

# システム操作動画から要約一枚画を半自動生成するシステムの提案

理学専攻 情報科学コース 2340672 沼澤 翠 (指導教員：五十嵐 悠紀)

## 1 はじめに

情報科学分野の中でも、特にヒューマンコンピュータインタラクションやコンピュータグラフィックスに関する研究分野では、研究者自らが新しくシステムを提案し開発することが非常に多い。このような研究者にとって、外部への研究発表では、自身が開発したシステムについての説明することが欠かせない。研究発表はスライド資料を使った口頭発表だけではなく、論文や研究報告書によるものも含まれる。口頭発表では、システムの挙動を動画を用いて詳しく説明することが可能である。一方、論文や研究報告書は紙や PDF 形式として閲覧者に提供され、動画を用いた詳しい説明をその閲覧者に直接行うことが困難である。このような場面では、静止画と文章を用いて説明を行うしかない。しかしながら、このための静止画作成は手作業で行われることが多く時間が費やされる。その原因の例として、アプリケーションソフトを使って動画内で表されている物を模した図形を作成する事が挙げられる。また、時系列的な変化を示すために複数枚用意することが多い。

動画の内容を一枚の静止画に要約する関連研究として、Correa らの動画から動的な物語生成を行うシステム [1] が挙げられる。この研究では、シーンの移り変わりを一枚の静止画に表す手法が提案されている。また、動画内の連続的な変化を静止画に要約する関連研究として、Goldman らの動画から図式的なストーリーボードを生成するシステム [2] が挙げられる。この研究では、動画から任意のキーフレームを選択し、映されている背景の変化と人物の動きや移動を、効果的な矢印図形と共に静止画に表す手法が提案されている。

本研究の目的は、研究者自身が開発したシステムを説明するための画像を、手作業よりも容易かつ自由に作成できるようにすることである。本研究では、「動画内で位置変化のあったオブジェクトの軌跡を表すこと」が動画要約の役割を果たすという仮説を立てた。そこで、PC 上で扱われるシステムの挙動を示した動画から半自動的に軌跡を表す一枚画を作成する手法を提案し、その手法を実装したプロトタイプの開発を行った。本論文では、提案手法のアルゴリズムとプロトタイプ開発、提案手法の有用性の検証について行った結果を報告する。

## 2 提案手法

本章では、提案する二種類の経路要約一枚画作成手法とその実装について説明する。

### 2.1 提案手法の実装

本研究では、提案手法を実装したシステムのプロトタイプとして、ローカルサーバアプリケーションを開発した。フレームワークには Streamlit を使用した。Python を用いてバックエンドロジックを実装し、CSS を用いてユーザーインターフェースのスタイルを設定した。提案手法ごと

にインターフェース画面は分かれている。

### 2.2 提案手法 A: 半透明化したフレームの多重合成

提案手法 A は半透明化した画像を数枚重ね合わせて一枚にするものである。アルゴリズムを表 1 に示す。

表 1: 提案手法 A のアルゴリズム

手順	処理実行者	処理内容
1	ユーザ	MP4 形式の動画を入力する。
2	システム	入力動画を読み込み 10 枚/秒のフレーム切り出しを行う。
3	システム	全てのフレームをユーザに提示する。
4	ユーザ	重ね合わせに使用するフレームを選択する。
5	ユーザ	重ねる際の各フレームの不透明度を設定する。
6	システム	不透明度が設定されたフレームを全て重ねた一枚画を PNG 形式で保存し、ユーザに提示する。

手順 4 と手順 5 のそれぞれにおいて、ユーザが操作した結果を手順 6 によって即時にインターフェース画面に提示する。このことによって、リアルタイムでのインタラクティブな操作を実現している。

