

モノづくりワークショップの振り返り支援システム

中江 一哉^{1,a)} 塚田 浩二^{1,b)}

概要: 近年, FabLab などのモノづくりコミュニティが主催する, デジタル工作機器を用いたモノづくりワークショップが頻繁に開催されている. こうしたワークショップでは, 制作物だけでなく, その制作過程も重要な体験である. しかし, 主催者/参加者ともに, 自身の作業に追われてしまうため, 制作過程を手軽に記録して振り返ることは難しい. 本研究では, こうしたモノづくりワークショップの振り返り動画の作成を支援するシステムを提案する. 本システムは, 「ハンズフリー/複数視点からの撮影機能」と「振り返りと一体化した動画編集機能」から構成され, 動画編集機能には「ツール認識」「笑顔認識」「両手作業認識」等の画像認識を用いたシーン探索機能を備える. 本論文では, システムの設計/多視点動画の分析/実装について述べる. さらに, 評価実験を通してシーン探索機能の性能等を検証する.

Support System to Review Manufacturing Workshop

KAZUYA NAKAE^{1,a)} KOJI TSUKADA^{1,b)}

1. はじめに

近年, 3D プリンターやレーザーカッターなどのデジタル工作機器の低価格化により, 一般の生活者が個人レベルでのモノづくりに気軽に携わることができるパーソナルファブ리케이션時代 (以下, ファブ時代) が到来しつつある. ファブ時代の到来に伴い, FabLab^{*1}などのモノづくりコミュニティが主催するモノづくり体験ワークショップが多数開催されている (図 1).

こうしたワークショップでは, 制作物だけでなく, デジタル工作機器を使った制作過程も重要な体験となる. しかし, モノづくりワークショップの多くは, 主催者が複数人の参加者を相手に対応する必要があるため, ワークショップをスムーズに進行させるための準備をする必要があるなど, 制作過程の撮影や動画の編集まで手が回らないことが多い. また参加者は自身の作業に集中してしまうため, 制作過程の振り返りを意識することは少ない. このように振り返りが行われていないという状況は, 参加者/主催者の



図 1 モノづくりワークショップの例

双方にとって大きな機会損失であるといえる. そこで, 制作過程を映像などで手軽に振り返る方法があれば, (1) 参加者にとっては, 思い出の振り返りや自らの技術の向上に役立てたり, 将来的な制作意欲の向上に繋がり, (2) 主催者にとっては, ワークショップの体制/進行の振り返りや, ワークショップの様子を Web 等で共有する際のプロモーション動画の作成, さらにワークショップのマニュアル作成等に幅広く活用できると考えた. 本研究では, モノづくりワークショップの制作過程を (半) 自動的に記録し, 振り返りと一体化した動画編集を実現する, モノづくりワークショップの振り返り動画作成支援システムを提案する.

¹ 公立はこだて未来大学
Future University Hakodate

a) g2116032@fun.ac.jp

b) tsuka@acm.org

*1 <http://fablabjapan.org/>

2. 関連研究

本章では、本研究に関連する研究事例について「モノづくり支援を目的とした研究」と「動画編集／視聴の支援を目的とした研究」、及び「画像処理によるオブジェクト検出手法」の3つの領域から紹介する。

2.1 モノづくり支援を目的とした研究

ファブ時代の到来に伴い、デジタル工作機械を用いた個人レベルでのモノづくりを支援することを目的とした研究が多く行われている。FabNavi[1]はモノづくりの過程を記録／共有／再生することでモノづくりを支援するためのシステムである。「見たまま、まねて、組み立てる」というコンセプトをもとに、あらかじめ写真や動画で記録したモノづくりの過程を作業機の上にセットしたプロジェクターから投影することで、映像通りに作業を行うだけで組立てが出来ることが特徴である。Sketch Chair[2]はレーザーカッターで出力可能な椅子を設計するシステムで、椅子を横から見た図をスケッチすることでシステムが重心を計算し倒れないように修正しつつ、出力に必要な部品のデータを作成することができる。PacCAM[3]はレーザーカッターなどの2次元データの作成を行う際に、効率よくパーツを配置するシステムで、パーツ間に余分なスペースがある場合に、自動で隙間が最小限になるように再配置するシステムである。これによって材料の節約と、出力時間の効率化に貢献している。また梅谷らは、モデルの重心がどこにあるか常に表示し、設計した家具が自立不能な場合や物理的に成り立っていない場合に、幾つかの修正案をユーザに提示するシステムを提案している[4]。このように、データの作成から出力、組み立てまで様々な工程の支援を目的とした研究が行われている。

