枚と数が少なく網羅性が確保できていない．そのため，本評価結果は MDNet の性能を十分に引き出せていない可能性がある．それに対し，本論文の提案手法は想定画像による事前学習を必要としないため，開始フレームの位置を教示するだけで容易に利用できる点でも優位性を示すことができた．従来手法の中で一番性能

が良かった MDNet_capsule と提案手法の閾値 20pixel の時のフィッシャーの正確確率検定の生起確率はシーケンス毎の追跡誤差に対する比較で 0.0756 であり，有意水準 5% の場合に有意とは言えないが，フレーム毎の追跡誤差に対する比較では 0.0047 であり，有意差を示せた．

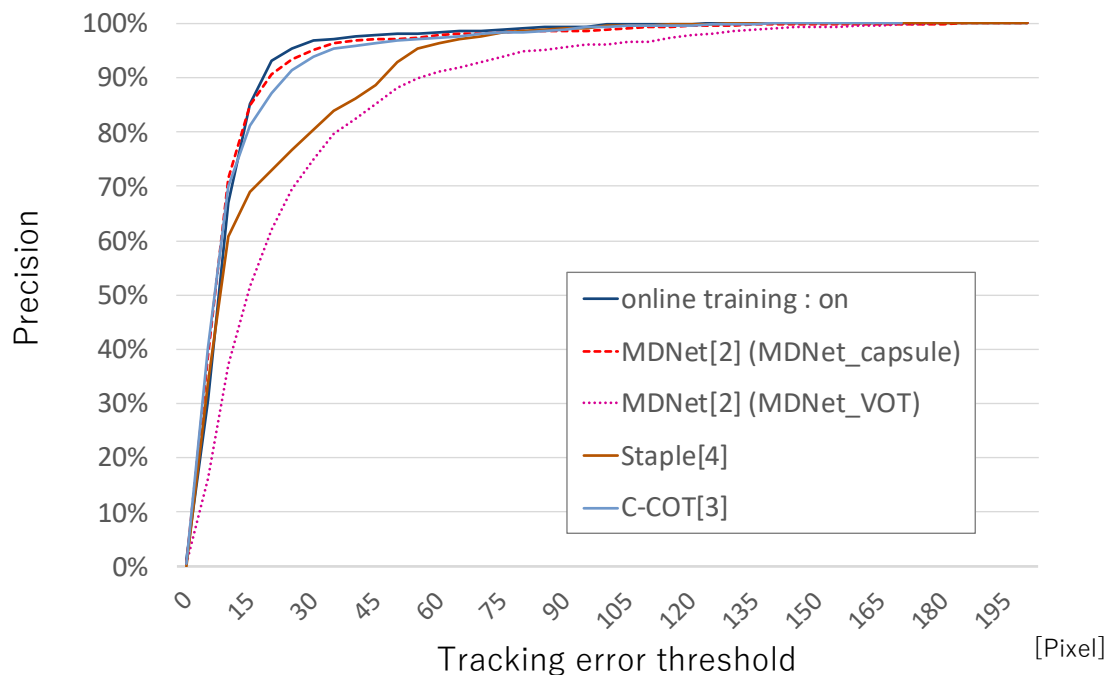


図 4.12: Comparison between MDNet[2] , C-COT[3], Staple[4] and the proposed method on relationship between precision and tracking errors on each frame.

図 4.14 に従来手法で一番精度の良かった MDNet_capsule と提案手法の出力結果の矩形を重畳したものを示す．図 4.9 と図 4.10 であげた成功，失敗例から小腸の変形が大きいものを示した．MDNet は追跡物体の見た目の大きさの変化に対応して矩形サイズを変更した出力を行うが，小腸の変形が大きいシーケンスでも矩形サイズ変化がほとんどないことがわかる．

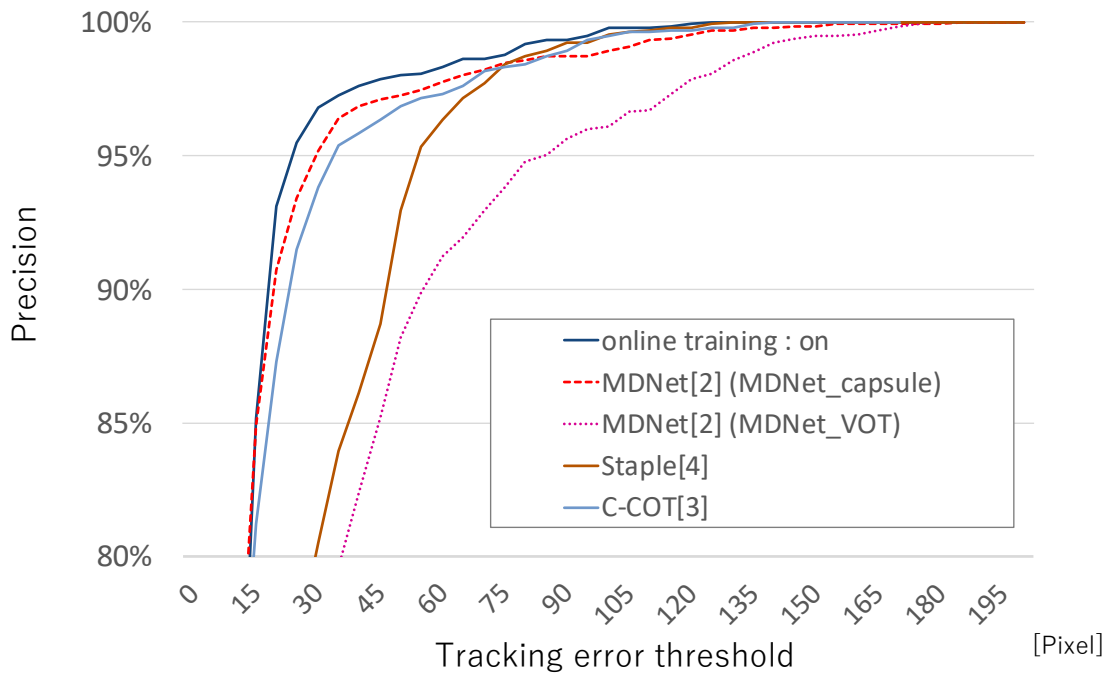


図 4.13: Focusing partial range of Fig. 4.12.

4.2.3 矩形サイズおよび初期位置依存における評価

提案手法は設定する追跡矩形サイズによって、学習に用いる画像枚数が変化する。CNN を用いた手法は、一般的に学習枚数が多くなれば精度が良くなると考えられている。提案手法は、設定矩形サイズが小さいほど学習枚数が多く得られる。一方、矩形サイズが小さすぎると、追跡に適した特徴が得られにくくなると考えられる。前節までは矩形サイズを 50 ピクセルとして評価を行ってきたが、本節では矩形サイズによる精度変化を比較する。4.15 に矩形サイズを変化したときのシーケンス毎の追跡誤差に対する正解率を表したグラフを、4.16 には各フレームごとの追跡誤差に対する正解率を表したグラフを示す。矩形サイズを 10 ピクセルとした時から 30 ピクセルまで順に精度が上がっていき、30 ピクセルをピークにサイズを大きくするほど精度が悪くなった。学習枚数と特徴を得るためのサイズのバラ

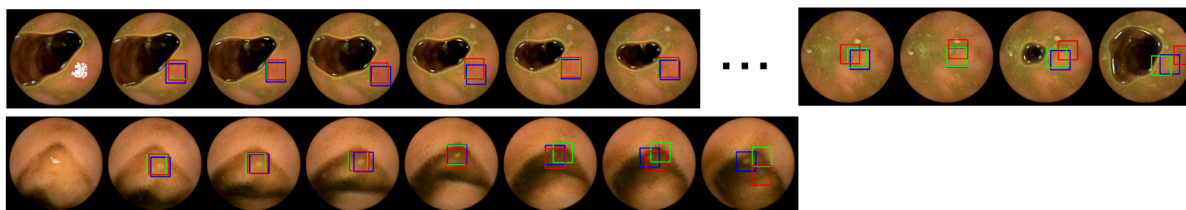


図 4.14: Comparison between MDNet[2] and proposed method on output rectangle.

