

# 学 士 論 文

題 目 熟練度の差に着目した機器操作  
のモデル化と体験映像の構造化

指導教員 中村 裕一 教授

京都大学工学部電気電子工学科

氏 名 上野 裕貴

平成 29 年 2 月 10 日

# Abstract

## Modeling Machine Manipulation and Structuring Experience Videos Focusing on Differences in Skill Level

Yuki Ueno,  
Undergraduate School of Electrical and Electric Engineering,  
Faculty of Engineering,  
Kyoto University

The development of wearable devices makes us able to easily record a wide range of daily experiences. However, the video recording itself is pretty redundant and needs great effort to review. Automatically extracting meaningful information from a large amount of experience video is necessary to reuse it in the future.

In this paper, we tackle this problem, considering sewing machine operation as a practical daily experience. Here important issues are how to organize video records to make useful instruction documents, and what features are effective to do that.

We focus on common and different manipulation between expert's and beginner's operations. The former can be used as necessary manipulations to archive a task, and the latter must be good information for support of wondering users. Touch, gazing and head movement will work to distinguish them. Both on those conditions, we design a state transition model to organize machine operations. In the experiment, we evaluate the automatic matching of expert manipulations and those of beginners as fundamental processing to extract common and different manipulations.

The result said that while the proposed method that uses touch features can detect most of the expert's manipulation, it remains many duplicate associations. Additional features are needed for more accurate detection.

# 目次

第 1 章	序論	1
第 2 章	関連研究	4
2.1	体験の利活用を扱った従来研究	4
2.2	機器操作を対象とした従来研究	4
2.3	熟練度の違いを扱った従来研究	5
第 3 章	機器操作のモデル化	6
3.1	対象とする機器と操作	6
3.2	体験間の共通点や差分に基づいた構造化	6
3.3	熟練度に着目した操作の特徴付け	7
3.4	機器操作の提案モデル	8
第 4 章	機器操作の記録と特徴抽出	9
4.1	記録機器	9
4.2	機器への接触検出	9
4.2.1	手領域の検出	9
4.2.2	前腕領域の除去	11
4.2.3	接触の判定と接触位置・面積の取得	12
4.2.4	接触位置の変換	12
4.3	注視位置推定	13
第 5 章	評価実験	15
5.1	操作記録データベースの作成	15
5.2	操作特徴の妥当性検証	16
5.3	上級者操作と初中級者操作の対応付け	17
5.3.1	対応検出手法	17
5.3.2	結果と考察	18

第 6 章 結論	20
謝 辞	21
参 考 文 献	22

# 第1章 序論

装着型カメラをはじめとするウェアラブルデバイスの技術発展により、映像・音声・加速度・位置情報・心拍・体温など、人間の振る舞いを様々なモダリティで記録することができるようになった。最新のウェアラブルデバイスは、誰にでも簡単に操作できる、行動を阻害することなく記録できる、小型軽量で使用する場所や時間の制約が少ないなどの利点がある。このような背景から、専門的な知識がなくても簡単に体験を記録・蓄積・再生できるようになった。特に映像記録は一般ユーザーが Youtube やニコニコ動画を代表とする動画 SNS へ簡単に投稿できるほど日常生活に浸透している。

また、体験映像を用いた人間の行動分析も盛んである。例えば、頭にカメラや各種センサを装着して記録した映像ライフログから日常生活全般を分析したり<sup>1, 2)</sup>、記憶に問題を抱えた人を補助することが試みられている<sup>3, 4)</sup>。さらにこのような映像は、フィールドワークやグループワーク等の体験活動の記録などにも用いられている。活動参加者個人の記録として用いるだけでなく、参加者間で共有したり、異なる参加者の記録を比較して共通点や違いを取り出したりと、記録を組み合わせることで総合的な分析も行われている<sup>5)</sup>。

これら体験映像記録の利活用における大きな課題の1つは、取得した大量の体験映像からどのように意味のある情報を取り出すのかである。なぜなら、ウェアラブルデバイスで記録される体験映像は長時間の取り流しになりやすく、適切な編集が施されていないため非常に冗長であり、目的の情報や閲覧して意義のある部分を探し出すには大きな労力を要するからである。体験映像を参照しやすい形で提示するために、その特性や利用目的に応じた構造化が求められる。

このような背景から、本研究では体験映像記録を構造化する上での指標、すなわち何に注目してどう分析すれば大量の記録から有用な情報を取り出せるのかについて基礎検討を行う。対象とする体験記録は、機器を扱っている場面とした。我々の身の回りには家電・オフィス機材・工作装置・電子案内板などの様々な機器が存在する。それらを使うには各々に適した操作方法や手順が求められる一方で、うまく扱うためのノウハウやコツ、うまくいかないときの対処法、同じ結果を得るにしても異なる操作がありうる、などの多様性も含んでいる。そのため、例えば

クラウドソーシングのような方法を通して多数のユーザーによる様々な操作記録を集めて構造化・分析を行えば、このような有用な情報を取り出して利活用できると期待されている。

本研究では機器操作記録を構造化する指標として、主に利用者の熟練度に着目する。すなわち、熟練者・中級者・初級者などの違いと操作方法・手順・振る舞いの関係を明らかにする。熟練度に着目した理由は以下の通りである。

- 構造化に対する良い指標であること

機器操作者の熟練度と、先に述べた操作の多様性には強い関係性があると考えた。例えば、熟練者は必要十分な操作だけを適切に行う傾向にあるが、初中級者の記録には余分もしくは無関係な操作が含まれている可能性が高く、操作方法も適切でないことが多い。

- 支援方法の選択に利用できること

熟練度の差から生じる操作方法・手順・振る舞いの違いは、記録の編集や提示方法にも利用できる。例えば、初めて使う人は使い方がわからず、説明書などを読んでも、細かいニュアンスなどが伝わらずイメージしにくい。これに対して、熟練者による最適な操作をマニュアルとして提示することは、操作方法の端的な理解に役立つ。熟練者と初心者の操作の違いから、間違いやすい操作や陥りやすい問題点を整理し、失敗を防ぐのも効果的と考える。

構造化に対する具体的なアプローチは次のとおりである。まず、熟練度の差を考慮した機器操作モデルを設計する。これは具体的な機器を想定した上で、種々のタスクを達成するための操作方法・手順・振る舞いなどが熟練度によってどのように異なり、なぜ異なるのかについて時系列に沿って整理することに相当する。続いてその違いを自動的に取り出すにはどのような特徴（物理量）を用いればよいのかについて検討する。ほとんどの場合、機器の操作を行うには手で触れる必要があるため、接触特徴を用いれば操作の有無や手順の違いが取り出せると考える。また、操作時にどこを見ているのかも熟練度や操作結果の良し悪しと関係が深いと思われる。そこで本研究では、機器への接触時刻と接触場所およびその時の注目位置を特徴とした自動検出を提案する。そして、実際の操作映像を入力としたときに複数の機器操作の違いを正しく取り出せるのか、熟練者と初心者の間で取り出された違いのパターンが提案したモデルで説明できるのかを検証する。