2.2 動画編集／視聴の支援を目的とした研究

動画の編集を行うためには、専門の知識や経験が必要であり、初学者が動画編集を行うためには、多大な労力と時間を必要とする。こういった問題を解決するために動画の編集によって発生する負担を軽減することを目的とした研究が数多く行われている。CinemaGazer[5]は、与えられた字幕情報をもとに動画の重要度が高い箇所と低い箇所の推測を行い、再生速度を変化させることで短時間での動画の視聴を可能にするシステムである。CookSum[6]は手軽に閲覧可能な「動画レシピ」を作成するためのシステムである。撮影した料理動画の音声情報から料理中の重要なタスク検出を行い、検出されたタスクに基づいて自動で字幕ファイルを作成する、そして情報の少ない部分を高速再生することにより、短時間でレシピの概要を伝える高速動画レシピを作成する。DemoCut[7]は、ユーザが指定したフレームの分析を行い、不要箇所の高速再生やカット、重要個

所の強調表示などの処理を半自動で行うことで、動画編集を支援するシステムである。QuickCut[8]は、ナレーションと動画フレームの関連性をユーザが選択することで、フレームの重要度を判断して不要部分をカットするシステムである。EgoScanning[9]は、撮影された一人称視点の映像を高速再生しながら視聴する際に、手の映り込みや、対話部分など閲覧者が着目したい箇所の再生速度を自動的に落として再生するシステムである。これにより、長時間の一人称視点の映像における、必要な情報へ素早いアクセスを可能にした。

2.3 画像処理によるオブジェクト検出手法

画像処理技術を用いた、オブジェクト検出の研究事例を紹介する。AI Goggles[10]はカメラを備えたゴーグルとタブレット型計算機とHMD(Head Mount Display)を用いたウェアラブルシステムで、カメラに映ったものが何であるかを推測し、HMDで表示することができる。追加学習も可能であり、学習されていない物体をカメラで撮影することで、新たな学習対象として追加することができる。すでに学習している物体でも、学習データを追加することで認識の精度を高めることができる。Pixel-Level Hand Detection[11]では、一人称視点の映像において、撮影者の手領域の認識を行っている。色情報だけでは背景に肌色が含まれた場合に、誤認識してしまうという課題を解決するために、色に加えて、テクスチャーを学習要素とすることで、高精度な手の認識を可能にした。

2.4 本研究の特徴

本研究は、モノづくりワークショップに焦点を当て、参加者がワークショップを振り返りながら動画の編集を行う、動画編集支援システムを提案する点が特徴である。

3. 提案

本システムは、以下の2つのコンセプトをもとに開発を行う。

- (1) ハンズフリー／複数視点での撮影
- (2) 振り返りと一体化した動画編集

以下、それぞれについて詳細を述べる。

3.1 ハンズフリー／複数視点の撮影

1つ目は、撮影をハンズフリー／複数視点で行うという点である。モノづくりワークショップの振り返りを行う機会が少ない原因の一つとして、主催者/参加者ともに、自身の作業に集中してしまうため、制作過程を手動で撮影し続けることが難しいという点がある。そこで制作過程の動画をハンズフリーで撮影することで、主催者/参加者ともに自身の作業に集中できるよう配慮する。複数視点からの撮影については、前述したハンズフリーである点に加えて、

カメラの設置しやすさと、動画から得られる情報の有効性を考慮して、図2の位置関係でカメラ（スマートフォン）の配置を行う。すなわち、作業机の上部に固定した俯瞰視点、ワークショップ参加者の反対側に固定した第三者視点、そして参加者の胸部に固定した参加者視点の三点である。俯瞰視点では、ミスミフレームで組んだ筐体の先に自作の固定具を用いてスマートフォンを固定する。また、第三者視点の固定には小型の三脚を、参加者視点の固定には首から吊るせる自作のスマートフォンケースを利用する。図3に各視点から撮影した映像の事例を示す。複数視点からの映像を用いることで、ワークショップを多角的に振り返ることができる。と考える。