ンスが良いのは 30 ピクセルであった。

また、提案手法は初期値の与え方を手動で与えることを想定しているので、医師によって誤差があると考えられる。そこで、初期位置をずらした場合の評価を行った。4.17 にずらし量別のシーケンス毎の追跡誤差に対する正解率を表したグラフを、4.18 には各フレームごとの追跡誤差に対する正解率を表したグラフを示す。それぞれ初期位置をずらさなかった場合 (shift:0) と中心位置を方向はランダムに 3 ピクセル (shift:3pixel) から 10 ピクセル (shift:10pixel) ずらした時の評価結果である。このとき病変のサイズは変更していない。3 ピクセルのずれでは精度がほとんど変わらず、5 ピクセルずれから精度が悪くなった。3 ピクセル程度の誤差であれば精度悪化することなく追跡が可能であった。

図 4.19 にオンライン学習有りの時の初期サイズのずらし量のシーケンス毎の追跡誤差に対する正解率を表したグラフを、4.20 には各フレームごとの追跡誤差に対する正解率を表したグラフを示す。初期サイズをずらさなかった場合 (scaling:0) と中心位置はずらさずに、縦か横のどちらかをランダムに 3 ピクセル (scaling:3pixel) から 10 ピクセル (scaling:10pixel) それぞれサイズ変化させた時の評価結果である。初期サイズのずれによる制度の影響はあまり見られない。これは、矩形の 1 辺の最小サイズを 50 ピクセルに設定しており、50 ピクセル以下の病変は 50 ピクセルと設定されるため、サイズのずれが影響しない。また、50 ピクセル以上の病変は 1 辺 50 ピクセルに正規化するため、ずれの影響があまりないためだと考えられる。

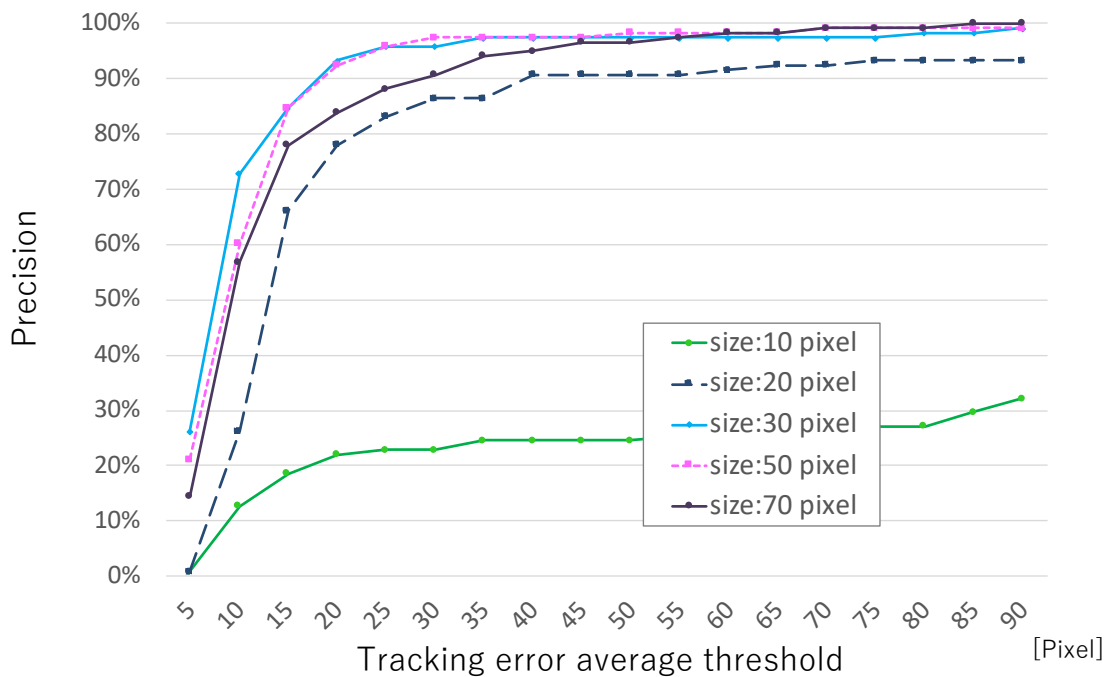


図 4.15: Comparison about sizes of the tracking rectangles on relationship between precision and mean value of tracking errors on image sequences.

人による初期値入力のずれがどれくらいあるかを評価するために、1 枚あたり 10 名に医師の入力を参考に初期値を入力した結果の、中心位置のずれを XY 軸にプロットした結果の楕円誤差の長軸の 2σ は 3.64 であった。この評価結果から 3 ピクセル誤差は許容できる数値といえる。また、この初期値入力ずれの評価結果から評価に用いたデータベースの手動正解値設定においても 3 ピクセル程度の誤差が含まれていることが考えられる。

4.3 まとめ

本章ではカプセル内視鏡画像の病変領域の追跡に CNN を用いた追跡方法を提案した。本手法は従来の CNN を用いた追跡手法の最大の課題である、事前学習に想

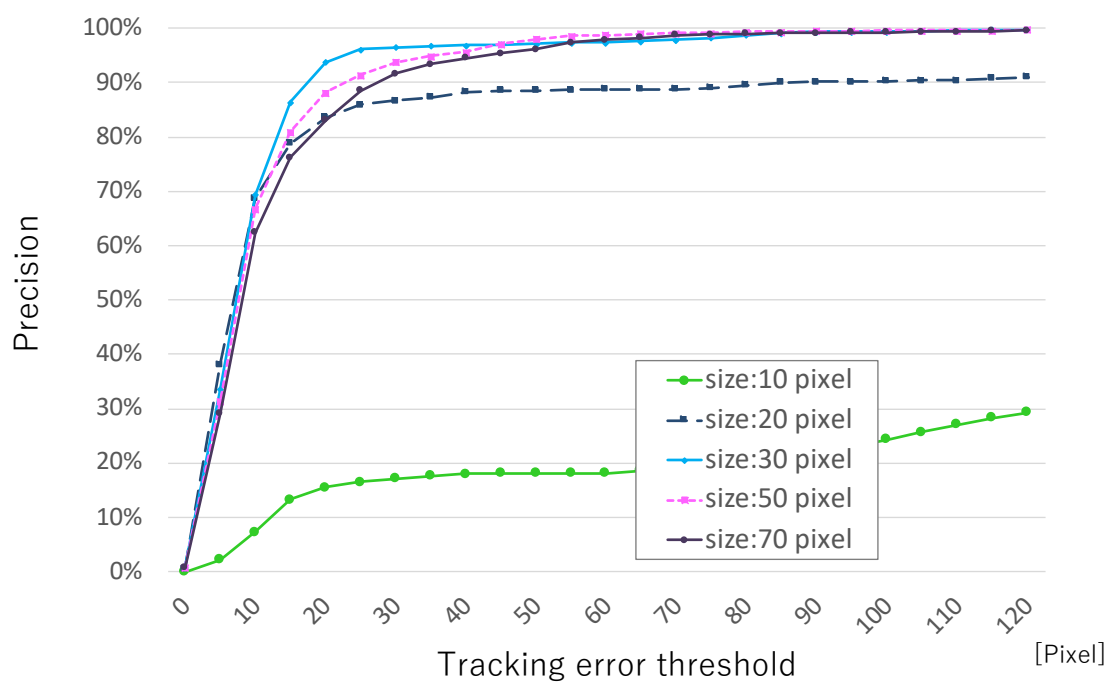


図 4.16: Comparison about sizes of the tracking rectangles on relationship between precision and tracking errors on each frame.

