様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　７ 年　 ５ 月 ３０ 日現在

機関番号：３２６１２
研究種目：基盤研究(C)（一般）
研究期間：2022～2024
課題番号：２２Ｋ０８７２０

研究課題名（和文）安全な内視鏡外科手術を実現する「目」の解明と技術応用基盤の構築

研究課題名（英文）Elucidation of the "eye" for safe endoscopic surgery and construction of a foundation for the technological application

研究代表者

三原　規奨（Mihara, Kisyo）

慶應義塾大学・医学部（信濃町）・共同研究員

研究者番号：８０４６４９５７

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：内視鏡外科手術における熟練スコピスト技術や予測の知識をAIを利用するなどして解析し、スコピストナビゲーションのプラットフォームを構築することを試みた。
Visual SLAMを用いた方法では、腹腔鏡の回転推定方法を用い、既存のAIモデルによる方法との比較を行なった。AI学習モデルの学習には用いられなかったデータセットではAIモデルよりも高い精度を、学習に用いられたデータセットでは特にZ軸周りに関してはAIモデルと同等程度の精度があることが分かった。術者の視野推定では、熟練医の視覚的顕著性領域を出力するフレームワークを構築し、腹腔鏡手術映像における視覚的顕著性領域の推定を実現することができた。

研究成果の学術的意義や社会的意義
触覚の乏しい内視鏡外科手術において、安全性を担保する根幹となる操作は、目で確認すること（視認）である。視認のためのスコープ（内視鏡カメラ）を操作するスコピストは、術者の目として大切な役割を果たす。本研究では、「手術の安全性を担保する最適なカメラワークとは何か」を解明するため、経験豊富なスコピストの持つ手（操作技術）と頭（操作予測）を解析し、暗黙知の可視化を試みた。この研究の解析データは、術者の意図する視野を提示する革新的スコピストナビゲーションシステムやオートメーションロボットカメラの開発に大きく貢献すると考えられる。

研究成果の概要（英文）：　We attempted to build a platform for scopist navigation by analyzing the knowledge of skilled scopist techniques and predictions in endoscopic surgery, including the use of AI.
The Visual SLAM-based method was used to estimate the rotation of the laparoscope and compared with the existing AI model. was found to be as accurate as the AI model, especially around the Z axis.　For the estimation of the surgeon's field of view, we were able to construct a framework that outputs the visual saliency region of a skilled physician and realize the estimation of the visual saliency region in laparoscopic surgery videos.

研究分野：外科学

キーワード：内視鏡外科　スコピスト　腹腔鏡下胆嚢摘出術　AI　Visual Odometry (VO)

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様　式　Ｃ－１９、Ｆ－１９－１（共通）

１．研究開始当初の背景

超高齢化社会の到来にともない、身体的負担が軽く、術後社会復帰が早い内視鏡外科手術の需要が高まっている。一方で、我が国では外科医不足が深刻化しつつあり、内視鏡外科手術におけるスコープ(内視鏡カメラ)を操作するスコピストは臨床研修医や外科修練医が担当する(図 1)ことが多い。また、2021 年には臨床工学技士法の改正に関連した政省令等が交付され、スコピストを臨床工学技士が担当することも可能になった。

触覚の乏しい内視鏡外科手術における安全性の担保は、目で確認すること（視認）でなされ、スコピストはその根幹となる役割を担う。未熟なスコープ操作は、視認操作に時間を要し、術者のフラストレーションの増大や手術時間の延長につながる。熟練医と修練医のスコープ操作技術を比較すると、術野の中心化・水平保持・手術機器の視認・術者の言語あるいは動作による指示の回数の全ての項目で、有意な差を認めた。スコープ操作には、目と手のコーディネーションや手術手順の予測といった複雑な精神運動タスクが要求される。経験を補うため、多くのトレーニングシステムが報告されている。

２．研究の目的

本研究では、内視鏡外科手術における熟練スコピストの持つ手（操作技術）と頭（操作予測の知識）を AI を利用して解析し、スコピストナビゲーションのプラットフォームを構築することを試る。熟練スコピストのスコープ操作の位置情報のデータ化と予測される未来のスコープ操作の解析を行う。これによって、「どの位置からみた視野が最適か」と「次の視野展開のためにどこにスコープを移動するのか」の解明が可能になると考える。なお、本研究は、「Visual SLAM を用いた方法」と「術者の視野推定」の 2 つのアプローチから進められた。