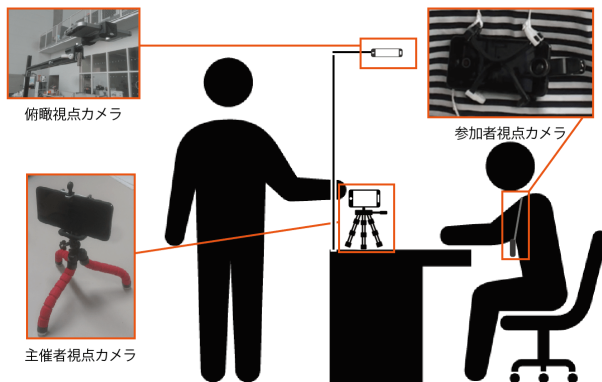


図2 各カメラの位置関係図



図3 各視点の映像例。(1) 俯瞰視点、(2) 第三者視点、(3) 参加者視点。

3.2 振り返りと一体化した動画編集

2つ目は、振り返りと一体化した動画編集が可能であるという点である。システムのユーザは、撮影された複数視点の動画をタブレットなどで視聴しながら、関心の高い箇所を選択することで動画の切り出しを行う。これにより、参加者が自身のワークショップの様子を振り返りながら、関心が高い箇所を集めたワークショップの振り返り動画を作成することができる。詳細なインターフェースや機能については第5章で記述する。

4. 動画分析

本章では実際のワークショップで撮影した動画分析の目的／手法／結果について述べる。

4.1 目的

動画を分析する目的は、俯瞰視点、第三者視点、参加者視点で記録された映像に対して、記録された映像情報と実際の作業内容を分析することで、各視点の映像の特徴や適切な画像認識技術を調査することである。

4.2 手法

ここでは、対象としたワークショップの内容と、動画分析の手順について説明する。表1に、ワークショップの概要について示す。3つのワークショップの撮影と分析を行ったが、本稿は分量の関係上、型押しレザーストラップワークショップに絞って説明する。

これらのワークショップにおいて、実験者は図2に示す3つのスマートフォン（カメラ）を導入し、俯瞰視点／第三者視点／参加者視点から撮影を行った。また、参加者／主催者には特にカメラや撮影を意識しないように指示しており、実験者はカメラを設置後はワークショップの進行には一切介入しなかった。全ての動画は本システムを用いて実験者が内容を確認／分析した。これらのワークショップに対して、ワークショップ全体の流れを時系列でまとめたものに撮影動画の分析を行い、各視点の特徴や作業内容との対応について考察する。

4.3 結果

ここでは、型押しレザーストラップワークショップ動画の分析結果を中心に説明する。このワークショップは、主催者が自作した複数のレザーパーツと押し型を組み合わせたストラップ制作を行う。参加者は、初めにレザーの色／形状と押し型を選択し、プレス機を用いてレザーに型押しをする。その後、革紐でレザーパーツ同士を結合させ、ストラップを組立てる。細かい作業が多いため、作業中は主催者が参加者を見守り、常に指示を出していた。図4は参加者の作業をタイムラインとしてまとめたものである。以下、各視点の映像について分析を行う。



図4 型押しレザーストラップワークショップのタイムライン。画像横の番号は、図a~lに対応している。

4.3.1 俯瞰視点

俯瞰視点では、図5(a)のように机の上の作業領域のほぼ全体が写っており、ワークショップで利用している部品や道具を一望できる。特に今回は研究室での開催で環境を整えやすかったため、机の上が整理されており、利用し

作業内容	所要時間	撮影人数	参加者の属性	開催場所
型押しレザーストラップ	約 10 分	3 人	大学生	大学研究室
3D ペンワークショップ	約 40 分	2 人	親子連れ	はこだてみらい館
3D モンスター作成ワークショップ	約 20 分	3 人	子供	はこだてみらい館

表 1 分析対象としたワークショップの概要. 本稿では型押しレザーストラップワークショップに絞って説明する.

ている道具や部品等をはっきりと確認出来る. プレス機やはさみなどの工具を使用している間は手で隠れているため (図 5 (b) (d)), あらかじめワークショップで使用される工具リストを作成し, 写っている工具の認識を行うことで, 使用している工具を推測できる可能性がある.

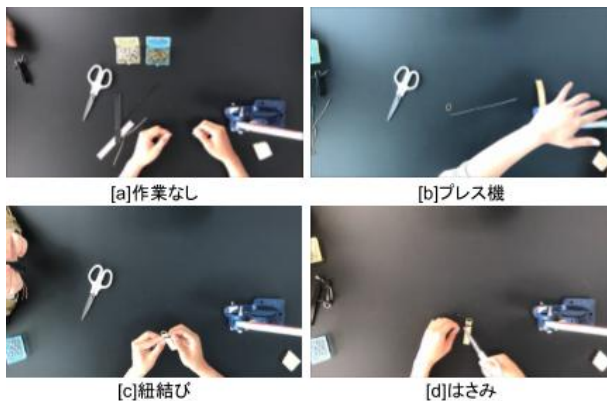


図 5 俯瞰視点の映像例

4.3.2 第三者視点

第三者視点では, 図 6 (e) のように, 参加者の様子が常時写っており, 表情や作業状態を確認出来る. 例えば, 参加者が笑顔浮かべているシーン (図 6 (f)) や, 完成して拍手をしている状態などをはっきり確認することができた. また, 今回は主催者が常駐しているため, 参加者が主催者に向かって質問をする場面が頻繁に確認された (図 6 (g)). よって参加者の表情の認識を行うことでワークショップ中の笑顔シーンや, 会話シーンなどの抽出ができる可能性がある. また, 作業中に場所の移動があるワークショップでは, 顔の検出により, 参加者の有無を推測できる.