定される追跡画像と同じような画像群と正解ラベルを大量に準備しなければならない課題に対して，正解ラベルを工夫することによって開始フレームと追跡対象の初期位置のみのイントラフレーム学習法を提案した．評価実験によりターゲットのサイズが大きく変わらない低フレームレートで撮影された小腸内視鏡画像の病変追跡においては，本手法の有効性を示すことができた．

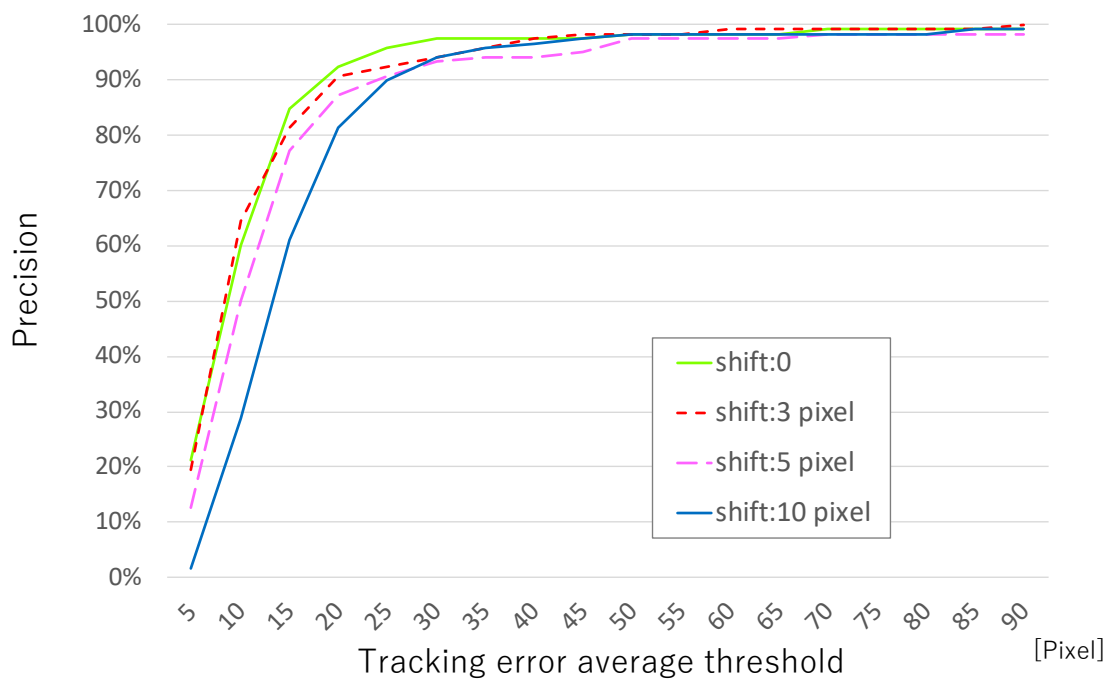


図 4.17: Evaluation for misaligned position in a starting frame on relationship between precision and mean value of tracking errors on image sequences.

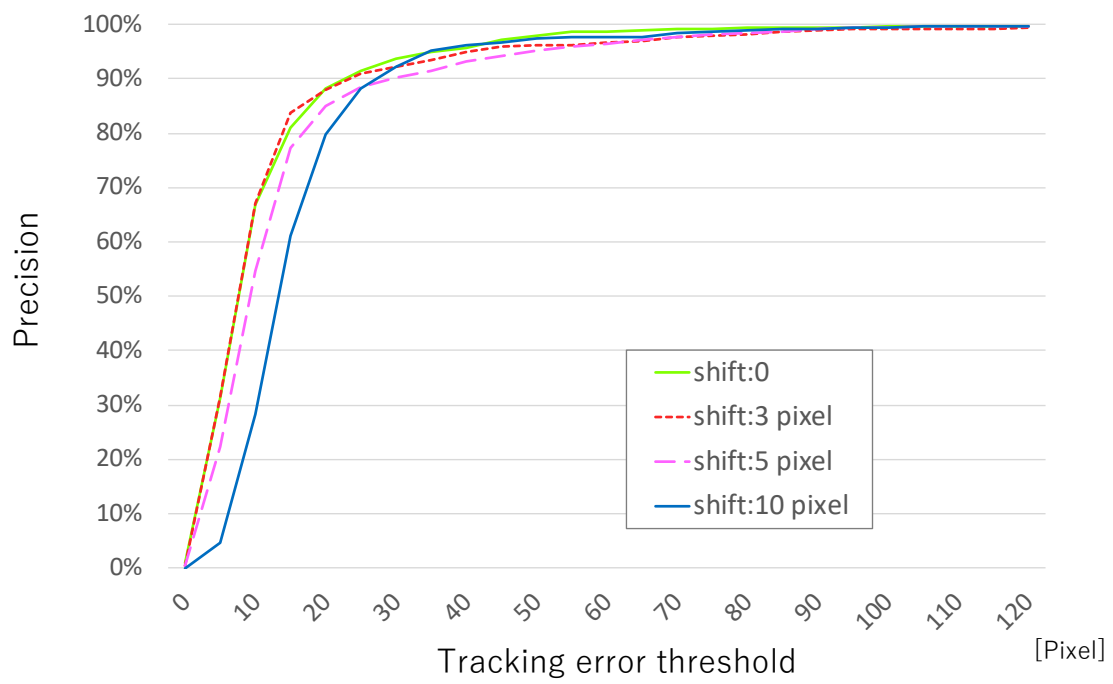


Fig. 4.18: Evaluation for misaligned position in a starting frame on relationship between precision and mean value of tracking errors on image sequences.

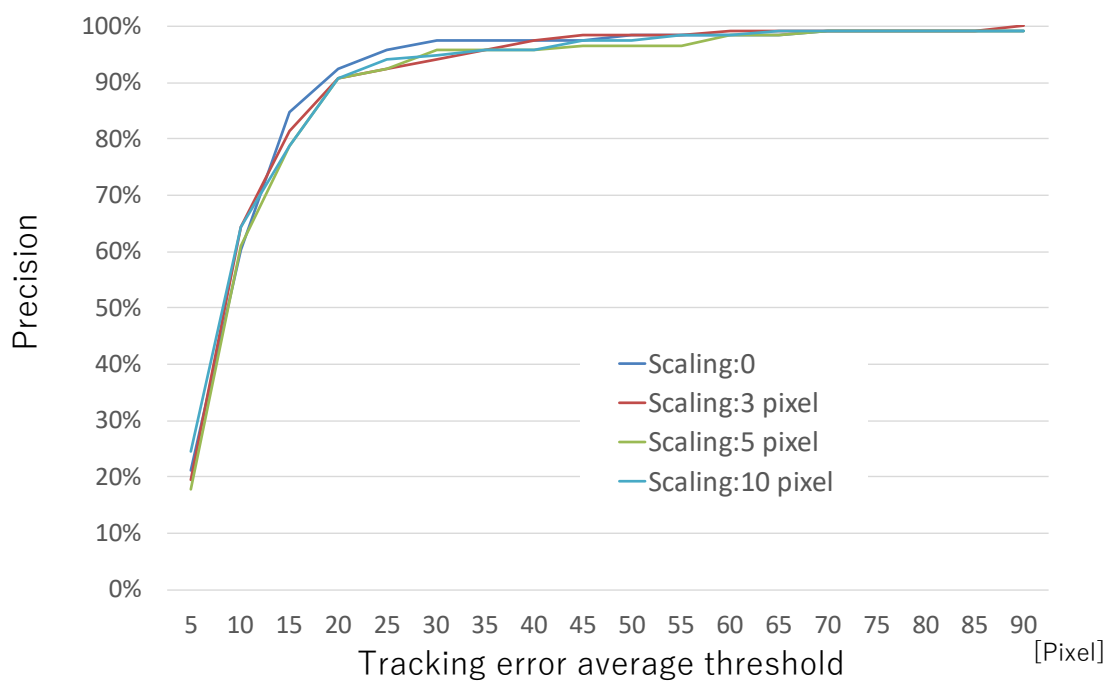


図 4.19: Evaluation for misaligned position in a starting frame on relationship between precision and tracking errors on each frame.