３．研究の方法

（１）Visual SLAM による研究

Visual SLAM による研究では、２枚の画像間における相対姿勢の回転成分を推定し続けることによって世界座標系におけるカメラ姿勢の回転成分を更新し続け、術中に途切れることなく腹腔鏡の回転を推定・可視化することによりスコピストの腹腔鏡操作を支援することを目的とした。まず最初に、画像の位置合わせの手法である[1]によって２画像間の平面射影変換行列と、変換による画像の位置合わせの正確さを示す相関係数を得、次にこの相関係数がある一定の閾値より低い場合は２画像間の相対姿勢の回転成分はないと推定し、高い場合は平面射影変換行列から２画像間の相対姿勢を計算し、さらに計算された相対姿勢をカルマンフィルタ[2]に通すことで最終的な２画像間の相対姿勢を推定する手法を提案した。

EndoSLAM データセット[3]と SCARED データセット[4]を用いて提案手法の評価を行った。EndoSLAM データセットは手術動画と、動画内のフレームに対応するカメラ姿勢の正解データも付随したものであり、今回は HighCam と LowCam と呼ばれる、ロボットアームに装着された内視鏡を用いて撮影された 2 種類のデータを評価に使用した。今回の実験に際して 2 種類の前処理を行った。1 つ目がフレームと姿勢正解データの同期である。フレームと正解データはデータ公開元がすでに[5]を用いてオプティカルフローとカメラの光軸方向の並進運動を用いて同期しているが、実際にフレームと対応する姿勢正解データの回転を同時に可視化すると動きが合っていないことがわかった。そのため、可視化されたフレームと姿勢正解データを光軸周りの回転を基準にして目視で同期させた。いくつかのシークエンスでは正確に同期させることができなかったため、それらのシークエンスは本実験には用いなかった。2 つ目の前処理として、連続したフレームの切り取りを行った。データセット内でいくつかのフレームが欠損していたため、フレームが２枚以上連続している部分のみを切り取り、実験に使用した。

SCARED データセットは AF-SfMLearner[6]の学習に用いられたデータセットであり、手術動画と動画内のフレームに対応するカメラ姿勢の正解データも付随している。SCARED データセットに関しては AF-SfMLearner の学習に使用されなかったテストデータを使用した。実験の比較対象としては AF-SfMLearner を用いた。この手法は手術動画を用いた自己教師あり学習で画像に対応する深度マップと２枚の画像を入力した際にはカメラの相対姿勢を３次元のオイラー角と並進ベクトルを出力する深層学習モデルである。今回の実験では、学習させるサブモジュールによって学習のフェーズを２段階に分ける方法を採用した事前学習が行われたモデルを使用した。このモデルの学習には SCARED データセット[4]が用いられている。

（２）術者の視野推定の研究

術者の視野推定の研究では、腹腔鏡手術映像を複数人の熟練医が観測し、熟練医の視覚的顕著性領域を予測することを目的とした。しかし医療分野における動的な視覚的顕著性データセットはこれまで存在しておらず、手術動画は DHF1K などの従来のデータセットで扱われている一般的な動画とは大きく異なる特徴を持つ。そのため、DHF1K で学習したモデルは腹腔鏡手術の動画において高いパフォーマンスを発揮できない。そのため、本研究では、腹腔鏡手術に特化したデータセット腹腔鏡手術における手術工程の分類などで使用されているベンチマークデータセットである Cholec80 [7]を活用した。Cholec80 には、80 本の高品質な手術動画が含まれてい

る。本研究では、これらの手術動画に熟練医の視点データをアノテーションすることで、動的な視覚的顕著性データセットを作成した。DHF1K をはじめとする一般的な視覚的顕著性予測データセットは、15 秒程度の短い動画を対象としている。一方で、腹腔鏡手術映像では、約 30 分間にわたる長時間の動画が対象となる。熟練医は、モニターに映し出された手術映像を観察し、モニター上の視点を記録した。視点計測に使用した手術映像は、Cholec80 を活用し、視点計測には Tobii Pro Nano を使用した。

本研究では、TASED-Net [8]と提案データセットを用いることで、実験と評価を実施した（図 1)。具体的には、Cholec80 の 32 フレームのクリップを入力とし、最終フレームの Saliency map を出力した。さらに、モデルからの出力と提案データセットの Saliency map を比較し、KLD(Kullback-Leibler Divergence) を用いて損失を計算した。評価時には、スライディングウィンドウ方式でクリップを取得することにより、全フレームの Saliency map を出力した。