図 6 第三者視点の映像例

4.3.3 参加者視点

次に, 参加者視点では, 図 7 (k) のように, 参加者の手元が大きく映し出されており, 手に持った道具や, 手元の細かな作業等を確認できる. 主催者が作業内容を説明している間は, 参加者の手は机の上に置かれているため, 映り込みは小さい (図 7 (i)). それに対して参加者が実際にプレス機を使用していたり, 紐を結ぶなどの作業をしている間は, 参加者の手が大きく映り込んでいる (図 7 (j) (k)). このように参加者視点映像における参加者自身の手の映り込みによって, 作業を行っているかどうかや, 作業内容の推測が行える可能性がある.

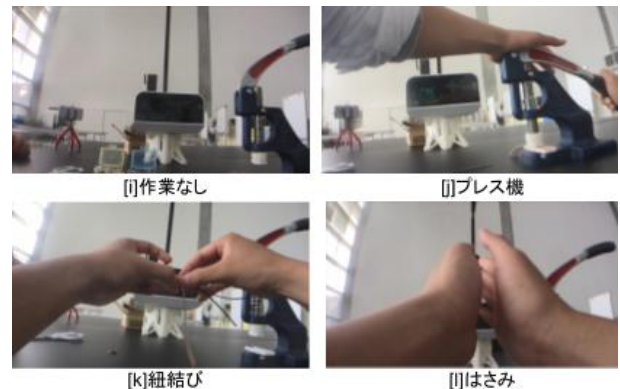


図 7 参加者視点の映像例

5. 実装

本章では動画編集システムの実装について述べる.

5.1 動画編集システム

ここでは, 振り返り動画編集システムの詳細を説明する. 具体的には, システムの基本的な機能と, 前章の動画分析に基づいて実装したシーン探索機能について述べる. システムは Windows タブレット上で動作するソフトウェアであり, ユーザは基本的にタッチパネルを通して操作する. システムのインターフェースは, 動画表示部とシーン探索部の 2 つから主に構成されている (図 8). まず, 「SELECT」ボタンを押してワークショップ動画が保存されているフォルダを選択することで, 3 つの視点の動画が動画表示部に自動的に読み込まれる. 次に, ユーザは再生された動画を視聴し, ワークショップ工程を振り返りながら, 興味の有る箇所/重要と思う箇所をタップすることで, タイミング情報/視点情報を記録できる. さらに, 注目動画 (大きく表示

される動画)を任意に切り替えることでも、視点情報は記録される。動画の振り返りが終了したら、「CREATE」ボタンを押すことで、システムはタイミング情報/視点情報に基づいて、各視点の動画の切り出しと結合処理/フェード処理を行い、一つの振り返り動画を作成する。

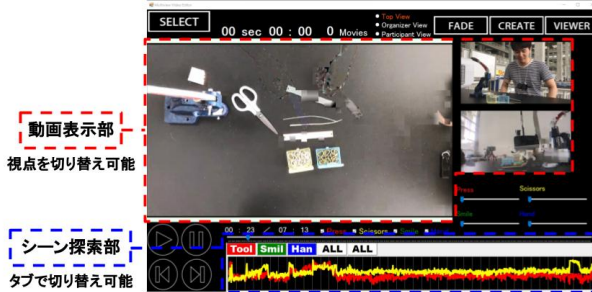


図 8 本システムのインターフェース

5.2 シーン探索機能

ここでは4章の動画分析の結果に基づいて実装したシーン探索機能について説明する。シーン探索機能は、シークバーと同期しており、各動画の様々な情報を視覚化することで、振り返りに必要なシーンの探索を支援する機能である。シーン探索機能として、工具利用検出(俯瞰視点)、笑顔検出(第三者視点)、両手作業検出(参加者視点)の3つの機能を実装した。またこれらの機能はタブで切り替えることが可能であり、必要に応じて組み合わせて利用することができる。