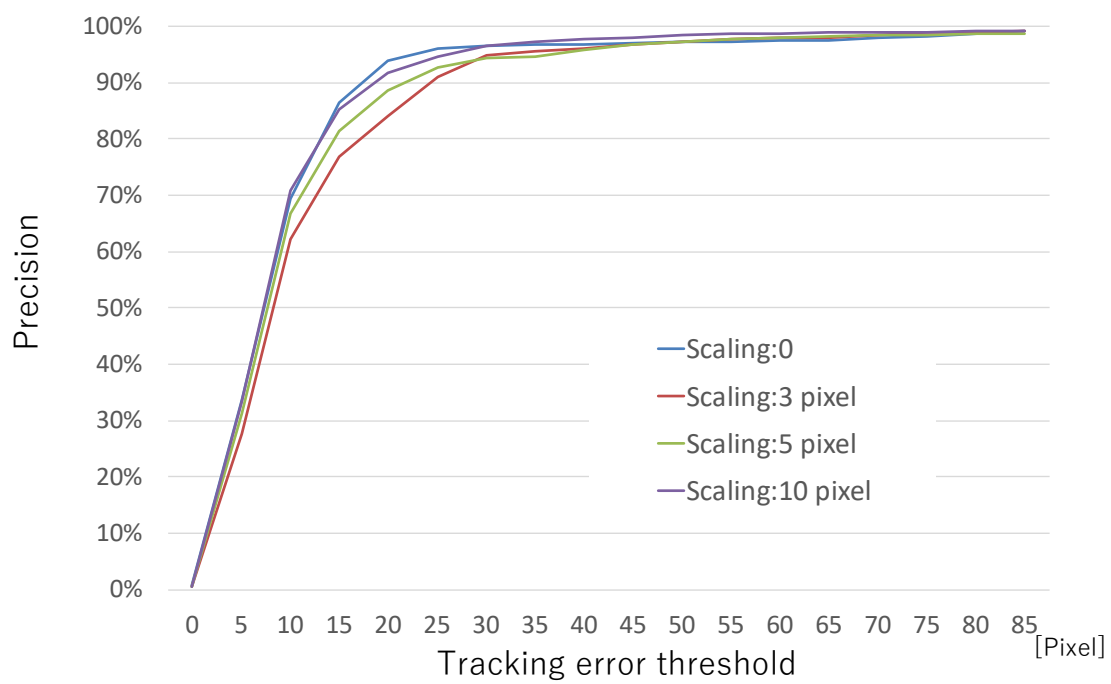


Fig 4.20: Evaluation for misaligned position in a starting frame on relationship between precision and tracking errors on each frame.

第5章 特徴点ベースの追跡手法と病変とCNNを用いた追跡手法の比較

3章で特徴点ベースの追跡手法を，4章でCNNを用いた追跡手法を提案した．本章では，提案した二つの病変追跡手法の比較を行う．

5.1 病変追跡性能の比較

図5.1に特徴点ベースの追跡手法の二つの幾何拘束条件とCNNを用いた追跡手法のオンライン学習のありなりのシーケンス毎の追跡誤差に対する正解率を表したグラフを示す．CNNを用いた追跡手法のオンライン学習ありのものが最も優れていることが読み取れる．一方で，追跡誤差が20ピクセル以内を正解とした時に，特徴点ベースの追跡手法でアフィン行列係数を拘束条件とした場合と，CNNを用いた追跡手法のオンライン学習なしのものを比較した場合，特徴点ベースの追跡手法の方が優れている．これは，CNNを用いた追跡手法のオンライン学習なしの場合，病変領域と周辺領域とを識別する特徴量を初期フレームの病変と周辺情報を用いて学習しており，追跡が進むにつれて病変と周辺の識別に適した特徴量が変化しているにも関わらず，更新していないために対応できていないためである．一方で，特徴点ベースの追跡手法はステージ3の詳細な位置推定では初期フレー

ムでの病変の特徴量を用いているが、周辺特徴点のマッチングには三角形の幾何拘束条件と共に、対象フレームとその1つ前のフレームの特徴量を用いているため、追跡が進んでも対応できているためと考えられる。

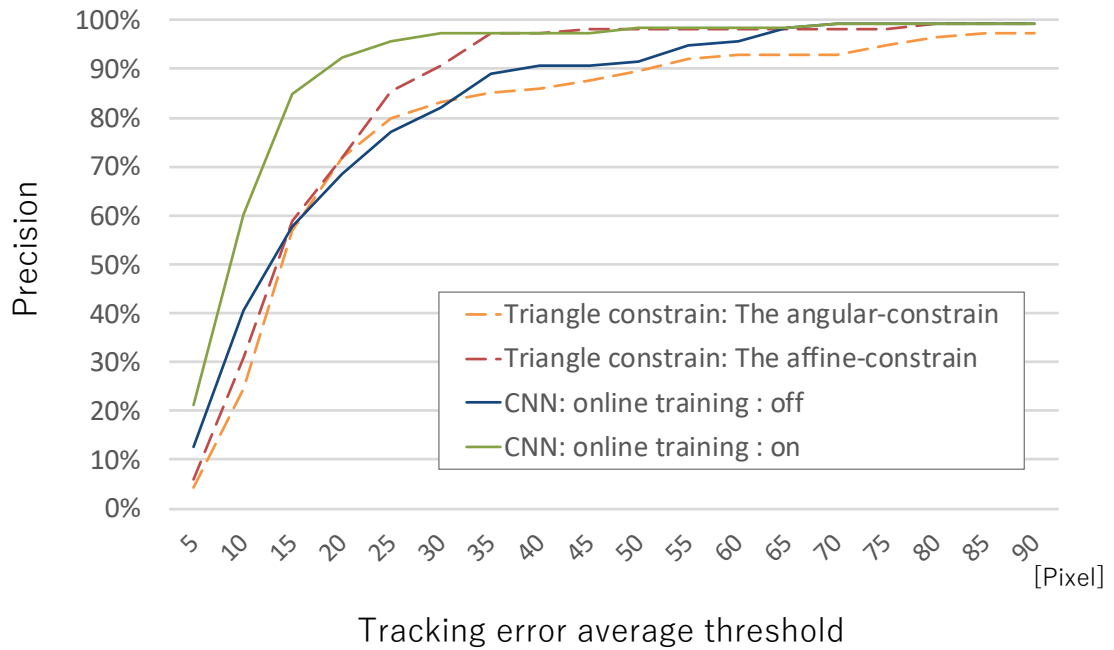


図5.1: Comparison between feature points matching based method and CNN based method on relationship between precision and tracking errors on each frame.

図5.2, 図5.3, 図5.4は特徴点ベースの追跡手法でアフィン行列係数を拘束条件とした場合の結果とCNNを用いた追跡手法のオンライン学習なしの場合の結果を示す。緑枠が正解の病変位置, 黄色枠が特徴点ベースの追跡手法, ピンクの枠がCNNを用いた追跡手法の結果である。

図5.2は特徴点ベースの追跡手法の方がよかった例である。オフライン学習なしのCNNを用いた追跡手法では初期フレームで学習した特徴量を用いて追跡を行っている。そのため、フレームが進むにつれて特徴量が変わることで追跡に失敗する。一番上の段と三段目はフレームが進むにつれてカメラ位置が変わり、病変と

背景の色味が変わって失敗している．二番目の段は小腸の蠕動運動によって病変も背景も特徴が変わり失敗している．一番下の段はエッジが明確な浮遊物の位置に引っ張られて失敗している．一方で，特徴点ベースの追跡手法は特徴点のマッチングを連続する2フレーム間で行うことで特徴の変化にロバストであると共に，画像の全体で位置の最適化をしているため，一番下の段のように浮遊物のように局所的なエッジに惑わされない．しかし，下から二段目に見られるように，CNNを用いた追跡手法のオンライン学習なしの場合は過去の移動情報を追跡に用いていないため，色味が同じようになると，追跡が元に戻るという特徴もある．CNNを用いた追跡手法はオンライン学習ありにすることによって，徐々に変化する色味の違いには対応できるようになる．しかし，図4.10の失敗例にもあるように連続フレーム間での急激な色味の変化では対応できないことがある．その点，特徴点ベースの追跡手法は画像全体の動きを見ることになるので，色味の変化にはロバストであると言える．