### 2.3 提案手法 B: 動的オブジェクト抽出による配置合成

提案手法 B では、動いたオブジェクトを自動抽出し、ユーザが任意の位置に配置する。アルゴリズムを表 2 に示す。

表 2: 提案手法 B のアルゴリズム

手順	処理実行者	処理内容
1	ユーザ	MP4 形式の動画を入力する。
2	システム	入力動画を読み込み 10 枚/秒のフレーム切り出しを行う。
3	システム	最初のフレーム画像と最後のフレーム画像を用いて、OpenCV による差分検出を行う。
4	システム	得られた差分から、OpenCV による色空間の処理を行い、動きのあったオブジェクトのみを抽出し PNG 画像として保存する。
5	システム	得られたオブジェクト画像をユーザに提示する。
6	ユーザ	要約一枚画に使用するオブジェクトを選択し、不透明度を設定する。
7	ユーザ	不透明度が設定されたオブジェクトを、最初のフレーム画像の前面の任意の位置に配置していく。
8	システム	随時その状態を一枚画として PNG 形式で保存し、ユーザに提示する。

手順 6 と手順 7 のそれぞれにおいて、提案手法 A と同様にリアルタイムでのインタラクティブな操作を実現している。

### 3 実験

本章では、提案手法 A・B の有用性を確かめるために著者自身が行った二つの実験について報告する。

#### 3.1 実験 X: 提案手法による表現可能性の検証

一つ目の実験では、提案手法 A・B を用いて作成された一枚画において、システム操作動画内で動きのあったオブジェクトの軌跡がどれくらい理解し易く表現されるか調査した。また、提案手法による一枚画作成の標準的な手間を検証した。いずれの一枚画においても、オブジェクトの移動が進むにつれて濃く見えるように作成した。

提案システムへの検証用入力動画として、元村らの小型人形用衣服デザインシステム [3] を使用する際の画面収録を用いた。特に、ユーザによる衣服デザインにおいて、特徴となるパーツを衣服上に配置する部分を入力動画とした。入力動画内では、文字オブジェクトである「YALE」が左方向に移動し、その後下方向に移動する。

提案手法 A を用いて作成した、使用フレーム数が異なる四種類の一枚画を図 1 に示す。

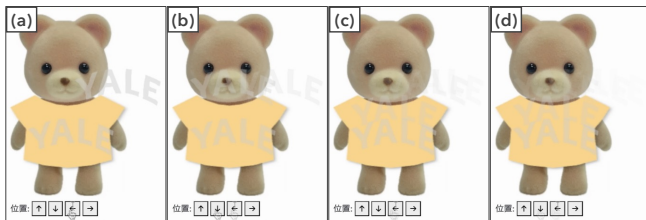


図 1: 実験 X において提案手法 A を用いて作成した一枚画。(a) は 2 枚のフレームを使用。(b) は 3 枚のフレームを使用。(c) は 4 枚のフレームを使用。(d) は 5 枚のフレームを使用。

一枚画作成に要した時間は (a) と (b) が約 1 分、(c) が約 2 分、(d) が約 3 分であった。図 1 に示した結果から、重ね合わせに使用するフレーム枚数と一枚画の理解容易度の関係性を得た。使用するフレーム数が少ないほど各フレームでのオブジェクトの位置は分かりやすいが、オブジェクトの軌跡は理解し難い。一方で、フレーム数が多いほど動画の始まりに近いフレームが非常に薄くなり、オブジェクトの位置が見づらくなるが、軌跡は理解し易くなる。

提案手法 B を用いて作成した、5 段階時系列での軌跡を示す一枚画を図 2 に示す。

図 2 の作成に要した時間は約 20 分であった。また、「YALE」の文字 4 個をそれぞれ 5 回配置したため、配置したオブジェクトは 20 個であった。さらに、図 1(d) からわかるように、提案手法 A を用いた場合、5 段階時系列では軌跡の初めの方を理解し難くなっていた。一方で、提案手法 B を用いた場合、5 段階時系列の場合でも、軌跡の全てを理解し易くできることが分かった。



図 2: 実験 X において提案手法 B を用いて作成した一枚画。

#### 3.2 実験 Y: 一枚画作成時間比較

二つ目の実験では、提案手法 A・提案手法 B・手作業作成の三種類の方法で、同等の一枚画を作成するのに要する時間を比較した。手作業作成は Microsoft PowerPoint (以下、PPT とする) を用いて行った。