5.2.1 工具の映り込みによる探索支援機能

俯瞰視点映像に対して工具の映り込みを検出することで、現在使用している工具を推測することが目的である。動画分析の結果より、俯瞰視点の映像には工具が常時映し出されており、工具が使用されるタイミングで工具が手で隠れてカメラの視界から消えるという点に着目した(図9)。工具の検出手法としてSIFTを利用し、予め動画から切り出して作成したテンプレート画像と動画の各フレームのマッチングを行う。マッチングした特徴点が多いほど、画像内に対象の工具が映っている確率が高いと判断する。システム上でグラフを表示する際はマッチング数に-1をかけて反転させることで、ユーザが直観的に工具を使用している箇所(=工具が手で隠れている箇所)を認識しやすいようにした。

5.2.2 笑顔による探索支援機能

第三者視点の映像に対し、参加者の表情の認識を行うことで、各フレームに対し笑顔のスコアを決定する。笑顔スコアを時系列に沿って可視化することで、笑顔シーンへのアクセスが容易になり、参加者は自身の楽しんでいる様子を、主催者は参加者の笑顔シーンを集約したプロモーション用の動画等を作成することができる。表情の認識には



図 9 工具のマッチング例

物体認識やパターン認識などが行えるライブラリであるOpenCV^{*2}のhaarcascadesライブラリを使用した。以下に表情の認識の手順を示す。今回使用した学習用のデータ(xmlファイル)はOpenCVと共に配布されているものを使用した。なお、複数の顔領域を検出した場合は、最も高い得点を笑顔スコアとして決定する。

- (1) 画像全体から顔の領域を検出する(haarcascade_frontalface_alt.xml)
- (2) 検出された顔領域の中から目の領域を検出する(haarcascade_eye.xml)
- (3) 目が検出された顔領域に対して目尻の角度や口角によって笑顔スコアを0~100で決定する(haarcascade_smile.xml)



図 10 笑顔認識結果例

5.2.3 手の映り込みによる探索支援機能

主観視点の映像に対し、手の映り込みを検出することが目的である。動画の分析結果より、手の映り込みが大きいシーンの多くは、3Dペンを使用しているシーンや、革紐を結んでいるシーンなどの作業である。すなわち、実際にワークショップの参加者が制作作業を行っているシーンの探索に利用できると考えた。以下に手の映り込み認識の手順を示す。本機能では、肌色面積の割合によって、手の映り込みを推測する。色の検出はHSV色空間を用いて行う。フレーム画像の各ピクセルに対し、肌色の閾値(色相: $15 < \text{Hue} < 35$, 明度: $58 < \text{Saturation} < 173$, 彩度: $88 < \text{Value} < 255$)を暫定的に設定し、含まれている場合は1, 含まれていない場合は0とし、ピクセル全体に対する1の割合が大きいほど、手が大きく映り込んでいると判断する。ただし、この手法では、手以外の背景などを誤認識してしまい、ノイズとなる危険性がある。そのため肌色部分のみを抽出した画像に対し収縮処理をかけてノイズの除去を行った後に、膨張処理を用いて、元のサイズに戻す必要がある。図11は実際の参加者視点の画像に対して、

^{*2} <https://opencv.org/>

収縮処理を3回かけて、ノイズを除去したのちに、膨張処理を3回行ったものである。このように、処理前は背景の段ボールの色などのノイズが含まれているが、処理後にはノイズを軽減できている。

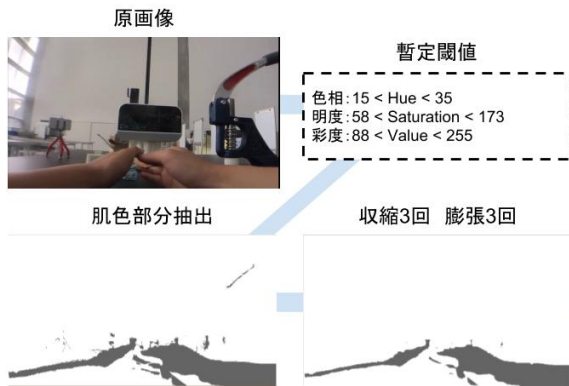


図 11 肌色領域の検出例

5.2.4 シーン探索機能の組み合わせ利用

複数のシーン探索機能を組み合わせて使用できる機能を実装した。単体でのシーン探索と異なり、「参加者の手の映り込みが大きく、対象の工具を使用している」のようにより具体的なシーン探索が可能になる。図 12 の上部に、前述した3つのシーン探索機能を組み合わせた例を示す。さらに、同時に複数の結果を表示する場合、表示が煩雑になり視認性が悪くなることを考慮して、システムの検出結果に対して任意の閾値を設定することで、「検出の有無」のみを二値化して表示する機能を実装した。図 12 の下部は、同図の上部の結果を二値化して表示した例である。