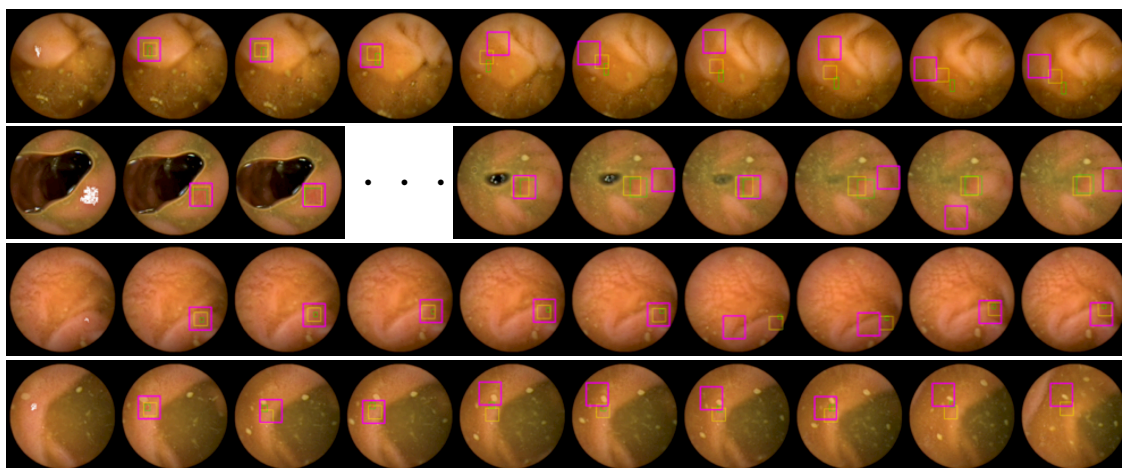


図 5.2: Comparison between feature points matching based method and CNN based method on output rectangle : Feature points matching method is better then CNN based method in those cases.

図 5.3 は CNN を用いた追跡手法のオンライン学習なしの場合の方がよかった例

である．特徴点ベースの追跡手法は画像の全体を使って位置の最適化をしているとも捉えることができるため，画像全体で局所ごとに違う動き・変化をするこれらの例は苦手である．一方で，CNN を用いた追跡手法は CNN を用いた検出による追跡手法のため，全体の変化に強い．

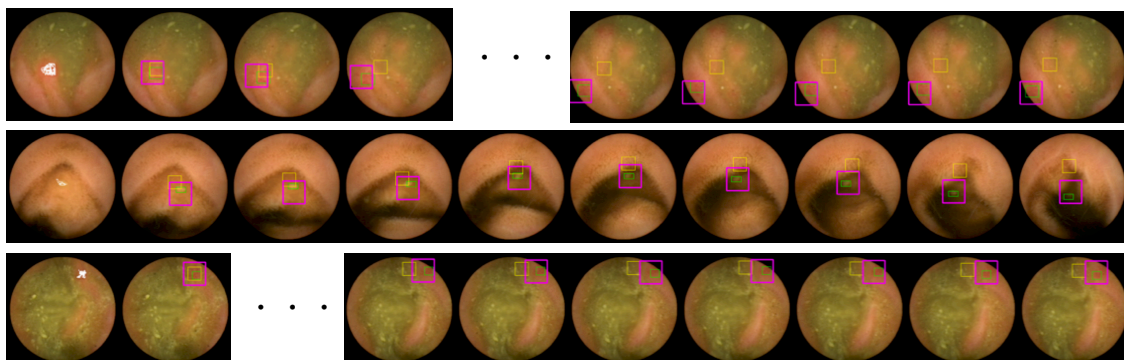


図 5.3: Comparison between feature points matching based method and CNN based method on output rectangle : CNN based method is better then feature points matching based method in those cases.

図 5.4 は両方とも成功している例である．

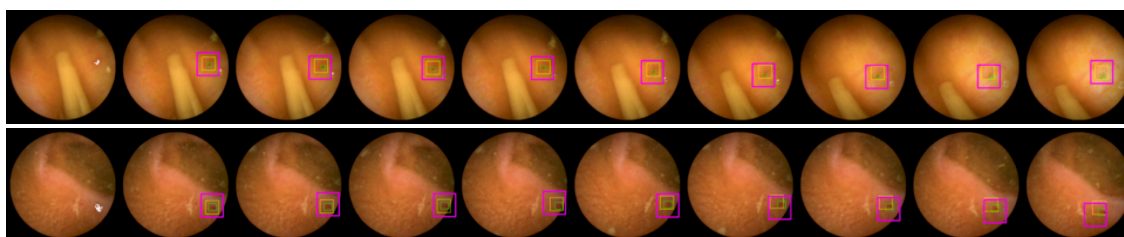


図 5.4: Comparison between feature points matching based method and CNN based method on output rectangle : In those cases , both method are correct.

5.2 消失フレームの判定

カプセル内視鏡画像を用いた医師の診断を助けるアプリケーションや、医師の診断能力を上げるトレーニングシステム構築の正解データ作成のためには病変領域の追跡だけでなく、病変が現れたフレームから消滅したフレームまでを特定することが重要になる。本論文で提案した二つの手法ではカプセル内視鏡画像列の中で医師または技師が病変を手動で指定またはコンピュータにて自動検出したあと、前に追跡しながら遡り、病変領域がなくなったフレームを見つけることで出現フレームを特定し、開始フレームから後ろに追跡を進め、病変領域がなくなったフレームを見つけることで消失フレームを特定することができる。

表 5.1 は二つの手法におけるカプセル内視鏡画像列の追跡に対して、病変領域の消滅を判定できたかどうかの結果を示している。「早期消失 (Lost)」はフレーム内に病変領域が存在しているにもかかわらず、消失したと判断したケースに当たる。特徴点ベースの追跡手法の 2 つの拘束条件による結果と CNN を用いた追跡手法を比較した結果、CNN を用いた追跡手法の方が成功数が多く優れていた。しかし、両提案手法の消失フレーム判定の成功率は高くなく訓練システムの正解データ作成という目的を満たすものではない。これは、図 5.5 の最後のフレームに示すように領域の追跡は成功しているが、小腸壁の動きやカメラの角度によって病変のみが見えなくなってしまった「滞在 (Stay)」のケースが多いためである。

表 5.1: The results of experiments related to detecting the disappearance of the abnormality.

	特徴点ベースの追跡手法		CNN を用いた 追跡手法
	Affine constraint	angle constraint	
成功	50	46	56
失敗：滞在 (Stay)	64	58	52
失敗：早期消失 (Lost)	6	16	11

図 5.5 は CNN を用いた追跡手法による追跡結果の例である．本論文における追跡手法は追跡に最適なように特徴量を構築しており，ターゲットを見つけることに主眼がおかれている．そのため，有無を判定するには不向きな特徴量になっていると考えられる．このことから，カプセル内視鏡の病変の追跡においては追跡と消失判定には別の特徴を用いる方が良いと考えられ，今後の検討としたい．

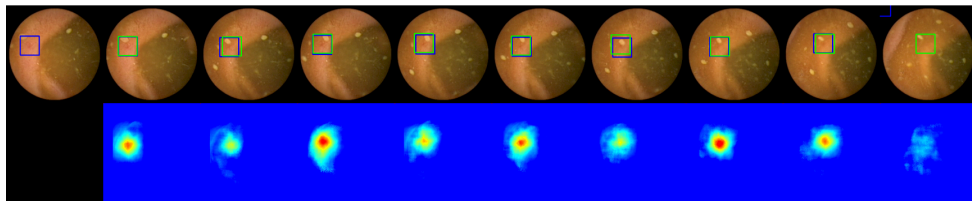


図 5.5: An example of "stay" as a classified result of CNN based method because of disappeared abnormality.