本稿の以降の構成は次のとおりである。まず第2章では関連研究について説明する。続いて第3章で機器操作のモデル化について述べる。第4章では機器操作の記録と特徴抽出について述べ、第5章で実験とその結果・考察について述べる。最後に第6章で結論を述べる。

## 第2章 関連研究

### 2.1 体験の利活用を扱った従来研究

体験記録の利活用を扱う試みは本研究だけでなく過去にも多く提案されている。相澤らは体験映像とともに脳波を取得することで、興味や関心などの個人の主観を反映した映像の要約を行っている<sup>6)</sup>。また、中村らはカメラ装着者が何かに注目しようとした場面を用いて、行動を要約するのに適切なシーンを検出することで映像の要約を行っている<sup>7)</sup>。

このように重要度の低い部分を削除して注目すべき時空間を取り出すことで記録全体を要約することは従来研究によってほぼ達成できているといってもよい。しかし、重要だとみなされた複数の体験の間にどのような関係があるのか、異なる体験は何を意味するのかにまで踏み込んで構造化してはいない。扱っている体験記録やその利用目的がそもそもそのようなことを求めているのが大きな理由の一つと考える。例えば、自身の日常生活をアルバムのように保存して事後に見返したりするのであれば、重要部分の検出と日付や場所といった簡単なインデックス付きで十分かもしれない。それらの分析や意味づけは専らユーザーが行えばよいといった考え方である。しかし、自身の一生の記録や複数の人間の体験記録まで規模を広げたり、自身でなく他者（遠い未来の自身も他者とみなせるかもしれない）が利用する場合、意味づけを伴う構造化を自動的に行う仕組みは必要不可欠である。

### 2.2 機器操作を対象とした従来研究

意味づけを伴う構造化が必要な体験記録の例として、本研究では機器操作に着目しており、その根拠は先に述べた通りである。機器操作の記録を扱った従来研究には以下のようなものが挙げられる。上野らは操作対象機器の時系列変化を手領域の出現と関係づけることで、作業記録の時間分割を行っている<sup>8)</sup>。また、三角らは作業中にどの時点でどのオブジェクトが使われたのかをRFIDタグを用いて記録し、異なる作業間の比較や過去に行われた作業から現在進行中の作業に適

したものを自動的に検索・提示している<sup>9)</sup>。Damenらは視線情報と一人称映像を組み合わせることで、例えば「ドアノブ」を回して引くことで「扉を開けた」といったように、何に対してどのような操作が発生したのかを自動で検出して体験映像の構造化を行っている<sup>10)</sup>。

これらの従来研究では重要部分の自動検出やタグ付けに基づいた単純な記録の再生はできるものの、操作方法や手順の違い、その理由の意味づけなどの自動分析は手掛けられていない。そのため、本研究がそれらを扱う意義は大きいと考える。

## 2.3 熟練度の違いを扱った従来研究

一方で教育や介護の分野では、対象（ここでは人物）へのアプローチの違いを分析する試みがいくつか報告されている。その多くはやはり熟練度に関するものである。例えば、玉宮らは教員と生徒の相互作用として視線や注意に着目し、教育経験と視線との関係を分析している<sup>11)</sup>。文献中では、教育経験の豊富な教員は生徒全体に対して注意を向けていたが、教育経験に乏しい教員は自分の正面に座っている生徒に注意が向きがちであるという結果が報告されている。また、後藤らは経験年数の差が介護動作に与える影響について、要介護者への接触方法やその時の注視位置に着目して調査している<sup>12)</sup>。経験年数の多い熟練者は要介護者を抱きかかえるときに手や足などの体の一部だけでなく体全体に視線が向けられていたのに対し、経験年数の少ない介護者は体の一部に集中して視線を向けがちであるという結果が報告されている。

このように、熟練度の差と接触及び注視には大きな関係があると考えられるため、本研究においてもこの方法を機器操作記録に適用することで有意な分析・構造化が達成できるのではないかと期待している。

## 第3章 機器操作のモデル化

### 3.1 対象とする機器と操作

本研究では操作対象の機器としてミシンを扱う。ミシンは身近な機器であるにも関わらず正しい操作方法を知っている人が少ない。その上1つ1つの操作がある程度複雑で、タスクをこなすのに複数のステップをクリアする必要がある。また、ボタンやレバーなど多くの操作対象を持つ上に、単に縫うだけでも様々な方法があるため操作体験に多様性がある。このような理由から、多数の機器操作体験の構造化を通して使い方やコツなどを取り出し、それを再利用することの意義が大きいと考えた。

### 3.2 体験間の共通点や差分に基づいた構造化

本研究ではどのような操作が複数体験間に共通して現れるのか、逆に同じ作業をこなすにもどのような違いがあるのかに基づいて構造化を行う。なぜなら多数の操作記録から差分や共通点を取り出せれば、さまざまな用途に応じた映像マニュアルを作成できると考えたからである。具体的には以下のような利用法が期待できる。

**共通点の利用** 複数体験間に共通する操作は、タスクを達成するための必要十分な操作とみなすことができる。これらを用いれば、一連の流れを簡易な手順で示す、必ず行わなければならない重要な操作を教示する、といった機器を操作する上での基礎的な利用が可能となる。

**差分の利用** 差分として現れた操作や振る舞いはタスク達成には冗長かもしれないが、利用者を支援する目的ではむしろ有効な情報を多く含んでいる。例えば、押すボタンを間違える、ボタンがどこにあるのか探す、とりあえず押してみるといった操作は一見すると余分である。しかし、よく見ながら操作する、機器に起こる変化を確認するといった体験記録はマニュアルとしては非常に

有用である。ボタンがどこにあるのか、このボタンを押せば機器のどこが変化するのかなどは初心者を助ける良い情報である。

このように共通点から取り出した必要十分な操作によって全体の枠組みを作り、差分として抽出された振る舞いでそれを補完していくような映像マニュアルを作ることを目指す。

### 3.3 熟練度に着目した操作の特徴付け

共通点や差分を取り出すためには、対象の2つの操作が同じなのかそれとも異なるのかを判別する特徴が必要となる。ここでは熟練度の違いが操作に与える影響を考えることで、機器操作を特徴付ける良い情報にはどのようなものがあるのかについて検討する。熟練度を大きく初級者・中級者・上級者の3段階に分類したときの操作の違いは以下のように考えられる。なお、初級者を「対象の機器を操作したことがなく、説明書も読んでいない利用者」、初級者よりも少し熟練度が高い中級者を「説明書を読み、ある程度操作方法を理解している利用者」、上級者を「何度も対象の機器を操作した経験があり、使い慣れている利用者」とそれぞれ定義している。

