

ものづくりの過程における動画要約システムの作成

小菅 みのり (指導教員: 五十嵐 悠紀)

1 はじめに

ものづくりにおける過程は重要である。完成品だけでなく、技術を受け継ぐためにも、材料から出来上がっていくその過程を保存することは必要である。しかし、一つの作品が出来上がるまでに膨大な工程と多くの時間が費やされている作品も多く、それらをすべて見ることは難しい。そこで、我々は要約動画に焦点を当てた。要約動画によって、人に手作りの贈り物をする際の付属品として作品に物語を付けることで、作品に対する付加価値を与えるという効果も期待できる。技術を学ぶための教材として用いるなどの教育現場での使用も可能である。

近年増えている1~3分程度のショート動画は、要約動画の最たる例であり、世界的にも視聴数が多い。Alphabet社は2024年の第3四半期決算発表会¹で、毎月Youtubeにアップロードをするチャンネルの内70%がShort動画をアップロードしていると発表した。また、2023年の第3四半期決算発表²で、Short動画の1日の平均再生回数は700億回を超え、毎月20億人以上のユーザーが視聴していると発表した。また、Youtubeのデータ分析ツールであるKamuiTrackerを用いてEVIRY社が2023年にした調査³では、ショート動画の月間投稿本数推移は図1のようになっている。2021年1月から2022年12月の2年間でショート動画の投稿本数だけでなく、ショート動画の占める割合も増加しており、月間投稿数がおよそ10倍となっていることがわかる。しかし、動画編集は手間や負担が大きいいため、手軽に作成するのは難しい。

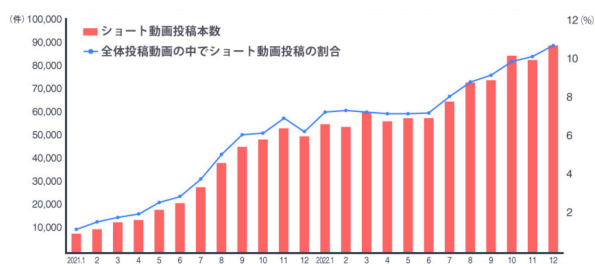


図1: Youtubeへのショート動画の月間投稿本数推移。

動画を要約する研究は数多く行われている。三浦らは、料理動画に特化した要約動画を作成する手法を提案した[1]。この先行研究では、動画の中で重要な部分を動きに基づいて抜き出し、料理番組の制作者による要約動画と、先行研究内の手法で作成された要約動画を比較して評価している。Chiらは、DIYのチュートリアル動画の編集を半自動で行うシステムを提案した[2]。このシステムでは、ユーザーが動画内に設置したマーカーに基づき、音声と映像の分析を行い、編集効果を適用している。沼澤らは、動画の内容を要約した

静止画の作成をする手法を提案した[3]。このシステムでは、入力動画から一定間隔で切り出したフレームを抽出し、重ね合わせることで一枚の静止画を出力している。

我々は、ものづくりの動画を要約した動画を自動で出力するシステムを提案する。提案システムでは、動画内のフレーム間の変化を計算し動きのないフレームを除去することにより、要約動画を出力する。本稿では、身近なものづくりの代表例である手芸を対象にシステムを開発した。ユーザーは要約動画の長さや変化の閾値を設定する。本稿では開発したシステムに、刺繍をする動画と、ぬいぐるみを作成する動画の2種類の動画をを入力してそれぞれの要約動画を作成し検証したため報告する。

2 要約動画作成システム

要約動画作成システムは、GoogleColab上でPythonを用いて実装した。

2.1 アルゴリズム

提案システムのアルゴリズムの概略図を図2に示す。3段階の工程で入力動画から要約動画を作成する。工程2つ目である「チャプター作成のための物体検出」では、ものづくりにおいて、糸や布を切るために場面転換のトリガーになりやすいハサミを、画像認識AIであるYOLOv10を用いて判定した。ハサミが出てきたフレームをチャプターとして登録することで、それぞれのチャプターに対して早送りなどの編集を追加することができる。

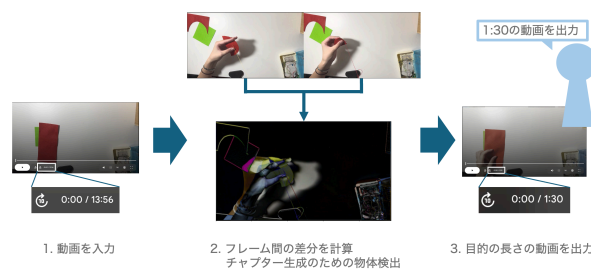


図2: 提案システムのアルゴリズム。

2.2 ユーザインタフェース

提案システムでは、ユーザーが自由に変更できる設定項目を、図3、4に示すように2つ設けている。

1つ目の項目(図3)は、システムの工程2つ目の「フレーム間の差分を計算」の段階で計算した差分をグラフにし、どのくらいの差分があるフレームを抽出するかを決定するための閾値をスライダーで調整可能とした。縦軸が差分の大きさ、横軸がフレーム番号、オレンジ色の線が差分の平均値、赤色の線が現在選択している差分の大きさである。スライダーの初期値は平均