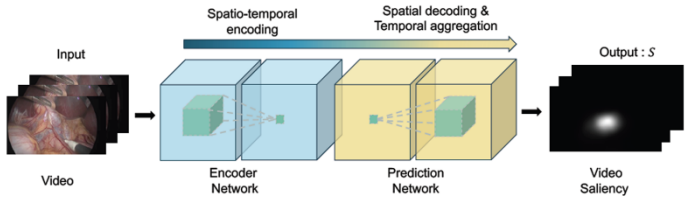

図 1　モデル概要

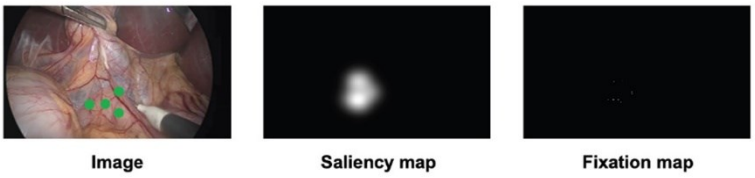

図 2　提案データセット

図 2 に示すように、計測した注視点をプロットしたものを Fixation map とし、その上に二次元ガウシアンを適用したものを Saliency map とした。ガウシアンのパラメータは、DHF1K に基づくものを使用した。Saliency map は Ground Truth (GT) として学習に用い、Fixation map は評価指標の計算に用いた。評価指標としては、予測された Saliency map と GT (Ground Truth) との類似度を示す NSS (Normalized Scanpath Saliency)、SIM (Similarity)、および GT と予測結果の相関関係を示す CC (Correlation Coefficient) を使用した。ここで、S は予測された Saliency map、S(xi) は座標 xi における顕著性値、G(xi) は GT (Ground Truth) における値、μ は平均値、σ は標準偏差をそれぞれ表している。NSS は、視点のアノテーションがあるピクセルにおける Saliency map の正規化スコアを基に計算され、視点情報を直接利用する指標である。SIM は、NSS の計算を全てのピクセルで行ったものであり、予測された Saliency map と GT の類似性を定量的に評価する。CC は、ピクセルごとの顕著性値の相関係数を計算し、予測結果と GT の間にどの程度の線形関係があるかを示す指標である。これらの指標を組み合わせることで、異なる観点から評価することが可能である。特に、NSS は視線情報に基づく直接的な評価を行い、SIM や CC は Saliency map 全体の分布特性を詳細に比較するための補完的な役割を果たす。これにより、予測精度の詳細な比較が可能となり、モデルの性能を評価することができる。作成したデータセットが視覚的顕著性データセットとして成立しているかを評価するために、一般的な顕著性データセットと提案データセットを使用し、各データセットを学習とテストに分割して、モデルを訓練・評価した。具体的には、DHF1K[9]、Hollywood2[10]、UCFSports[11]を使用することにより、提案データセットの性能を評価した。

４．研究成果

（１）Visual SLAM による研究

表 1 と表 2 は、各種提案手法と AF-SfMLearner[6]の正解データに対する平均 L2 ノルム誤差平均誤差をそれぞれ HighCam データと LowCam データにおいて比較している。それぞれの数値の単位はラジアンである。LowCam データにおける Stomach-III シークエンスでは前処理にて正確に同期させることができなかったため比較には用いなかった。表 1 と表 2 の結果を総合してみると、相関係数に対する閾値とカルマンフィルタ[2]の両方を活用した提案手法が 9 種類のうち 6 種類の内臓において最も誤差が小さいという結果になった。この結果より、他のデータセットで学習された深層学習モデルである AF-SfMLearner よりも回転成分の推定精度が高いこと、そして相関係数に対する閾値とカルマンフィルタの活用が有効であることがわかった。図 3 と図 4 では EndoSLAM データセットにおける AFSfMLearner と提案手法の挙動の一例を示している。図 3 と図 4 から、提案手法が Af-SfMLearner よりも精度良く回転推定できていることが分かる。図 5 では SCARED データセットにおける AF-SfMLearner と提案手法の挙動の一例を示している。

AF-SfMLearner は SCARED データセットを学習に使用しているため、提案手法よりも正解データ
図 6 想定している腹腔鏡の回転の可視化方法。(左上、右上、左下)それぞれカメラの後ろ、上、横から見た時のカメラ姿勢の可視化;(右下) 腹腔鏡で写された体内[3]に近い回転推定ができている。しかしながら、Z 軸周りの回転に関しては提案手法も同等程度の精度の回転推定ができていることがわかった。