**初級者の操作** 初級者はどこにボタンがあるのか、どのボタンを押すべきかなど、基本的な操作方法が全く分からないため「探す」「試す」といった行動が頻繁に見られるだろう。そのため、必要のない対象に触れることが多く、探すという行動によって頭部運動も大きくなると考えられる。

**中級者の操作** 初級者とは異なり無知ではないため、探す、試すという操作はあまり見られないと考える。しかし、説明書通りに操作できているのかを気にして対象を注視する、対象に触れた後の機器の変化をよく確かめるというように、操作が慎重になると予想される。そのため、機器に触れる場所とその時に見ている位置はかなり近いと考えられる。

**上級者の操作** ボタンの位置を覚えており操作後に機器に起こる変化を事前に知っているため、見ずに操作できたり、変化を確かめずに次の操作の準備を行う傾向にある。

以上の考察から，熟練度による操作の違いは「触れる」と「見る」の時空間変化に現れやすいことがわかる．そこで本研究ではこれらを定量的に示した，接触位置・注視位置・頭部運動の3つの量を体験の構造化における操作特徴とする．

### 3.4 機器操作の提案モデル

上記のような考察を踏まえて機器操作体験を構造化するモデルを設計する．ミシンを用いた縫製を含め一般に機器の操作は順序に従ってこなしていくもので，同時に複数の操作を行うことはほとんどない．つまり，時系列に沿ったモデルが適していると考えられる．そこで本研究では，一連のタスクは小さなタスクの連結で構成されているとみなし，必要十分な操作によってそれぞれの小タスクが順に達成されていくような状態遷移で機器操作をモデル化する．すなわち従事している小タスクを状態，状態遷移時に出力されるシンボルを操作とみなす．この仕組みは図3.1のように表すことができる．

モデル中の操作  $x_p, x_a$  は4章で述べる接触位置  $h_{ref}$ ，注視位置  $c^t$ ，および頭部運動  $H^t$  の操作特徴で構成されたベクトルである．ここで注意したいのは，必要十分とされる共通操作  $x_p$  はタスク全体が進む方向への状態遷移を起こす反面，そうでない操作  $x_a$  は必ずしも異なる状態に遷移するわけではなく，元の状態に戻ってしまう点である．そのような遷移が冗長・余分な操作を行ったことに対応する．ただし，似たような操作が異なる意味を持っていたり，逆に一見異なって見える操作が同じ状態遷移を起こす可能性もある．そのため，操作と状態遷移の関係を決定論的に扱うのがよいのか，確率的に扱ったほうが良いのかについては更なる議論が必要であると考えられる．

## 第4章 機器操作の記録と特徴抽出

### 4.1 記録機器

機器操作体験を記録する方法，および前章で述べたような構造化を自動的に行うための特徴抽出について述べる．本研究では機器操作体験を主に映像として記録する．これは人間が機器操作体験を参照する際に最も効果的なメディアと考えられる上に，自動分析に用いる特徴の1つである接触情報を効果的に記録できるからである．接触の有無を直接検出するには一般にタッチセンサなどを用いるが，機器表面および手の三次元位置から推定することもできる．

そこで FullHD RGB カメラ，赤外線カメラ，赤外線投射装置を搭載した Intel REALSENSE SR300（以降単にカメラと呼ぶ）を用いて映像およびシーンの三次元情報を同時に記録する．このカメラは小型軽量で簡単に頭部に装着することができるため，機器操作者の目の前で発生するイベントを漏らさず記録することができると考えた．また，カメラに3軸加速度センサー，3軸ジャイロセンサー，3軸地磁気センサーを搭載した LOGICAL PRODUCT 社製の9軸モーションセンサー（以降単に IMU と呼ぶ）を装着することで頭部の動きも合わせて計測する（図4.1）．これは自動分析のためのもう1つの特徴である注視位置を推定するためである．カメラで記録された映像の中心を注視位置とみなすこともできるが精度があまり良くないため，頭部運動に基づいた注視位置推定手法を用いることにした．

### 4.2 機器への接触検出

#### 4.2.1 手領域の検出

機器への接触特徴を取り出すために，カメラで記録された映像からまず手領域を検出する．検出には主に肌色情報を用いる．すなわち，画素単位で色に関する閾値処理を施して肌色の検出を行い，一定以上の大きさを持つ肌色領域を手領域とみなす．

しかし、入力画像で定義されている RGB 色空間での閾値処理は照明条件の変化に敏感なため、一般シーンに対して適切な閾値を設定することが難しい。そこで本研究では、RGB 色空間から YCbCr 色空間へと変換して閾値処理を行う。YCbCr 色空間とは輝度 Y、青の色差 Cb、赤の色差 Cr の 3 値で色を表現する方法である。照明変化の影響はほとんどが Y 値の変化として現れるため、残りの Cb, Cr 空間を用いることで安定した検出が望める。

RGB 色空間から YCbCr 色空間への変換は以下の式を用いる。

$$Y = 0.299R + 0.587G + 0.144B \quad (4.1)$$

$$Cb = -0.172R - 0.339G + 0.511B \quad (4.2)$$

$$Cr = 0.511R - 0.428G - 0.083B \quad (4.3)$$

YCbCr 色空間における肌色領域は予備実験の結果に基づいて次のように設定した。

$$Y \neq 0 \quad (4.4)$$

$$77 \leq Cb \leq 127 \quad (4.5)$$

$$137 \leq Cr \leq 173 \quad (4.6)$$

この方法は事前に設定された色を持つ領域を手として検出するが、最もよく現れた色を手領域として検出するような手法を併用することで精度を高める。まず、前述の方法で得られた手領域に属する画素を HSV 色空間へと変換する。HSV 色空間とは色相 H、彩度 S、明度 V の 3 値で色を表現する方法であり、波長による色の違いが H 値に現れる。そのため CbCr 空間上での矩形領域選択とは異なる方向から色領域を評価できる。RGB 色空間から HSV 色空間への変換は以下の式を用いる。

$$V = \max(R, G, B) \quad (4.7a)$$

$$S = \begin{cases} 255 \times \frac{V - \min(R, G, B)}{V} & (\text{if } V \neq 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (4.7b)$$

$$H = \begin{cases} 30(G - B)/S & (\text{if } V = R) \\ 60 + 30(B - R)/S & (\text{if } V = G) \\ 120 + 30(R - G)/S & (\text{if } V = B) \end{cases} \quad (4.7c)$$