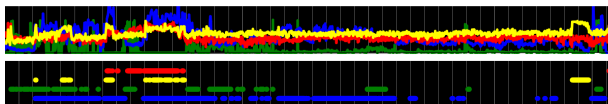


図 12 シーン探索機能の組み合わせ利用例（上）と二値化表示例（下）。赤：プレス機，黄：はさみ，緑：笑顔，青：両手作業。

6. 評価実験

ここでは、シーン探索機能の機能評価実験について述べる。

6.1 目的

実験の目的は、実装したシーン探索機能（工具利用検出機能／顔検出機能／両手作業検出機能）が、目的のシーンを検出できているかを評価する。工具利用検出機能では対象の工具を使用しているシーン、笑顔検出機能では、ワークショップ参加者が笑顔を浮かべているシーン、両手作業検出機能では、工具や両手を使用して作業を行っているシーンをそれぞれ正しく検出できているか評価する。ま

た、シーン探索機能の二値化処理に用いる最適な閾値についても検討する。

6.2 手法

実験の手法は、各システムによって計算された結果（実験データ）と、実験者が動画を見ながら手動で行ったタグ付け結果（正解データ）を比較する。システムで計算された値は任意の閾値を用いて二値化処理を行い、正解データと比較する。

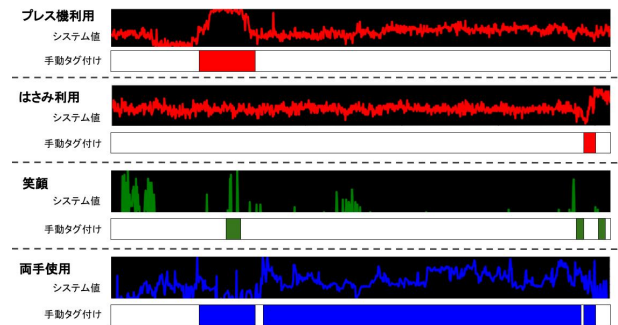


図 13 システムの検出データの一例と正解データ

閾値は複数用意した上で、それぞれの結果に対し、適合率（正解と予測して、実際に正解である割合）と再現率（実際に正解であるもののうち、正解と予測できた割合）の2つの値から機能の信頼性を評価する。実験に使用する動画データは、動画分析で用いた型押しレーザーワークショップの事例から2つの動画データを使用した。

6.3 結果

6.3.1 工具利用検出機能

ここでは、工具利用検出結果について述べる。本実験では、型押しレーザーワークショップで使用されているプレス機と、はさみの利用検出を行う。システムが計算した値は、サンプル画像とのマッチング数であり、対象の工具を使用している場合は、手で隠れるためマッチング数は少なくなる。表2にプレス機の検出結果を、表3にははさみの検出結果をそれぞれ示す。なお、括弧内の数字はシステムが計算したマッチング数データに対して二値化処理を行う際に用いた閾値である。

6.3.2 笑顔検出機能

ここでは、笑顔検出結果について述べる。実験者が第三者視点の映像を視聴して、参加者が笑顔であると判断した箇所に対してタグ付けを行った。その後システムの検出結果を閾値を用いて二値化したデータと、手動でタグ付けした正解データを比較した結果を表4に示す。

6.3.3 両手作業検出機能

手の映り込みが大きいシーンと、参加者が手元で作業を行っているシーン（プレス機を使用しているシーン、紐を

表 2 プレス機使用検出結果

	適合率／再現率（動画 1）	適合率／再現率（動画 2）	適合率／再現率（平均）
マッチング数 24	0.79/0.81	0.88/0.99	0.84/0.90
マッチング数 21	0.92/0.77	0.95/0.94	0.94/0.86
マッチング数 18	0.99/0.71	0.98/0.92	0.99/0.81
マッチング数 15	0.99/0.69	0.99/0.70	0.99/0.70

表 3 はさみ使用検出結果

	適合率／再現率（動画 1）	適合率／再現率（動画 2）	適合率／再現率（平均）
マッチング数 12	0.54/1.00	0.92/0.92	0.73/0.96
マッチング数 10	0.63/0.94	0.98/0.87	0.81/0.91
マッチング数 8	0.82/0.76	0.99/0.77	0.91/0.77
マッチング数 6	0.97/0.71	0.99/0.64	0.98/0.68