表 1 HighCam データにおける平均 L_2 ノルム誤差の比較

Model	Stomach-I	Stomach-II	Stomach-III	Colon-IV	Small Intestine
AF-SfMLearner[9]	0.4343	**0.3025**	0.2201	0.2443	0.3162
Ours w/o threshold and Kalman filter	1.647	0.685	0.2772	0.2174	0.4497
Ours w/o threshold	1.639	0.6417	0.2604	0.1905	0.4178
Ours w/o Kalman filter	0.4401	0.4151	0.1956	0.1612	0.2645
Ours	**0.4254**	0.427	**0.1908**	**0.1441**	**0.2578**

表 2 LowCam データにおける平均 L_2 ノルム誤差の比較

Model	Stomach-I	Stomach-II	Stomach-III	Colon-IV	Small Intestine
AF-SfMLearner[9]	**0.0944**	**0.4393**	-	0.3607	0.4197
Ours w/o threshold and Kalman filter	1.098	1.427	-	0.4296	0.9948
Ours w/o threshold	0.7742	1.409	-	0.3817	0.8576
Ours w/o Kalman filter	0.5997	0.647	-	0.3248	0.4289
Ours	0.4359	0.6265	-	**0.3009**	**0.4102**

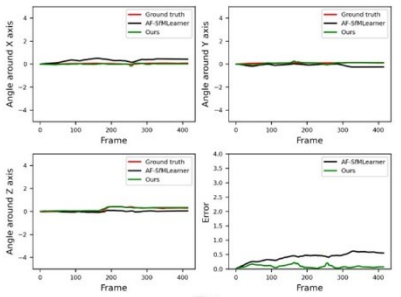

図3

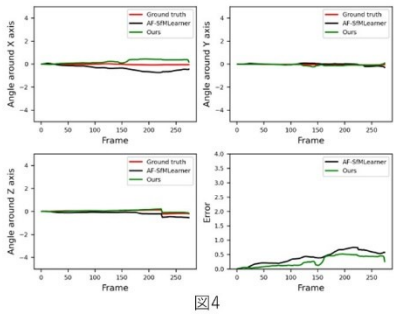

図4

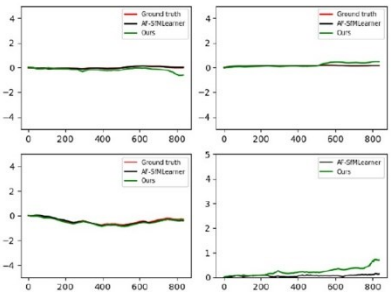

図5

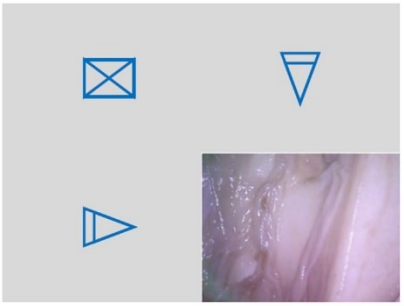
図6

図 3：HighCam データの Colon-IV シークエンスの一部におけるカメラ姿勢の各軸周りの回転成分と正解姿勢との L2 ノルム誤差の比較.
図 4：LowCam データの Small Intestine シークエンスの一部におけるカメラ姿勢の各軸周りの回転成分と正解姿勢との L2 ノルム誤差の比較
図 5：SCARED データセットのテストデータにおけるカメラ姿勢の各軸周りの回転成分と正解姿勢との L2 ノルム誤差の比較.
図 6：想定している腹腔鏡の回転の可視化方法.(左上, 右上, 左下)それぞれカメラの後ろ, 上, 横から見た時のカメラ姿勢の可視化;(右下) 腹腔鏡で写された体内

（2）術者の視野推定の研究

TASED-Net を提案データセットで学習させた結果は表 3 に示す通りである。従来のデータセットと比較しても同等の結果が得られていることから、本研究で作成したデータセットが顕著性予測における有用なデータセットであることが示唆される。 また、図 7 は手術映像中の一部のフレームを示しており、この図からも、単に術具の先端周辺領域を推定しているだけではなく、広範な術野に対する推定が行われていることが確認できる。これらの結果から、提案したデータセットを用いることで、腹腔鏡手術における熟練医の顕著性領域の推定が実現できていることが示された。