続いて得られた H の値を用いて色相空間における正規化ヒストグラム（全体を 1 としたときの分布）を作成する。このとき、各画像上で検出された肌色領域毎の

ヒストグラム  $H_i$  だけでなく、記録されている全ての画像フレーム上の肌色検出領域に対するヒストグラム  $H_{all}$  も同時に作成する。  $H_{all}$  は記録全体での色分布を表すため、  $H_i$  の各ビンに  $H_{all}$  で重みをつけた重み付き和

$$\omega_i = \sum_{k=1}^N H_i(k)H_{all}(k) \quad (4.8)$$

を計算するとそれは対象領域の現れやすさを意味する。一定以上の現れやすさを持つ領域を手として検出するために、その値に閾値処理を施すことで手領域か否かの判別を行う。本実験ではヒストグラムのビン数  $N$  を 36 とし、閾値は予備実験の結果から 0.3 に設定した。

このような色情報のみに基づいた手法のみでは依然として肌色に近い背景が誤検出されてしまうため、作業中に手が現れる空間領域に限定するために深度値に対して閾値処理を施す。予備実験の結果に基づいて、カメラから 20cm~100cm の距離を閾値として設定した。

#### 4.2.2 前腕領域の除去

前節で述べた手領域の検出方法では、場合によっては前腕部分も検出されてしまう。しかし機器に腕部を触れさせて操作することは稀であるため、前腕部分を取り除いて手の部分だけ取り出す。この際、手の部分は前腕部分よりも円に近いという幾何学的性質を用いる。

具体的な検出手順は次の通りである。まず、前節で取得した手領域の中央を通る直線を抽出する。このとき、手腕部分は画像下端から中央・上部に向かって伸びると仮定して左右方向の輪郭を取り出す。そして行毎の輪郭中点の分布を直線に近似することで中央線を得る。手領域検出結果は手領域を 1、背景を 0 とした二値の行列として考えることができる。左右方向の輪郭画素はこの行列をそれぞれの行で水平スキャンすることで検出される。手領域を囲む最小矩形の大きさを  $(w, h)$  pixel とすると、手領域は  $hand(u, v) = \{0, 1\}, 0 \leq u < w, 0 \leq v < h$  と表すことができるため、行  $v$  を固定した上で列  $u$  を 0 から  $w - 1$  まで変化させていき、初めて  $hand(u, v) = 1$  が成立する列  $Left_v$ 、最後に成立する列  $Right_v$  を輪郭位置として取り出せる。この処理を  $0 \leq v < h$  の範囲で繰り返す。それらの中点は  $Center_v = \frac{Left_v + Right_v}{2}$  で導出され、図 4.2 に示すような分布となる。この中点の集合を直線で近似して中央線を得る。

次に抽出された直線上に中心を持つ直径 $2r$ の円の面積に対する手領域 $hand(u, v) = 1$ の割合を計算し(図 4.3), その値が最も大きいときに対応する円と手領域の共通部分を前腕部を除去した手領域として出力する(図 4.4). このとき $r$ を手領域の幅の最大値すなわち

$$2r = \max |Right_v - Left_v| \quad 0 \leq v < h \quad (4.9)$$

とすることで最も円に近い手領域を取り出すことができる.

### 4.2.3 接触の判定と接触位置・面積の取得

手がシーン中のオブジェクトに触れているならば, 手領域とその周囲の領域はカメラから見て似た距離に存在するはずである. この仮説に基づいて, 図 4.5 に示すように手領域の輪郭点 $h_{cont}(u, v)$ 毎に近傍矩形内の手領域の深度値の平均 $D_h$ と手領域周辺の深度値の平均 $D_o$ の差 $|D_h - D_o|$ を導出し, その値に対して閾値処理を施すことで接触の有無 $h_{cont}(i, j) = \{0, 1\}$ を判定する. 閾値は, 予備実験の結果に基づいて $7\text{mm}$ とした. この処置を手領域全体とそれを含む最小矩形で行わないのは, 指の先だけで触れるといった操作が多いからである. 接触部検出の結果の例を図 4.6 に示す. 接触特徴である接触位置は接触判定された画素位置の重心で与え, 接触面積は接触位置と接触判定画素との距離の中で最も長いものを半径とした面積で定義する. ただし, この特徴量計算は各画像上ではなく, 次節で述べる共通座標系に変換した上で行う.

### 4.2.4 接触位置の変換

機器操作は頭部搭載したカメラで記録されるため, 抽出した接触位置は異なる姿勢のカメラから撮影された画像上での座標であって基準点が共通ではない. 複数の接触の比較を行うためには共通の座標系への変換が必要である. 本研究では, 操作対象全体を撮影した画像 $I_{ref}$ に対して各時刻の画像 $I_t$ がどのような位置関係で対応付くかを導出した上で, それに基づいて座標変換を行う. 本来ならばシーンは3次元であるため3次元剛体変換を用いるべきであるが, 簡単化のため操作対象は平面であると仮定して2次元平面での平面射影変換(ホモグラフィ変換)を用いる.

シーンの中の同一点が  $I_{ref}, I_t$  上で現れている座標をそれぞれ

$$\mathbf{x}_{ref} = \begin{bmatrix} u_{ref} \\ v_{ref} \\ 1 \end{bmatrix} \quad \mathbf{x}_t = \begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} \quad (4.10)$$

とおくと、これらの関係はホモグラフィ行列

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (4.11)$$

を用いて

$$s\mathbf{x}_{ref} = \mathbf{H}\mathbf{x}_t \quad (4.12)$$

のように定式化される．ただし  $s$  は非線形射影であるホモグラフィ変換を線形形式で記述するための媒介変数である．なお，ホモグラフィ変換はスケーリングの不定性を持つため，その自由度は8である．一般に， $h_{33}$  の値を1に固定した形で表すことが多い．

任意の  $\mathbf{x}_t$  に対する変換座標  $\mathbf{x}_{ref}$  を求めるには既知の対応点から  $\mathbf{H}$  を推定しておく必要がある．本研究では回転・並進・スケーリングに不変な特徴点を SURF (Speeded-Up Robust Features) を用いてそれぞれの画像から抽出し，それらの特徴量の距離に基づいて対応点を自動決定する．1組の対応点から2つの制約条件が得られるため， $\mathbf{H}$  を決定するには最低4組の対応点が必要である．ただし，対応点を自動的に取り出す方法では無視できない割合で間違っただ対応も出力してしまう．そのため4組よりも十分に多い対応点を取り出した上で，その中からもっともらしいペアと変換モデルを RANSAC(Random Sample Consensus) を用いてロバストに推定する．