5.3 特徴点ベースの追跡手法と病変と CNN を用いた追跡手法を組み合わせた手法の提案

前節までで二つの提案手法である特徴点ベースの追跡手法と CNN を用いた追跡手法を説明し，それぞれの特徴を述べた．この二つの手法は完璧ではなく，まだ改良の余地がある．そこで，本節では二つの手法を組み合わせた手法のアイデアを提案する．特徴点ベースの追跡手法は個々の特徴点のマッチング，病変の詳細位置の高精度推定には GI color histogram を特徴量として用いている．GI color histogram は病変を視認しやすいように設計された色空間におけるヒストグラム特徴量であり，追跡に最適に定義された特徴量ではない．そこで，この特徴量を CNN を用いた追跡手法と同じように，入力フレームを用いて学習するということが考えられる．特徴点の追跡においては点ごとに学習することも考えられるが，そ

の場合，検出される特徴点の数が多くなれば特徴点の数に比例して学習時間が増加することになる．そこで，検出された特徴点とその周辺画像を用いて学習データを作成し，検出された特徴点全点分を全て用いて学習することを考える．そうすることで，1点ごとに学習するよりも特徴量の表現力は落ちることになるが，学習時間の短縮ができ，追跡対象画像に最適な特徴量を学習することができる．また，ステージ3の高精度位置推定には病変と周辺領域のみの情報を用いて学習した結果を用いるとさらに精度が高められると推察できる．

第6章 結論

カプセル内視鏡は，従来の内視鏡を用いた検査では難しかった小腸の病変の有無を検査でき，患者への負担も少ないことから活用が進んでいる．患者が飲み込んだカプセル内視鏡は消化管の蠕動運動によって推進し，胃腸管を通過する 7-8 時間の間，画像を撮影し続ける．そのため，カプセル内視鏡検査で撮りためられる画像は膨大な量になる．診断する医師はこの膨大な量の画像の中から些細な変化を見逃さずに病気の有無を診断する必要がある，負担が大きい．また，カプセル内視鏡診断は，従来の内視鏡検査と違ったスキルを必要とする．そこで，医師の診断を助けるサポートシステムや医師の技術を向上させる教育システムの研究がすすんでいる．本論文では，サポートシステムでの効率的な表示や教育システムの訓練データ作成に有益なカプセル内視鏡画像の病変領域の追跡に焦点を当て，カプセル内視鏡画像中の病変追跡手法を提案した．

本論文ではカプセル内視鏡画像中の病変追跡を対象として周辺状態を考慮した領域追跡手法として 2 つの手法を提案し，その有効性の検証を行った．

- 病変の周辺の特徴を手がかりとした，周辺の特徴点と病変の幾何拘束に基づく領域追跡（特徴点ベースの追跡手法）[16],[17]
- 病変とその周辺領域を用いたイントラフレーム学習による領域追跡（Convolutional Neural Networks（CNN）を用いた追跡手法）[18]

第 1 章では，まず，カプセル内視鏡画像における病変領域の追跡についての重要性を述べ，本研究の目的を明らかにした．さらに，病変の追跡を実現するため

に必要となる要素を整理し、それを実現するための大まかなアプローチについて論じた。

第2章では、カプセル内視鏡画像を用いた研究と画像の物体追跡手法の従来研究の紹介を行った。

第3章では、病変の周辺の特徴を手がかりとした、周辺の特徴点と病変の幾何拘束に基づく特徴点ベースの領域追跡の手法を提案した。カプセル内視鏡画像は病変と周辺部分との特徴差があまりみられないため、病変部のみを用いての追跡は難しい。一般的な画像を用いた対象追跡では追跡対象と背景が別々の動きをするため、背景の位置変化情報を追跡に利用するのは難しいが、カプセル内視鏡画像における病変領域の追跡においては、局所的にみると、病変と背景が同じ動きをするという特性を利用し、病変の周辺の特徴点を用いて追跡を行った。しかし、カプセル内視鏡画像は強いエッジを持たないため、特徴が不鮮明であり、個々の特徴点のマッチング結果が安定しない。という課題がある。そこで、提案手法は近接フレームでは病変部と周辺の狭い領域の各特徴点の幾何学的位置関係が保持される。という仮定に基づき、特徴点のマッチングに三角形の幾何拘束を適用することを提案した。三角形の幾何拘束条件として、三つの特徴点のなす角度による拘束と三つの特徴点によるアフィン行列係数による拘束の二つのアプローチを考察し、評価実験では実際の病変を含むカプセル内視鏡画像列を用い、幾何拘束条件がない場合と二つの三角形の幾何拘束条件を適用した場合についての比較実験ならびに、従来研究との比較実験を通して、特徴点ベースの領域追跡において三角形の幾何拘束を用いることの有効性評価を行った。これにより、提案手法は、動きの変位が大きく、肉眼では検出が困難な場合でも、病変を追跡することができることを示した。

第4章では、CNNを用いた病変とその周辺領域を用いたイントラフレーム学習による領域追跡の手法を提案した。カプセル内視鏡画像では病変部と周辺の特徴

の差が明確でないため、追跡に有効な特徴を予め定義するのは困難である。一方で、CNN は検出や追跡に最適な特長量を自動的に学習して定義することができる。しかし、従来の CNN を用いた物体追跡手法はあらかじめ大量の学習データが必要であるという大きな課題があった。これまで提案されている CNN を用いた追跡手法の多くは大量の正解ラベルが付与された対象画像と類似したデータを用いた事前学習が前提になっていた。本論文では類似データを用いた事前学習を行わず、追跡開始時に与えられる開始フレームと追跡ターゲットである病変の初期位置のみを用いて CNN の学習に必要なデータを構築することで、事前に大量の学習データを必要としないイントラフレーム学習による追跡手法を提案した。実際の病変を含むカプセル内視鏡画像列を用い、従来研究との比較実験を行い、これらの評価実験を通して、イントラフレーム学習による領域追跡手法の有効性評価をおこなった。ターゲットのサイズが大きく変わらない低フレームレートで撮影された小腸内視鏡画像の病変追跡においては、本手法の有効性を示すことができた。

第 5 章では 3 章で提案した特徴点ベースの追跡手法と 4 章で提案した CNN を用いた追跡手法の比較を行った。CNN を用いた追跡手法のオンライン学習を行う場合の精度が一番良いという結果になった。しかし、オンライン学習を行わない場合は、二つの手法での性能差はすくなく、特徴点ベースの追跡手法は色味が一律に変化する等の全体的に同じような特長量の変化や浮遊物が動く等の局所的な変化に強く、CNN を用いた追跡手法は蠕動運動などによって生じる凹凸の変化等の全体的なパターン変化に強いことが分かった。また、病変の消失判定については特徴点ベースの追跡手法と CNN を用いた追跡手法の両方とも十分な精度を出すことができなかった。これは追跡に用いている特長量が有無の判定に向かないことを示している。

さらに、これらの結果を受けて二つの手法を合わせた新たな手法の提案も行った。本論文では、評価画像として 2 fps で撮影した画像を使用した。現在診断に使

われているカプセル内視鏡は 2-6fps で画像を撮影している．特徴点ベースの追跡手法はフレームレートによって拘束条件のパラメータを変更することによって 6fps での画像にも対応可能であるし，CNN を用いた追跡手法についてはフレームレートに依存していないため，そのまま適用可能である．

現状，特徴点ベースの追跡手法は処理時間として 1 フレームあたり約 2 秒，CNN を用いた手法は約 20 秒かかっており，リアルタイム処理には向かない．また，提案した二つの手法は一般的な物体追跡の課題では必須項目である，追跡対象の大きさの変化には対応していない．しかし，カプセル内視鏡画像の診断サポートシステムや教育システムの訓練画像作成に対しては有益な研究結果が示せたと考えられる．

高速化については，Winner update algorithm[79] などの高速化手法を用いること，CNN の学習については，追加学習時に後段のレイヤーのみを再学習するなどの高速化の検討が考えられる．今後，処理速度の削減によるリアルタイム処理化や追跡対象のサイズ変更への対応などを検討し，一般物体認識への応用研究に取り組んでいきたい．

謝辞

本研究を推進するにあたり，大阪大学産業科学研究所 八木康史教授には御多忙な中，終始熱心な御指導を賜りました．篤く御礼申し上げます．遅々として研究の進まない，著者に対して辛抱強く，数多くの御指導・御指摘・御助言を賜りました，大阪電気通信大学 大学院 工学研究科 情報工学専攻 越後富夫教授に心から感謝致します．本論文をまとめるにあたり，大変有益な御指導と御助言を多数賜りました，大阪大学大学院情報科学研究科コンピュータサイエンス専攻 伊野文彦教授に心より感謝申し上げます．本研究を推進するにあたり，日頃からの御協力と数多くの御助言・御指摘を賜りました，大阪大学産業科学研究所 榎原靖准教授に深く感謝致します．本論文は，筆者がオムロン株式会社に在職しながら，2011年に大阪大学大学院情報科学研究科 博士後期課程に入学して以来行ってきた研究をまとめたものです．博士課程への進学するために退職することを検討していた著者に対して，在職したままの進学の道を整えてくださった，オムロンサイニックス株式会社 社長 諏訪正樹氏に深く感謝申し上げます．業務の合間に研究を遂行することを容認していただいたことも含め，様々なサポート，御助言を賜りました，オムロン株式会社 技術・知財本部 長谷川友紀氏，木下航一氏にお礼を申し上げます．進学先に悩んでいた著者に対して八木教授を紹介してくださり，さらに，入学後もサポート，多くの御助言・御指導を賜りました．広島市立大学 情報科学研究科 満上育久准教授に心から感謝致します．本研究に対して，多岐に渡って多くの御助言・御指摘を賜りました，Hanoi University of Science and Technology