	NSS (↑)	CC (↑)	SIM (↑)
Ours	2.05	0.622	0.497
DHF1K	2.667	0.470	0.361
Hollywood2	**3.302**	**0.646**	**0.507**
UCFSports	2.920	0.582	0.469

表 3 実験結果

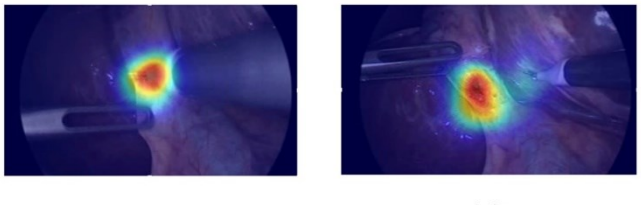

図 7 定性評価

①仮定の検証
検証条件：出力された顕著性領域内における計測した術者の注視点がどの程度の割合で存在しているかを、閾値を用いて検証した。出力される Saliencymap は、各ピクセ 23 ルに 0 から 255 までの値が割り当てられており、値が大きいほど視覚的に顕著な領域を示している。本研究では、閾値として 50、100、150、200 を使用した。
検証結果：熟練医の視覚的顕著性領域が術者の注目領域であるという仮定の検証結果について、閾値を 150 や 200 といった厳しく設定した場合には精度が低下するが、閾値を 50 程度に設定すると高精度で注目領域の推定が行えている。なお、閾値 50 は緩い条件のように思えるが、実際には画像の全ピクセル数の約 1 割に該当し、かなり限られた領域である。したがって、注視点ではなく注目領域の予測という観点では、十分に適切な閾値であると言える。術者の注目領域の推定が正確に行われていることが確認され、仮定が正しいことが明らかとなった。
②中心付近注視との比較
手術現場におけるスコピストの多くは訓練を積んでおり、優れたカメラワークを行うため、術者が多くの場面で中心付近を注視しているのではないかという疑問が生じる。そのため、顕著性予測が注目領域の予測に有効ではないのではないかと考えられる。そこで、本章では、各フレームの中心を最大値とするガウシアンマップを作成し、モデルの出力と比較を行った。作成したガウシアンマップ内に術者の注視点がどの程度の割合で含まれているかを比較した結果から、顕著性予測が単なる中心注視よりも高精度であることが確認され、術者が必ずしも中心付近を注視しているわけではないことが明らかとなった。
③術具の先端付近注視との比較
術者は術具の先端だけを注視しているのではなく、次の手術工程を見据えた箇所も注視していることがわかっている。しかし、実際には大抵術具の先端を注視しているのではないかという疑問が生じる。そこで本章では術具の先端を最大値とするガウシアンマップを作成し、モデルの出力と比較した。本研究で使用している Cholec80 には術具位置が記録されていないため、既存の術具検出モデルを用いて術具の位置を取得した。近年、ゼロショットの優秀な物体検出モデルが提案されている。しかし、手術環境という特殊なドメインにおいては精度が低いことが判明した。そのため、本研究では腹腔鏡手術における術具検出手法として Conv-LSTM トラッカー[11]を採用した。Conv-LSTM トラッカー[11]弱教師あり学習を用い、CNN で空間特徴を抽出し、LSTM で時系列情報を処理する手法では、腹腔鏡手術動画データセットである Cholec80 を用いて学習されており、腹腔鏡手術に特化した追跡能力を有している。不完全なラベルデータを活用しながら、高い追跡精度を実現しており、アノテーションが少ない状況下でも優れた性能を示していた。
④術具の先端
腹腔鏡手術では、術野に複数の術具が存在する場合がある。重要な手技、例えば切断などを行う術具は、右側から挿入されることが一般的である。これは、腹腔鏡手術の器具がそのように設計されており、手術全体を通じて一貫しているためである。したがって、複数の術具が検出された場合には、右側の術具を選択する。一般的に、術具の挿入方向は一定である。そこで、本研究では、検出されたバウンディングボックスと術具の中心の中点を先端として定義した。また、バウンディングボックスの左右どちらの境界を選択するかは、術具の挿入方向に基づいて決定した。最後に、定義した先端を中心とするガウシアンを作成した。医師が必ずしも術具の先端を注視しているわけではないことが明らかとなり、術者の注目領域の推定には視覚的顕著性がより効果的であることが示された。