実験において，カメラで撮影された画像上での接触座標を  $h_t$ ，この点に対応付く共通座標系での接触座標を  $h_{ref}$  とした．また，実際に対応付けた様子を図 4.7 に，その後画像を重ね合わせた例を図 4.8 に示す．

### 4.3 注視位置推定

一般に顔向きと注視方向は必ずしも一致しないが，右に振り向くときには顔向きよりもさらに右方向を注視しやすいといったように，それらの関係は頭部運動

に強く依存することが知られている。Leelasawassukらはこの特性を利用することで、IMU センサを用いて計測した頭部の加速度および角加速度から頭部カメラ映像上の注視点の位置を推定している<sup>13)</sup>。本研究ではこの方法を用いて注視点位置の推定を行う(図 4.9)。

まず、時刻  $t-1$  においてユーザーの注視位置が画像上の座標  $c^{t-1} = (x^{t-1}, y^{t-1})$  にあるとする。また、IMU を用いて時刻  $t-N$  から時刻  $t$  までに取得した頭の動きを行列  $\mathbf{H}^t = [\mathbf{h}_t \ \mathbf{h}_{t-1} \dots \mathbf{h}_{t-N}]$  とおく。行列中の各ベクトル  $\mathbf{h}$  は

$$\mathbf{h} = \begin{bmatrix} \theta \\ a \\ \omega \end{bmatrix} \quad (4.13)$$

のように定義する。ここで、 $\theta$  は頭の方角、 $a$  は加速度、 $\omega$  は角速度である。

これら頭の動き  $\mathbf{H}^t$  と現在の注視位置  $c^{t-1}$  を入力として、回帰関数  $\Phi$  から注視位置の変化量  $\Delta c^t$  を推定する。つまり、

$$\Delta c^t = \Phi(\mathbf{H}_t, c^{t-1}) \quad (4.14)$$

のようにモデル化されている。 $\Delta c^t$  を用いて現在の注視位置  $c^t$  が

$$c^t = c^{t-1} + \Delta c^t \quad (4.15)$$

のように導出される。回帰関数  $\Phi$  の具体的な形に関しては文献を参照されたいが、大まかに言えば過去  $N$  時刻における頭部運動と注視位置の関係を経過時刻に応じた重み付き平均するものである。なお提案手法は頭部の加速度・角速度と注視位置の間の回帰モデルとなっているため、事前に回帰関数のパラメータ同定を行う必要がある。実験では被験者の正面で棒の先に巻き付けた赤いテープを様々に動かし、それを眼球運動だけでなく頭部運動を伴った自然な動作で注視させることで同定のためのサンプルデータを得た(図 4.10)。

## 第5章 評価実験

### 5.1 操作記録データベースの作成

分析・評価を行うために、熟練度が異なる利用者がミシンを使った縫製作業を行った時の操作を頭部搭載カメラで記録し、操作記録データベースを作成する。被験者にはミシンを使って布に描かれた線に沿って縫うことをタスクとして課した(図 5.1)。被験者は普段からミシンを使うことはなく、正しい操作方法を知らない男子大学生2名(被験者 A, 被験者 B)である。初級者・中級者・上級者に対応した操作記録を取得するために、被験者には次のような条件を与えることで熟練度を制御をした。

**初心者の再現** 対象の機器を操作したことがなく、説明書も読んだことがない初心者を再現するために、被験者 A に操作箇所の名前や場所についての説明を全く行わず操作手順のみ事前に伝えた状態で縫製作業を行わせた。操作手順が多いため、それを忘れた場合にのみ適宜教示し、作業終了まで行わせた。

**中級者の再現** 説明書を読み、ある程度操作方法を理解している中級者を再現するために、被験者 B に対して操作手順や各操作箇所の名前と場所などの一通りの説明を与えた上で縫製作業を行わせた。

**上級者の再現** 普段からミシンを使っており、操作に慣れている上級者を再現するために、被験者 B に同様の縫製作業を 30 回程度行わせて、操作に十分慣れた後の操作を記録した。

被験者への説明に用いたミシンの操作手順は表 5.1 の通りである。操作箇所の名称については図 5.2 を参照されたい。

被験者には縫製作業開始前に IMU を装着したカメラを頭につけてもらい、注視位置推定のためのパラメータ同定を行うサンプルデータを取得している。また、被験者に自身の操作体験記録を提示した上で、迷った操作や探したボタンなどの操作中に陥った問題点について聞き出して記録した。

## 5.2 操作特徴の妥当性検証

実際の機器操作体験における熟練度に応じた操作の違いを 3.3 節の考察と比較することで、設定した操作特徴の妥当性を検証する。ただし、比較は目視で行った。結果を以下に述べる。

初心者は、1 番目の「主電源を入れる」や 5 番目の「縫製速度調整のつまみを動かし、縫製速度を最低にする」というタスクにおいて正しい対象に触れるまでに、頭が上下左右に大きく揺れ注視位置が定まらず必要のない対象に触れていた。実験後に被験者にこの様子を見てもらったところ、どこに対象があるのか、どれが正しい対象なのかがわからず、探したり、とりあえず触れてみたりしたということだった。中級者は頭が大きく動く、操作に余分な対象に触れるということとはなかった。しかし、例えば 1 番目の「主電源を入れる」というタスクでは、ミシンの右側面にある主電源を入れる前に頭が右方向に動き右側を注視していた。また、注視位置と接触位置は近く、他にも主電源を入れることによって点灯する液晶画面と注視点との距離も近かった。被験者に記録を見てもらったところ、慎重に操作していたため対象や機器の変化を確認していたようだった。上級者は基本的に頭が動かず、3 番目の「装飾縫製用のダイヤルを回し、番号を「2」にする」というタスクを除いてミシン中央に注視位置があり、対象を目視せずに操作していることが多かった。上級者が対象を目視せずに行っていたタスクは 1 番目の「主電源を入れる」、2 番目の「布抑えレバーを下す」、5 番目の「縫製速度調整用のつまみを動かす」タスクであった。被験者に記録を見てもらったところ、対象の場所はある程度覚えており、主電源やレバー、つまみといった対象は形が他の対象に比べて特徴的であり見なくても操作できるということだった。また、1 番目の「主電源を入れる」タスクにおいて、液晶画面の変化に対して注視位置はミシン中央から変化していないことから、上級者の被験者は変化を確認しないことも確認できた。