Vu Hai 氏，大阪大学データビリティフロンティア機構 武村紀子准教授に深く感謝申し上げます。本研究において貴重な御助言・御指導を賜りました，奈良先端科学技術大学院大学情報科学研究科 向川康博教授，大阪大学産業科学研究所 村松大吾准教授に心から感謝致します。本研究を進めるにあたり，多くの御討論や御助言を賜りました大阪大学産業科学研究所 複合知能メディア研究分野 八木研究室の諸氏，オムロン株式会社の同僚各位に心より感謝申し上げます。最後に、研究の遂行と本論文の執筆に関して様々な面で支えてくれた両親に心から感謝します。

参考文献

- [1] R. T. Collins: “Mean-shift blob tracking through scale space”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 234–240 (2003).
- [2] H. Nam and B. Han: “Learning multi-domain convolutional neural networks for visual tracking”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4293–4302 (2016).
- [3] M. Danelljan, A. Robinson, F. S. Khan and M. Felsberg: “Beyond correlation filters: Learning continuous convolution operators for visual tracking”, European Conference on Computer Vision, pp. 472–488 (2016).
- [4] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik and P. H. Torr: “Staple: Complementary learners for real-time tracking”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1401–1409 (2016).
- [5] G. Iddan, G. Meron, A. Glukhovsky and P. Swain: “Wireless capsule endoscopy”, *Nature*, **405**, 6785, p. 417 (2000).
- [6] “日本消化器内視鏡協会”, <https://www.jges.net/citizen/check-cure/no3-3>.
- [7] 山本, 喜多, 砂田, 矢野, 林芳, 佐藤, 岩本, 井戸, 菅野: “小腸内視鏡検査”, *日本内科学会雑誌*, **93**, 6, pp. 1189–1199 (2004).

- [8] H. Yamamoto, C. Ell, H. Kita and A. May: “Double balloon endoscopy”, Video Capsule Endoscopy, Springer, pp. 113–118 (2014).
- [9] 渡辺: “カプセル内視鏡による炎症性腸疾患の診断”, 消化器内視鏡, **22**, pp. 323–328 (2010).
- [10] 白川: “カプセル内視鏡による小腸疾患の診断”, 消化器内視鏡, **17**, 4, pp. 473–479 (2005).
- [11] S. Suman, F. A. B. Hussin, N. Walter, A. S. Malik, S. H. Ho and K. L. Goh: “Detection and classification of bleeding using statistical color features for wireless capsule endoscopy images”, International Conference on Signal and Information Processing, pp. 1–5 (2016).
- [12] Y. Yuan, B. Li and M. Q.-H. Meng: “Bleeding frame and region detection in the wireless capsule endoscopy video”, IEEE Journal of biomedical and health informatics, **20**, 2, pp. 624–630 (2016).
- [13] Y. Yuan, B. Li and M. Q.-H. Meng: “Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images”, IEEE Transactions on Automation Science and Engineering, **13**, 2, pp. 529–535 (2016).
- [14] H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa and Y. Yagi: “Controlling the display of capsule endoscopy video for diagnostic assistance”, IEICE Transactions on Information and Systems, **92**, 3, pp. 512–528 (2009).
- [15] “日本カプセル内視鏡学会 内視鏡読影トレーニング”, <https://www.ce-learning.jp/login/index.php>.

- [16] Y. Yanagawa, T. Echigo, H. Vu, H. Okazaki, Y. Fujiwara, T. Arakawa and Y. Yagi: “Tracking abnormalities in video capsule endoscopy using surrounding features with a triangular constraint”, IEEE International Symposium on Biomedical Imaging, pp. 578–581 (2012).
- [17] Y. Yanagawa, T. Echigo, H. Vu, H. Okazaki, Y. Fujiwara, T. Arakawa and Y. Yagi: “Abnormality tracking during video capsule endoscopy using an affine triangular constraint based on surrounding features”, IPSJ Transactions on Computer Vision and Applications, **9**, (2017).
- [18] 柳川, 越後, 宮崎, 武村, 八木 : “イントラフレーム学習による convolutional neural networks を用いたカプセル内視鏡画像における病変領域の追跡”, 人工知能学会論文誌, **33**, 6, pp. C-I33_1 (2018).
- [19] B. Li and M. Q.-H. Meng: “Texture analysis for ulcer detection in capsule endoscopy images”, Image and Vision computing, **27**, 9, pp. 1336–1342 (2009).
- [20] Y. Yuan, J. Wang, B. Li and M. Q.-H. Meng: “Saliency based ulcer detection for wireless capsule endoscopy diagnosis”, IEEE transactions on medical imaging, **34**, 10, pp. 2046–2057 (2015).
- [21] A. Kundu, A. Bhattacharjee, S. Fattah and C. Shahnaz: “An automatic ulcer detection scheme using histogram in yiq domain from wireless capsule endoscopy images”, Region 10 Conference, TENCON 2017-2017 IEEE, pp. 1300–1303 (2017).
- [22] A. R. Hassan and M. A. Haque: “Computer-aided gastrointestinal hemorrhage detection in wireless capsule endoscopy videos”, Computer methods and programs in biomedicine, **122**, 3, pp. 341–353 (2015).

- [23] X. Xing, X. Jia and M.-H. Meng: “Bleeding detection in wireless capsule endoscopy image video using superpixel-color histogram and a subspace knn classifier”, IEEE Engineering in Medicine and Biology Society, pp. 1–4 (2018).
- [24] X. Jia and M. Q.-H. Meng: “A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images”, IEEE Engineering in Medicine and Biology Society, pp. 639–642 (2016).
- [25] T. Aoki, A. Yamada, K. Aoyama, H. Saito, A. Tsuboi, A. Nakada, R. Niikura, M. Fujishiro, S. Oka, S. Ishihara, et al.: “Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network”, *Gastrointestinal Endoscopy* (2018).
- [26] S. Sawant and M. Deshpande: “Tumor recognition in wireless capsule endoscopy images”, *International Journal of Computer Science and Network Security*, **15**, 4, p. 85 (2015).
- [27] W. S. L. Jebarani and V. J. Daisy: “Assessment of crohn’s disease lesions in wireless capsule endoscopy images using svm based classification”, *International Conference on Signal Processing, Image Processing and Pattern Recognition*, pp. 303–307 (2013).
- [28] P. Szczypiński, A. Klepaczko, M. Pazurek and P. Daniel: “Texture and color based image segmentation and pathology detection in capsule endoscopy videos”, *Computer methods and programs in biomedicine*, **113**, 1, pp. 396–411 (2014).
- [29] H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa and Y. Yagi: “Detection of contractions in adaptive transit time of the small

- bowel from wireless capsule endoscopy videos”, *Computers in Biology and Medicine*, **39**, 1, pp. 16–26 (2009).
- [30] Y. Yagi, H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi and T. Arakawa: “A diagnosis support system for capsule endoscopy”, *Inflammopharmacology*, **15**, 2, pp. 78–83 (2007).
- [31] V. Hai, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa and Y. Yagi: “Adaptive control of video display for diagnostic assistance by analysis of capsule endoscopic images”, *International Conference on Pattern Recognition*, Vol. 3, pp. 980–983 (2006).
- [32] Y. Chan, A. Hu and J. Plant: “A kalman filter based tracking scheme with input estimation”, *IEEE transactions on Aerospace and Electronic Systems*, 2, pp. 237–244 (1979).
- [33] M. Isard and A. Blake: “Condensation - conditional density propagation for visual tracking”, *International Journal of Computer Vision*, **29**, 1, pp. 5–28 (1998).
- [34] B. Ristic, S. Arulampalam and N. Gordon: “Beyond the Kalman filter: Particle filters for tracking applications”, *Artech house* (2003).
- [35] T. Zhang, A. Bibi and B. Ghanem: “In defense of sparse tracking: Circulant sparse tracker”, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3880–3888 (2016).
- [36] K. Nummiaro, E. Koller-Meier and L. Van Gool: “An adaptive color-based particle filter”, *Image and Vision Computing*, **21**, 1, pp. 99–110 (2003).