¹<https://abc.xyz/2024-q3-earnings-call/>

²<https://abc.xyz/2023-q3-earnings-call/>

³<https://kamuitracker.com/blog/archives/17635>

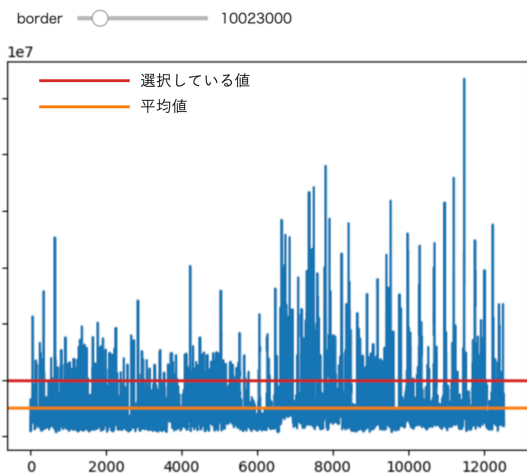


図 3: フレーム間の差分を示すグラフと差分の閾値を調整するスライダー。



図 4: 動画時間を入力するテキストボックス。

値とした。スライダー操作によって変化した値は、図に即座に反映され、ユーザが選択している値を確認することが可能である。

2つ目の項目(図4)は、工程3つ目である「目的の長さの動画を出力」の段階で、動画時間を入力によって決定するためのテキストボックスである。元動画より長い場合はその旨をユーザに指摘し修正を促すこととした。

3 要約動画作成結果と比較

我々は、刺繍をしている動画とぬいぐるみを作る動画を提案システムに入力した。これらの入力動画は10分程度の長さである。作成結果では、ユーザが自由に変更できるそれぞれの設定項目の入力に応じた、要約動画の生成ができた。

スライダーの初期値である、差分の平均値を用いて動画を生成した場合、658秒の動画が274秒に、830秒の動画が270秒になり、それぞれ4割と3割ほどの短さまで動画時間を縮めることができた。よって、ユーザが特に指定したい項目がない場合でも、半分以下の動画時間まで要約することが可能である。

1/5の長さに短縮したタイムラプス動画と、テキストボックスによる時間指定でタイムラプス動画と同じ長さの要約動画を生成し、その内容を比較した。図5は、それぞれの手法のフレーム抽出部分の図解と、動画内の同じフレーム番号の場面を抜き出した画像である。その結果、タイムラプスでは両手の写っていない、動きのないフレームが存在するが、提案手法ではそれが無いことが確認できる。

4 まとめと今後の課題

本稿では、ものづくりの過程の動画から自動で要約動画を作成できるシステムを提案した。現在は主にカット編集を自動で処理しているが、ズーム機能や、繰り返

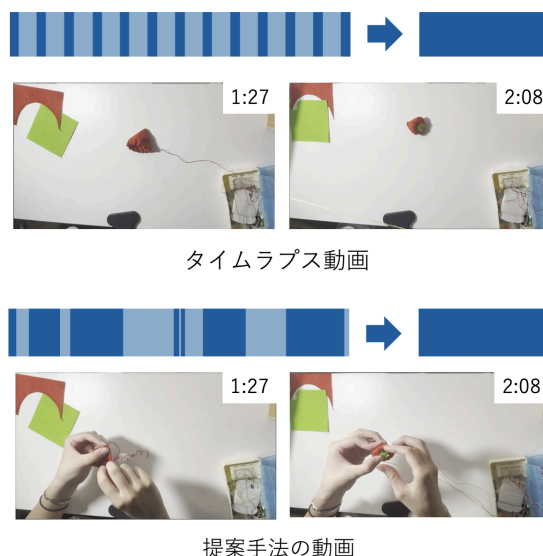


図 5: タイムラプス動画と提案手法による動画の比較画像。

返し、BGMの追加などの自動編集を追加することを考えている。また、ものづくりの過程で多く登場する繰り返しの動作について、1, 2回目がスロー、5回目以上が早送りなどのような、目的の部分にのみ特定の処理をすることができていない。よって、現在行っているハサミの検出だけでなく、さらに細かいチャプターの設定と、反復動作の検出やその回数のカウントを利用し、自動で多くの編集を施すシステムの実装を考えている。また、ものづくりにかかる時間は長く、現在のシステムでは1時間などの長い動画を処理するのに1時間以上かかってしまう。したがって、処理の速度を上げるために、チャプター分割を用いた動画読み込み方法の改善を行う予定である。さらに、ユーザ評価を行い、システムを用いて要約動画を作成してもらった上で、ユーザが必要としている情報が要約動画に含まれているか、動画として見やすいか、過剰にフレームが除去されていないか、ユーザが要約動画を巻き戻したり停止した部分があるかなどを調査する。

参考文献

- [1] 三浦宏一, 浜田玲子, 井手一郎, 坂井修一, 田中英彦.: 動きに基づく料理映像の自動要約, 情報処理学会論文誌: コンピュータビジョンとイメージメディア. 44(9), 2003.
- [2] Pei-Yu (Peggy) Chi, J. L. M. D. W. L. B. r. H., Joyce Liu: DemoCut: Generating Concise Instructional Videos for Physical Demonstrations, in Proceedings of the 26th annual ACM symposium on User interface software and technology, UIST ' 13, p. 141 - 150, New York, NY, USA (2013), Association for Computing Machinery
- [3] 沼澤翠, 五十嵐悠紀: 操作説明動画からの要約静止画作成支援システム. 情報処理学会第86回全国大会講演論文集 2024.