これらの結果より、3.3 節の考察で設定した操作特徴は妥当であることが言える。

## 5.3 上級者操作と初中級者操作の対応付け

### 5.3.1 対応検出手法

提案モデルに従って機器操作体験を構造化することは、機器操作体験における共通部分を取り出すことに対応する。ここでは、上級者の操作がタスク達成に対して必要十分な共通操作であるという仮定の下で、上級者による操作と同じ効果を持つ（同じ結果をもたらす）ものが初中級者の操作記録のどこに現れているかを自動的に探し出すことを試みる。このとき、対応しなかった操作が提案モデルにおける冗長な手順や動作とみなされる。この対応付けは、上級者の操作を生起時刻に沿って並べたクエリ  $M_{exp,i} (i = 1, 2, \dots, p)$ 、探索対象である初中級者の操作を並べたターゲット  $M_{nor,j} (j = 1, 2, \dots, q | q \geq p)$  が与えられたときに、同一効果を持つ操作のペア  $c(i, j) = \{0, 1\}$  を探す問題として定式化される。ただし  $c(i, j) = 1$  のとき  $M_{exp,i}$  と  $M_{nor,j}$  が対応付くことと定義する。対応付けには機器への接触特徴を用いる。これは同じ効果を持つ操作では同じ部位に触れているという仮定に基づく。3章で述べたモデル化では接触特徴だけでなく注視位置や頭部運動といった情報も対応付けに有効と考えているが、ここではまず単純な接触特徴だけを用いた場合にどれだけ正しく冗長な操作を取り出せるのかを検証する。複数のモダリティを併用した手法に関しては今後の課題としたい。提案手法では2つの操作の接触領域の重複の有無を

$$c(i, j) = \begin{cases} 1 & \text{if } |u(M_{exp,i}) - u(M_{nor,j})| < T_h, |v(M_{exp,i}) - v(M_{nor,j})| < T_h \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$

なる条件で判別して操作の対応付けを行う。ここで  $(u, v)$  は4.2節で述べた手法で取り出した接触位置を示す。また  $T_h$  は判別に用いる閾値、 $S$  は4.2節で述べた方法で取り出した接触面積を示しており、ミシンの構造上の特性より

$$T_h = \begin{cases} 7.5 & \text{if } S < 23 \\ 20 & \text{if } 23 \leq S \leq 52 \\ 140 & \text{if } S > 52 \end{cases} \quad (5.2)$$

と定めた。これは、接触面積が大きい接触対象ほど、接触位置のばらつきが大きくなるためである。提案手法では考えうる  $(i, j)$  のペアのそれぞれに対して対応の有無を判別するため、単一の上級者操作  $M_{exp,i}$  に対して複数の初中級者操作  $M_{nor,j}$

が対応づく可能性がある。初中級者が特定のプッシュボタンを何度も押すなどがその例である。そこで1対1で対応付いた操作があれば、それを基準として生起順序に矛盾があるペアを消去する。この追加処理は

$$c(i, j) = 0 \quad \text{if} \quad (i - m)(j - n) < 0 \quad \text{st.} \quad c(m, n) = 1, c(m, j \neq n) = 0 \quad (5.3)$$

のように定式化される。5.1節で作成したデータベースを入力とし4章で述べた方法で接触特徴を抽出して上記対応付けの処理を行った。

### 5.3.2 結果と考察

表5.2に初級者の操作に対応付けた結果、表5.3に中級者の操作に対応付けた結果を示す。縦軸は上級者の操作、横軸は比較対象者の操作、表中のセルの値 $\{0, 1\}$ は操作同士の対応付け $c(i, j)$ を表している。なお $-1$ は一度対応付くと判断されたものの、他の1対1で対応付いた操作を元に順序矛盾から最終的に非対応と判定されたことを示す。また、赤で塗りつぶされているセルは人手で与えられた対応付けの真値である。

初級者の操作は上級者よりも非常に多く、余分な操作が多いことがわかる。全く無関係な部分に触れている操作も見られるが、知識やこれまでの機器操作の経験を利用して正しい操作を想像した結果、上級者と同じ部分に触れるときもある。ただそのタイミングや扱い方が適切ではないため、冗長な操作になってしまっている。そのため、接触特徴だけからは1つの上級者操作に対して複数の初級者操作が対応付く結果となった。また、このデータセット対では全ての上級者操作に対して複数の初級者操作が対応付けられたため、1対1対応と順序矛盾に基づいたフィルタリングを施すことはできなかった。

一方、比較対象が中級者になると余分な操作数が激減しているが、そもそも縫製タスク中に同じ部分に触れる回数が複数存在する（スタートボタン・ストップボタン・レバーなど）ため、接触特徴だけでは対応付けを1対1まで絞り込むことは難しかった。なお上級者と中級者の対応付けには、1対1の対応が縫製速度調整のつまみに対して検出できたため、生起順序に矛盾がある対応を消去できている。しかし、順序関係の曖昧性（どちらを先に行ってもよい）を考慮していないため却って誤った対応付けを導いてしまった。具体的には、中級者の針上下ボタンを押すという操作は、上級者の最後の操作と対応付けられてしまい、中級者のそれ以降の操作が上級者のどの操作とも対応付かないことになってしまった。

対応付けの性能を，検出率（上級者操作を比較対象者操作から探し出せた割合）および，適合率（上級者操作と対応付くと判定された操作の正しさ）の観点から定量的に評価したところ，上級者と初級者の対応付けでは，検出率が80%，各操作の適合率の平均が37%となった．一方，上級者と中級者の対応付けでは，検出率が91%，各操作の適合率の平均が41%となった．

検出率が100%に至らなかった理由は，初級者については説明した押すべきボタンを押さずに作業を進めたためである．中級者については，一度対応付いたものの1対1対応と順序矛盾によるフィルタリングによって対応付けが消去されたためである．適合率がどちらの比較でも低い値となったことは，接触特徴だけで同じ操作かどうかを判断することは難しいことを意味している．それは，装飾縫製のダイヤルや縫製速度調整つまみといった1度しか現れない操作の適合率が高く，スタート・ストップボタンや針上下ボタンのように何度も行う必要がある操作の適合率が低いという結果からも見て取れる．接触特徴だけでなく，注視位置や頭部運動の情報を特徴量として用いることができれば，適合率が向上すると考えられる．

## 第6章 結論

本研究では、そのままでは冗長で参照しづらい機器操作体験映像を有効利用するためには、どのように構造化を行えばよいかについて検討した。複数の機器操作間に共通する操作や反対に異なる点に着目して構造化を行えば、様々な用途に役立つマニュアルを作成できると考え、具体的な機器操作としてミシンを使った縫製作業を想定した機器操作モデルを設計した。これは、一連のタスクを構成する小さなタスクを状態、状態遷移時の出力シンボルを操作とみなした状態遷移モデルである。また、利用者の熟練度の差が操作方法に大きな影響を与えることから、それらの分析を通して共通操作・異なる操作を自動的に識別するための操作特徴として機器への接触位置・注目位置・頭部運動を提案した。