結んでいるシーン、はさみを使用しているシーン）を比較した結果について述べる。画像全体に対して肌色部分が占める割合を閾値として、二値化処理を行った。正解データと比較した認識結果を表 5 に示す。括弧内の数字はシステムが計算した肌色面積の割合に対して、二値化処理を行う際に使用した閾値である。

6.4 考察

本章では、実験結果に基づいた考察について述べる。表 2、3 より、工具の映り込みによるシーン探索では、再現率は 64～100 % に分布しており、実際に工具を利用しているシーンの検出に概ね成功している。一方、ハサミの事例に注目すると、動画 1 において適合率が 54～97 % と低い問題が発生した。これは、図 14 のように、他の工具を利用する際に手で対象の工具が隠れてしまうことがあったためである。この問題を解決するために、工具の認識に俯瞰視点だけでなく参加者視点映像を組み合わせたり、工具のレイアウトを工夫する等の手法を検討したい。また、図 14 では、動画のはじめと終わり部分でも大きい値が記録されているが、これは動画の撮影時にスマートフォンに直接触れて撮影を開始しているため、振動によって映像がぶれているためであると考えられる。リモコンシャッターを利用したり、最初／最後のシーンをカットすることでこの問題は解決できると考える。

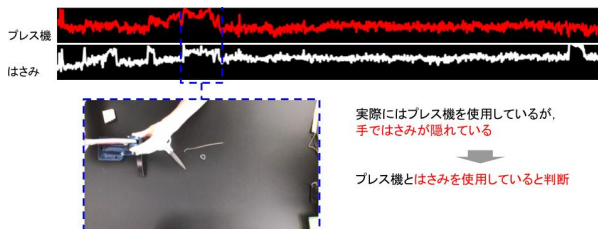


図 14 工具利用検出の失敗例

笑顔によるシーン探索機能では、表 4 より、閾値 20 と 50 においては再現率が 40～61% と低い結果となった。また閾値を最低値である笑顔スコア 0 にした結果、本来は再

現率がほぼ 100% になるはずであるが、実際は平均 77% という結果となった。この原因として考えられるのは、参加者が笑っているシーンで下を向いているなど、参加者の顔自体を正しく認識できていない場合があったことが挙げられる（図 15）。よって、カメラの設置方法や位置を見直すなどの工夫をして、安定して参加者の表情を記録する必要があると考える。



図 15 笑顔検出の失敗例

手の映り込みによるシーン探索機能では、適合率と再現率が低い値となり適切な値は見つけられなかった。実際の動画を確認した結果、映像に両手作業がしっかり記録されている場合は、肌色面積が 15% 以上であることが多いことが確認できた。そこで適切な閾値は 14～16% であると考えられるが、閾値を 16% 以上にした結果、再現率が小さな値になった。すなわち、両手作業をしているにもかかわらず、作業をしていないと判定される場合が多かった。動画 2 をもとに原因の分析を行った結果、紐を結ぶなどの両手を使用する作業を行っている箇所、値が小さくなっている箇所が散見された。これは参加者視点カメラが、作業機の下を映してしまい、作業スペースの記録ができていないことが一つの原因として挙げられる（図 16）。これは作業中の参加者の姿勢や、作業の癖などに依存していると考えられるため、装着型カメラの固定方法の見直しが必要だと考えられる。現在の首から固定具を吊り下げる手法ではなく、例えば、ハーネスのような専用固定具による撮影の検討が必要である。

また、全ての認識結果において、(1) 二値化の閾値を上げる程、適合率が上がり、再現率が下がる（≒実際に作業し

表 4 笑顔検出結果

	適合率／再現率（動画 1）	適合率／再現率（動画 2）	適合率／再現率（平均）
笑顔スコア 0	0.86/0.76	0.53/0.78	0.70/0.77
笑顔スコア 20	0.91/0.61	0.85/0.57	0.88/0.59
笑顔スコア 50	0.98/0.40	0.93/0.28	0.96/0.34

表 5 両手作業検出結果

	適合率／再現率（動画 1）	適合率／再現率（動画 2）	適合率／再現率（平均）
肌色面積割合 8%	0.62/0.93	0.57/0.94	0.60/0.94
肌色面積割合 12%	0.68/0.87	0.53/0.67	0.60/0.77
肌色面積割合 16%	0.85/0.56	0.53/0.49	0.69/0.53
肌色面積割合 20%	0.88/0.31	0.65/0.36	0.77/0.34