実験 Y で使用した入力動画は、地図アプリ上で現在地のピンアイコンが移動する場面を、イラストを用いて模したアニメーションである。背景の地図やピンアイコンのイラストは、画像生成 AI である DALL-E[4] を使用して作

成した。入力動画の概要を図 3(a) に示す。

三種類の方法によって作成した一枚画を図 3(b)~(d) に示す。また、それぞれの作成に要した時間を表 3 に示す。

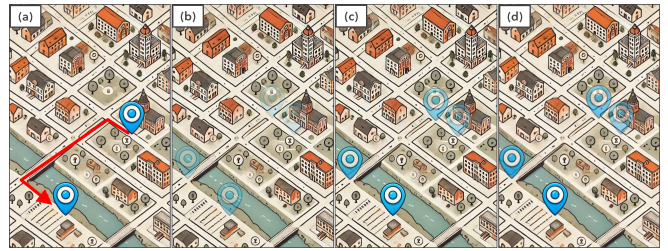


図 3: 実験 Y に関する画像。(a) は入力動画の概要であり、赤色矢印はピンアイコンの軌跡を示す。(b) は提案手法 A を用いて作成。(c) は提案手法 B を用いて作成。(d) は PPT を用いて作成。

表 3: 実験 Y において一枚画作成に要した時間

提案手法 A	提案手法 B	PPT
5 分	8 分	15 分

PPT を用いた図 3(d) の一枚画の作成手順は以下である。入力動画の最後の場面のスクリーンショットを撮影した。背景削除機能を使って、その画像からピンアイコンのみの画像を作成した。最後の場面のスクリーンショットをベース画像として、ベース画像の前面にピンアイコンを三箇所配置し、それぞれの透明度を変更した。実験 Y では、提案手法 A・B が PPT を用いる方法と比較して短い時間で一枚画を作成できることが、表 3 から明らかになった。

### 4 考察

実験 X と実験 Y で得られた結果から次の事が考えられる。提案手法 A は短い時間で一枚画を作成できるが、表す時系列の段階が 5 を超える場合には適していない。提案手法 B は表す時系列の段階が多くなっても、オブジェクトの軌跡の理解し易さを確保できる。しかし、配置するオブジェクト数が多い場合は一枚画作成に時間が掛かる。PPT を用いる方法は、配置をしたいオブジェクトの背景透過画像作成に時間が掛かる。

### 5 まとめと今後の展望

本研究では、システム操作動画において、動いたオブジェクトの軌跡の要約一枚画を半自動生成する手法を二通り検討した。今後は、提案手法 A・提案手法 B・PPT を用いる方法のそれぞれの長所を活かした手法の検討を行う。また、提案システムのインタラクティブ性やユーザエクスペリエンスをさらに高めることを目標としている。そのために、本研究で開発したプロトタイプをユーザに使用してもらい評価実験を行う予定である。

### 参考文献

- [1] Carlos D. Correa, Kwan-Liu Ma, 2010, Dynamic video narratives, ACM Trans, Graph. 29, 4, Article 88 (July 2010), 9 pages, <https://doi.org/10.1145/1778765.1778825>.
- [2] Dan B Goldman, Brian Curless, David Salesin, Steven M. Seitz, 2006, Schematic storyboarding for video visualization and editing, ACM Trans, Graph. 25, 3 (July 2006), 862–871, <https://doi.org/10.1145/1141911.1141967>.
- [3] 元村 愛美, 五十嵐 悠紀, 人間用衣服単一画像からの小型人形用衣服デザイン生成手法, 研究報告ヒューマンコンピュータインタラクション (HCI), 2024-HCI-210(28), 1-7 (2024-11-11), 2188-8760.
- [4] OpenAI, n.d., DALL-E: Creating images from text, GitHub Repository, <https://github.com/openai/DALL-E>.