評価実験では、3段階の熟練度に対応した機器操作体験を頭部に搭載したカメラで記録してデータベースを作成した。このデータを用いて操作の多様性を分析した結果、接触位置や注目位置、頭部運動が機器操作体験の構造化を行うための特徴量として妥当であることが確認された。また、接触特徴を用いて、上級者の操作と初中級者の操作を自動的に対応付ける方法を提案した。その結果、上級者の操作のほとんどを検出できた一方で、重複した対応が多く残ってしまうことも確認された。

今後の課題には次のようなものが考えられる。まず、自動対応付けの適合率を上げるために、接触位置に加えて注目位置や頭部運動といった特徴量を用いた自動対応付けを検討する必要がある。ただし、それらの特徴量をそのままに用いては熟練度間の違いがむしろ大きくなってしまう。操作だけでなく機器からの反応も含めた特徴量を設計すべきではないかと考えている。データベースの多様性についてもまだ十分に担保できていない。本研究で記録した高々3パターンでは統計的な分析ができないため、より多くの被験者による機器操作体験記録を収集することが求められる。なお、提案したモデルに基づいた自動構造化には状態遷移を確率的に扱う手法の1つである隠れマルコフモデルの適用を現在進めている。

## 謝 辞

本研究を進めるにあたり，多大なご指導をいただいた京都大学工学研究科電気工学専攻複合メディア分野の中村裕一先生に深く感謝いたします。

また，本研究の細部に至り，多くのご指導，ご助言をいただきました工学研究科電気工学専攻複合メディア分野の近藤一晃先生に深く感謝いたします。

また，研究室での生活や，本論文の作成など，様々な点でご協力いただいた研究室の皆様方に深く感謝いたします。研究上の様々な事務処理をしていただいた秘書の小幡佳奈子さんに深く感謝いたします。

## 参考文献

- 1) 堀鉄郎, 河崎晋也, 石川尊之, 相澤清晴. ライフログ応用に向けたコンテキストに基づく映像・データ検索 (ヒューマンコミュニケーショングループ (hcg) シンポジウム). 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎, Vol. 103, No. 745, pp. 55–60, 2004.
- 2) Ig-Jae Kim, Sang Chul Ahn, Heedong Ko, and Hyoung-Gon Kim. Automatic lifelog media annotation based on heterogeneous sensor fusion. pp. 703–708, 2008.
- 3) 上岡隆宏, 河村竜幸, 馬場茂幸, 吉村真一, 河野恭之, 木戸出正継. 日常生活における物探し支援のためのウェアラブルカラーカメラ. 情報処理学会シンポジウム論文集, Vol. 2005, No. 4, pp. 85–86, 2005.
- 4) 河村竜幸, 福原知宏, 村田賢, 武田英明, 河野恭之, 木戸出正継. 対象物に「触れる」行為と記憶の遍在化による日常記憶支援. 電子情報通信学会論文誌 D, Vol. 88, No. 7, pp. 1143–1155, 2005.
- 5) 津田侑, 森幹彦, 近藤一晃, 小泉敬寛, 喜多一, 中村裕一. 環境貢献ワークショップにおける個人視点映像の利用法. 人工知能学会全国大会論文集, Vol. 25, pp. 1–4, 2011.
- 6) 相澤清晴, 石島健一郎, 椎名誠. ウェアラブル映像の構造化と要約: 個人の主観を考慮した要約生成の試み. 電子情報通信学会論文誌 D, Vol. 86, No. 6, pp. 807–815, 2003.
- 7) Y.Nakamura, J.Ohda, and Y.Ohta. Structuring personal activity records based on attention-analyzing videos from head-mounted camera. In *15th Int'l Conference on Pattern Recognition Track4*, pp. 209–213, 2000.
- 8) 上野智史, 内藤整. 設備変化検出に基づく一人称視点映像の作業シーン分割の一検討 (一般セッション, 新メディア, 及び安心・安全社会に寄与するパターン

- 認識). 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 114, No. 356, pp. 75–78, 2014.
- 9) 三角徹, 松家好希, 緒方広明, 矢野米雄. 体験映像と実世界オブジェクトのリンクによるユビキタス学習支援 (ユビキタス・モバイル学習環境/一般). 電子情報通信学会技術研究報告. ET, 教育工学, Vol. 108, No. 470, pp. 205–210, 2009.
  - 10) Dima Damen, Teesid Leelasawassuk, Osian Haines, Andrew Calway, and Walterio W Mayol-Cuevas. You-do, i-learn: Discovering task relevant objects and their modes of interaction from multi-user egocentric video. Vol. 2, p. 4, 2014.
  - 11) 玉宮義之, 林安紀子, 田代幸代. 教育経験と教室における注意: 実際の教育場面における視線計測を通じて. *JCSS Japanese Cognitive Science Society*, pp. 218–220, 2014.
  - 12) 後藤彰彦, 山本晃嘉, 濱田泰以, 高井由佳. 車椅子介助における介護者の眼球運動解析. 日本人間工学会大会講演集, Vol. 48, No. 0, pp. 250–251, 2012.
  - 13) Teesid Leelasawassuk, Dima Damen, and Walterio W Mayol-Cuevas. Estimating visual attention from a head mounted imu. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 147–150. ACM, 2015.