本研究では、手術動画からスコープ操作のデータを解析することで次の操作の未来予測を試み、術者でなくスコピストに注目した。内視鏡外科手術の AI 動画分析に関する報告は、縫合や結紮などの小さな動作の認識や、手術動画の場面同定、手術中の臓器の位置推測などに止まり、未来予測という時間軸を加味した解析は行われていない。本研究では、近未来情報を解析することで、手術手順の予測を可能とする。また、これまでの手術支援の研究は術者支援に主眼が置かれてきたが、本研究ではスコピストに注目した。スコピストの多くは、外科チーム内で、最も経験の浅いメンバーが担当する。手術の質を効率よく向上させるには、チーム内で最も経験の浅いメンバーに焦点を当てることが肝要である。

この研究で用いたスコープの操作位置情報の解析データと未来予測データは、将来の知能をもったオートメーションカメラシステムへの発展性を備えている。

[1] IEEE, Vol. 30, No. 10, pp. 1858-1865 (2008).
[2] ASME-Journal of Basic Engineering, Vol. 82, No. Series D, pp.35-45 (1960).
[3] Medical Image Analysis, Vol. 71, p.102058 (2021).
[4] Stereo Correspondence and Reconstruction of Endoscopic Data Challenge (2021).
[5] P Artificial Intelligence - Volume 2, IJCAI'81, p. 674-679 (1981).
[6] Medical Image Analysis, Vol. 77, p. 102338 (2022).
[7] IEEE Transactions on Medical Imaging, Vol. 36, No. 1, pp. 86-97, 2017.
[8] I IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.
[9] Revisiting video saliency: A large-scale benchmark and a new model, June 2018.
[10] IEEE Transactions, Vol. 37, No. 7, pp. 1408-1424, 2015.
[11] Cognitive computation, Vol. 3, pp. 5-24, 2011.

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計5件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 三原規奨　野島大資、松田祐子、植松陽介、田中真之、北郷実、青木義満、北川雄光
２．発表標題 AIによる視点予測を利用した内視鏡外科手術支援
３．学会等名 第79回日本消化器外科学会総会
４．発表年 2024年

１．発表者名 廣瀬和輝、斎藤英雄、内山英昭、山口真弘、三原規奨、松田祐子、植松陽介、田中真之、北郷実、北川雄光
２．発表標題 スコピスト支援を目的とする画像の位置合わせを用いた腹腔 鏡の回転推定手法の提案
３．学会等名 第 27回 画像の認識・理解シンポジウム MIRU2024
４．発表年 2024年

１．発表者名 Kazuki Hirose, Hideo Saito, Hideaki Uchiyama, Masahiro Yamaguchi, Kisho Mihara, Sachiko Matsuda, Yosuke Uematsu, Masayuki Tanaka, Minoru Kitagou, Yuko Kitagawa
２．発表標題 Estimating Rotation of Laparoscope Using Image Alignment for Navigating Scopists
３．学会等名 The 20th Asian Conference on Computer-Aided Surgery (ACCAS 2024)（国際学会）
４．発表年 2024年

１．発表者名 野島大資、三原規奨、北郷実、北川雄光、島見晃平、青木義満
２．発表標題 腹腔鏡手術におけるスコピスト支援を目的とした熟練医の視覚的顕著性予測
３．学会等名 ViEW2024	ビジョン技術の実利用ワークショップ
４．発表年 2024年

１．発表者名 野島大資、三原規奨、萩原克洋、田中真之、北郷実、北川雄光、青木義満
２．発表標題 腹腔鏡手術映像における視点解析によるスコープオートメーション
３．学会等名 第36回日本内視鏡外科学会
４．発表年 2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	青木　義満 (Aoki Yoshimitsu) (00318792)	慶應義塾大学・理工学部（矢上）・教授 (32612)	
研究分担者	田中　真之 (Tanaka Masayuki) (30573414)	慶應義塾大学・医学部（信濃町）・助教 (32612)	
研究分担者	植村　宗則 (Uemura Munenori) (50636157)	神戸大学・医学研究科・非常勤講師 (14501)	
研究分担者	北郷　実 (Kitago Minoru) (70296599)	慶應義塾大学・医学部（信濃町）・准教授 (32612)	
研究分担者	斎藤　英雄 (Saito Hideo) (90245605)	慶應義塾大学・理工学部（矢上）・教授 (32612)	

６．研究組織（つづき）

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	川平　洋 (kawahira Hiroshi) (90447285)	自治医科大学・医学部・教授 (32202)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関