- [37] D. Comaniciu, V. Ramesh and P. Meer: “Kernel-based object tracking”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **25**, 5, pp. 564–577 (2003).
- [38] M. B. Kaâniche and F. Brémond: “Tracking hog descriptors for gesture recognition”, IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 140–145 (2009).
- [39] G. D. Hager and P. N. Belhumeur: “Real-time tracking of image regions with changes in geometry and illumination”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 403–410 (1996).
- [40] Q. A. Nguyen, A. Robles-Kelly and C. Shen: “Kernel-based tracking from a probabilistic viewpoint”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007).
- [41] R. T. Collins, Y. Liu and M. Leordeanu: “Online selection of discriminative tracking features”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 10, pp. 1631–1643 (2005).
- [42] Z. Kalal, K. Mikolajczyk and J. Matas: “Tracking-learning-detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **34**, 7, pp. 1409–1422 (2012).
- [43] J. Son, I. Jung, K. Park and B. Han: “Tracking-by-segmentation with online gradient boosting decision tree”, IEEE International Conference on Computer Vision, pp. 3056 – 3064 (2015).

- [44] H. Grabner, J. Matas, L. J. V. Gool and P. C. Cattin: “Tracking the invisible: Learning where the object might be”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1285–1292 (2010).
- [45] W. He, T. Yamashita, H. Lu and S. Lao: “Surf tracking”, IEEE International Conference on Computer Vision, pp. 1586–1592 (2009).
- [46] S. Avidan: “Support vector tracking”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **26**, 8, pp. 1064–1072 (2004).
- [47] M. Mueller, N. Smith and B. Ghanem: “Context-aware correlation filter tracking”, IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, p. 6 (2017).
- [48] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Čehovin, G. Fernandez, T. Vojir, G. Häger, G. Nebehay, R. Pflugfelder and et al.: “The visual object tracking vot2015 challenge results”, Visual Object Tracking Workshop 2015 at ICCV2015, pp. 564–586 (2015).
- [49] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin, T. Vojir, G. Häger, A. Lukežič, G. Fernandez and et al.: “The visual object tracking vot2016 challenge results”, European Conference on Computer Vision Workshops, pp. 777–823 (2016).
- [50] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin, T. Vojir, G. Häger, A. Lukežič, A. Eldesokey, G. Fernandez and et al.: “The visual object tracking vot2017 challenge results”, IEEE International Conference on Computer Vision Workshops (2017).

- [51] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pfugfelder, L. C. Zajc, T. Vojir, G. Bhat, A. Lukezic, A. Eldesokey, G. Fernandez and et al.: “The sixth visual object tracking vot2018 challenge results”, European Conference on Computer Vision Workshops (2018).
- [52] C. Tomasi and T. Kanade: “Detection and tracking of point features”, Technical report, International Journal of Computer Vision (1991).
- [53] H. Zhou, Y. Yuan and C. Shi: “Object tracking using sift features and mean shift”, *Comput. Vis. Image Underst.*, **113**, 3, pp. 345–352 (2009).
- [54] Y. Kanazawa and K. Uemura: “Wide baseline matching using triplet vector descriptor.”, The British Machine Vision Conference (Eds. by M. J. Chantler, R. B. Fisher and E. Trucco), pp. 267–276 (2006).
- [55] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei: “ImageNet large scale visual recognition Challenge”, *International Journal of Computer Vision*, **115**, 3, pp. 211–252 (2015).
- [56] A. Krizhevsky, I. Sutskever and G. E. Hinton: “Imagenet classification with deep convolutional neural networks”, *Conference and Workshop on Neural Information Processing Systems*, pp. 1097–1105 (2012).
- [57] K. Chatfield, K. Simonyan, A. Vedaldi and A. Zisserman: “Return of the devil in the details: Delving deep into convolutional nets”, *CoRR*, **abs/1405.3531**, pp. 1–11 (2014).

- [58] K. He, X. Zhang, S. Ren and J. Sun: “Deep residual learning for image recognition”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016).
- [59] K. Simonyan and A. Zisserman: “Very deep convolutional networks for large-scale image recognition”, CoRR, **abs/1409.1556**, pp. 1–14 (2014).
- [60] V. Badrinarayanan, A. Kendall and R. Cipolla: “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”, CoRR, **abs/1511.00561**, pp. 1–14 (2015).
- [61] S. Hong, H. Noh and B. Han: “Decoupled deep neural network for semi-supervised semantic segmentation”, Conference and Workshop on Neural Information Processing Systems, pp. 1495–1503 (2015).
- [62] J. Long, E. Shelhamer and T. Darrell: “Fully convolutional networks for semantic segmentation”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015).
- [63] M. Maire, T. Narihira and S. X. Yu: “Affinity cnn: Learning pixel-centric pairwise relations for figure/ground embedding”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 174–182 (2016).
- [64] R. Girshick: “Fast r-cnn”, IEEE International Conference on Computer Vision, pp. 1440–1448 (2015).
- [65] S. Ren, K. He, R. Girshick and J. Sun: “Faster r-cnn: Towards real-time object detection with region proposal networks”, Conference and Workshop on Neural Information Processing Systems, pp. 91–99 (2015).

- [66] H. Li, Y. Li and F. Porikli: “Deeptrack: Learning discriminative feature representations online for robust visual tracking”, *IEEE Transactions on Image Processing*, **25**, 4, pp. 1834–1848 (2016).
- [67] M. Danelljan, G. Hager, F. Shahbaz Khan and M. Felsberg: “Convolutional features for correlation filter based visual tracking”, *IEEE International Conference on Computer Vision*, pp. 58–66 (2015).
- [68] J. Fan, W. Xu, Y. Wu and Y. Gong: “Human tracking using convolutional neural networks”, *IEEE Transactions on Neural Networks*, **21**, 10, pp. 1610–1623 (2010).
- [69] S. Hong, T. You, S. Kwak and B. Han: “Online tracking by learning discriminative saliency map with convolutional neural network”, *International Conference on Machine Learning*, pp. 597 – 606 (2015).
- [70] N. Wang, S. Li, A. Gupta and D. Yeung: “Transferring rich feature hierarchies for robust visual tracking”, *CoRR*, **abs/1501.04587**, pp. 1–9 (2015).
- [71] M. Danelljan, G. Hager, F. Shahbaz Khan and M. Felsberg: “Learning spatially regularized correlation filters for visual tracking”, *IEEE International Conference on Computer Vision*, pp. 4310–4318 (2015).
- [72] D. G. Lowe: “Object recognition from local scale-invariant features”, *IEEE International Conference on Computer Vision*, pp. 1150–1157 (1999).
- [73] H. Bay, T. Tuytelaars and L. Van Gool: “Surf: Speeded up robust features”, *European conference on computer vision*, pp. 404–417 (2006).
- [74] H. Vu, T. Echigo, K. Yagi, H. Okazaki, Y. Fujiwara, Y. Yagi and T. Arakawa: “Image-enhanced capsule endoscopy preserving the original color tones”, in

Proceeding of Workshop on Computational and Clinical Applications in Abdominal Imaging 2011, LNCS 6668 (In Print) (2011).

- [75] S. Tokui, K. Oono, S. Hido and J. Clayton: “Chainer: A next-generation open source framework for deep learning”, NIPS Workshop on Machine Learning Systems (2015).
- [76] <https://github.com/HyeonseobNam/MDNet>.
- [77] <https://github.com/martin-danelljan/Continuous-ConvOp>.
- [78] <https://github.com/bertinetto/staple>.
- [79] Y.-S. Chen, Y.-P. Hung and C.-S. Fuh: “Fast block matching algorithm based on the winner-update strategy”, IEEE Transactions on Image Processing, **10**, 8, pp. 1212–1222 (2001).