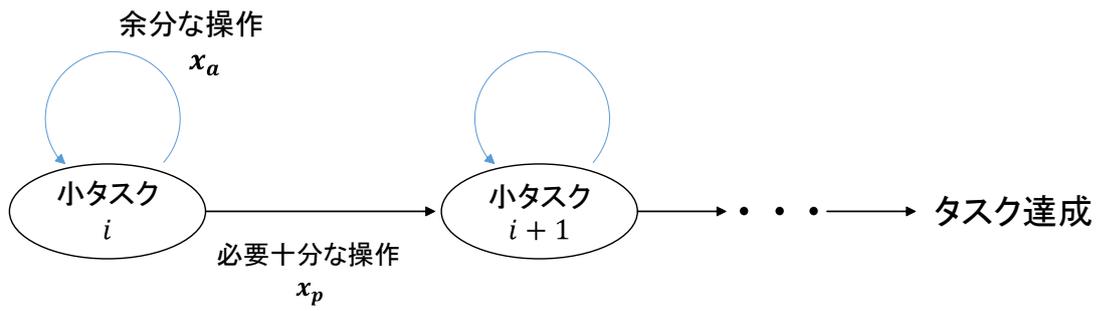


図 3.1: 機器操作の提案モデル



図 4.1: 記録機器

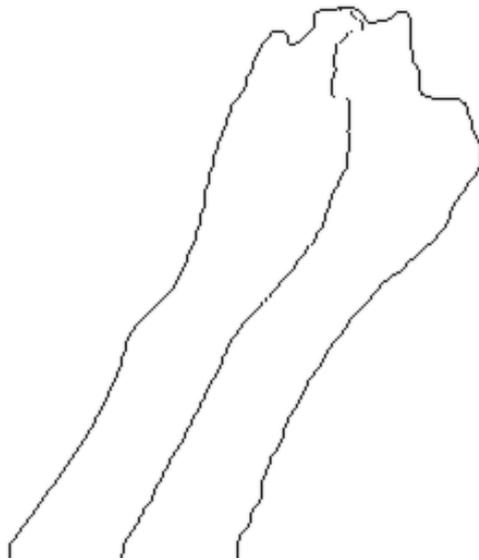


図 4.2: 中点の分布

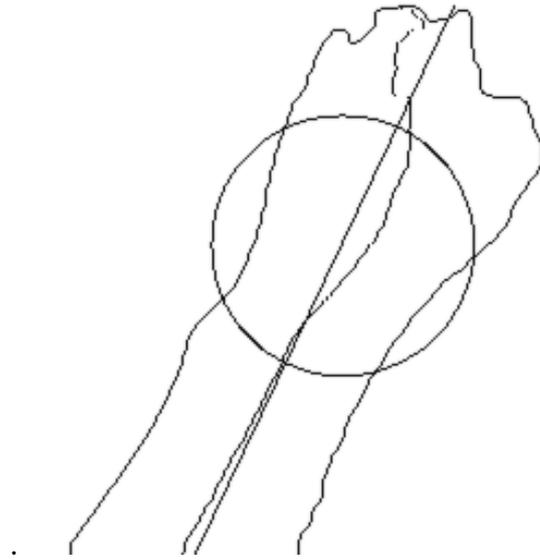


図 4.3: 中央線と円の作成



図 4.4: 前腕領域の除去

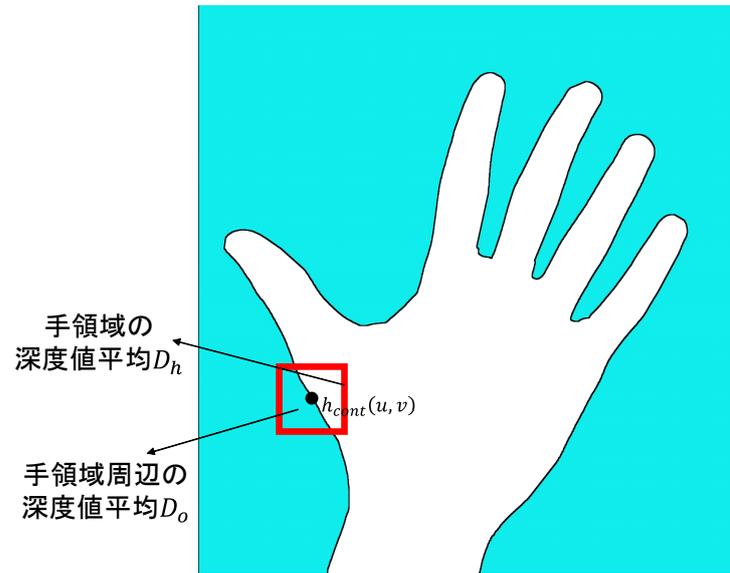


図 4.5: 深度値に基づいた接触検出



図 4.6: 接触検出

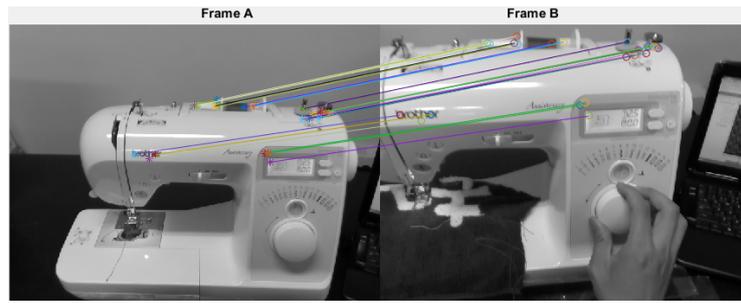


図 4.7: 異なる画像における特徴点の対応付け



図 4.8: 画像の重ね合わせ



図 4.9: 注視位置推定の様子



図 4.10: 回帰関数のパラメータ同定のためのサンプルデータの取得



図 5.1: 実験に用いた布

表 5.1: ミシンの操作手順

手順	内容
1	主電源を入れる
2	布抑えレバーを下げる
3	装飾縫製用のダイヤルを回し、番号「2」に設定する
4	針上下ボタンを押す
5	縫製速度調整のつまみを動かし、縫製速度を最低にする
6	スタートボタンを押し、縫製を開始する
7	ストップボタンを押し、縫製を停止する
8	布抑えレバーを上げる
9	布の向きを変える
10	布抑えレバーを下げる
11	スタートボタンを押し、縫製を開始する
12	ストップボタンを押し、縫製を停止する
13	布抑えレバーを上げる
14	針上下ボタンを押す



図 5.2: ミシンの操作カ所の名称

表 5.2: 上級者と初級者の対応付け

上級者\初級者	布	ダイヤル	操作キー	糸巻き軸	ミシン左上	布	ダイヤル	はずみ車	操作キー	スタート	つまみ	針上下	返し縫	はずみ車	布
装飾ダイヤル	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
針上下ボタン	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
速度調整つまみ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
スタートボタン	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
布	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
ストップボタン	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
布	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
スタートボタン	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
布	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
ストップボタン	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
針上下ボタン	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

ダイヤル	操作キー	糸通しレバー	天秤	返し縫	糸調子ダイヤル	つまみ	操作キー	布	スタート	布	ストップ	布	スタート	布	ストップ
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1
0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

表 5.3: 上級者と中級者の対応付け

上級者 \ 中級者	装飾ダイヤル	速度調整つまみ	針上下ボタン	スタートボタン	布	ストップボタン	布	スタートボタン	布	ストップボタン	針上下ボタン
装飾ダイヤル	1	0	0	0	0	0	0	0	0	0	0
針上下ボタン	0	0	-1	0	0	0	0	0	0	0	-1
速度調整つまみ	0	1	0	0	0	0	0	0	0	0	0
スタートボタン	0	0	0	1	0	1	0	1	0	1	0
布	0	0	0	0	1	0	1	0	1	0	0
ストップボタン	0	0	0	1	0	1	0	1	0	1	0
布	0	0	0	0	1	0	1	0	1	0	0
スタートボタン	0	0	0	1	0	1	0	1	0	1	0
布	0	0	0	0	1	0	1	0	1	0	0
ストップボタン	0	0	0	1	0	1	0	1	0	1	0
針上下ボタン	0	0	1	0	0	0	0	0	0	0	1