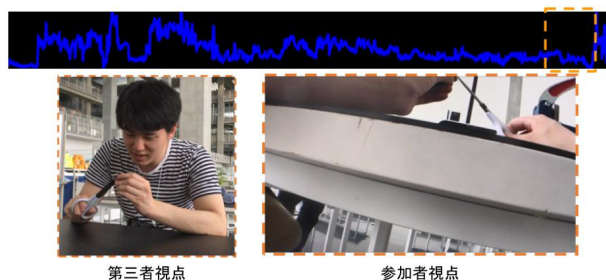


図 16 両手作業検出の失敗例

ている状況を見落とすケースが増える), (2) 閾値を下げる程, 適合率が下がり, 再現率が上がる (≒作業していない場面を誤検出するケースが増える) 傾向が見られた. よって, 使用目的に応じて柔軟に閾値を調整可能なインタフェースを今後設計していく. 例として, ユーザが手動で動画の切り出しを行う場合には, 誤検出部分は手動で省けるため, 閾値は低めに設定する. 今後システムが自動的に, 動画の切り出しを行う場合は, 全体の動画時間を短縮することを優先して, 閾値を高く設定する等の工夫が考えられる.

7. おわりに

モノづくりワークショップにおいて振り返りの機会が設けられることが少なく, 参加者／主催者ともに大きな機会損失であるという問題を解決するために, 本研究では, モノづくりワークショップの振り返り動画作成支援システムを開発した. システムは, ハンズフリー／複数視点での動画撮影, 振り返りと一体化した動画編集の 2 つのコンセプトを基に開発した. さらに, 振り返りに必要なシーンへ効率よくアクセスするためのシーン探索機能 (笑顔による探索機能, 手の映り込みによる探索機能, 工具の映り込みによる探索機能) の実装を行った. その後シーン探索機能の精度を明らかにするための機能評価実験を行った.

今後は, 実際のワークショップの参加者／主催者に対するユーザ評価実験を通して本システムの有効性を検証するとともに, 実際のワークショップでの利用を進め, データ収集を進めていきたい. 具体的には, 参加者／主催者それぞれの視点からの重要場面の切り出しデータ (≒人間が重

要と思う場面) を蓄積することで, コンピュータによるモノづくり動画の自動作成や, ワークショップのレシピ化の支援につなげていきたい.

謝辞 本研究は JST CREST の支援を受けた.

参考文献

- [1] K. Tsukada, M. Oki, K. Watanabe, and D. Akatsuka. FabNavi: Support system to assemble physical objects using visual instructions, Paper presented at Fab10, Vol. 2, pp. 2-8, 2014.
- [2] G. Saul, M. Lau, J. Mitani, and T. Igarashi. SketchChair: an all-in-one chair design system for end users, Proceedings of TEI'11. ACM, pp. 73-80, 2011.
- [3] D. Saakes, T. Cambazard, J. Mitani, and T. Igarashi. PacCAM: material capture and interactive 2D packing for efficient material usage on CNC cutting machines, Proceedings of UIST'13. ACM, pp. 441-446, 2013.
- [4] N. Umetani, T. Igarashi, and N. J. Mitra. Guided Exploration of Physically Valid Shapes for Furniture Design, Proceedings of SIG GRAPH 2012. ACM, 2012.
- [5] K. Kurihara. CinemaGazer: a system for watching videos at very high speed, Proceedings of AVI'12. ACM, pp. 108-115, 2012.
- [6] 橋本遼太郎, 塚田浩二, 栗原一貴. CookSum: 動画レシピ作成支援システム. WISS2014 論文集.
- [7] P. Y. Chi, J. Liu, J. Linder, M. Dontcheva, W. Li, and B. Hartmann. Democut: generating concise instructional videos for physical demonstrations, Proceedings of UIST'13. ACM, pp. 141-80, 2013.
- [8] A. Truong, F. Berthouzoz, W. Li, and M. Agrawala. Quickcut: An interactive tool for editing narrated video, Proceedings of UIST'16. ACM, pp. 517-528, 2016.
- [9] K. Higuchi, R. Yonetani, and Y. Sato. EgoScanning: Quickly Scanning First-Person Videos with Egocentric Elastic Timelines, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 6536-6546, 2017.
- [10] 原田達也, 中山英樹, 國吉 康夫. AIGoggles: Wearable Image Annotation and Retrieval System with Online Learning, 電子情報通信学会論文誌, pp. 1-13, 2009.
- [11] X. Zhu, X. Jia, and K. -Y. K. Wong. Structured forests for pixel-level hand detection and hand part labelling. Computer Vision and Image Understanding, pp. 95-107, 2015.