

ShowMe: 対話的な強調表示と拡大表示によるプレゼンテーションビデオの視覚的アクセシビリティの改善

SECHAYK Yotam* SHAMIR Ariel† 五十嵐 健夫*

概要. プレゼンテーションビデオを使った学習は広く一般的に行われている。講師は、ビデオ作成過程で、さまざまな視覚的補助動作を活用することが多い。具体的には、プレゼンテーション中のポインティング、マーキング、スケッチなどが、視覚的補助動作としてよく使われる。しかし、これらの動作は視覚的に認識が難しいことが多く、説明が不十分であることが多い。弱視の学習者は、このような動作に追従するために、常にプレゼンテーションのフレーム内を探索する必要があり、フラストレーションと疲労につながっている。我々は、この問題を理解し解決するために、3人の弱視ユーザとユーザ参加型デザインを実施し、その結果にもとづき、講師の視覚的補助動作を強調表示し、拡大表示するツール ShowMe を開発した。ShowMe は、弱視ユーザがプレゼンテーションをフォローできるように支援し、疲労とフラストレーションを軽減する。

1 はじめに

学習動画は、正式および非公式の学習環境において、現代の教育において極めて重要である [34] [1] [18] [31]。教師はしばしばスライド [43]、インタラクティブ黒板 [32]、画面共有 [10]、2D/3D アニメーション [19] などの様々な視覚的ツールを使用し、多くのプレゼンテーションはこれらを組み合わせている。電子ポインタあるいはペンも、ポイント・マーク・スケッチにより学習者を導くのによく使われる [35] (図 1 を参照)。これらのツールは多くの学習に役立つが、低視力 (LV) の学習者にとってリアルタイムで気付くことが困難である [45]。

学習のためのユニバーサルデザインの原則とアクセシブルプレゼンテーションのガイドライン [4] [29] [13] は、記述的なスピーチによるアクセシビリティを推奨している。しかし、実際には、多くの教師はポイント・マーク・スケッチの際に不完全な説明を提供することが一般的である。例えば、教師はポインティングしながら抽象的な指示詞 (例: 「これ」, 「ここ」) を使用したり、テキストを強調するためにマークしたり、図をスケッチしながら詳細を完全に説明しないことがある [13]。

これらの行動—ポイント・マーク・スケッチ—をビジュアルアクティビティと呼ぶ。これらは教師が学習者を導き、情報を伝えるのに役立つ。これらの活動はしばしば微細で、小さく、コントラストが低く、気付きにくい。その結果、低視力の学習者は常にこれらの手掛かりを探さなければならず、視覚的探索の問題を示している [53]。



図 1. ビジュアルアクティビティの例。(A) ポイント、(B) マーク、(C) スケッチ。

3人の低視力学習者と共に、視覚的探索の問題を理解し解決するためのユーザ参加型デザインを行った。その結果、ビジュアルアクティビティの非アクセシビリティが関与・注意・挫折のレベルにどのように影響を与え、認知的負荷を増大させていることが示された。ユーザ参加型デザインの結果に基づき、視覚的探索の問題に対処するための ShowMe を開発した。ShowMe は2つの部分で構成されている: (1) 活動認識, (2) 活動可視化である。ShowMe はビジュアルアクティビティをより見やすく、追いややすくするのに助ける。

2 関連研究

2.1 動画におけるアクセシビリティ

低視力 (LV) のビデオアクセシビリティは重要な問題である [36]。解説放送 (AD) は伝統的に言語的な説明を追加する [38] [46] [41] [33] [39]。最近

Copyright is held by the author(s).

* 東京大学

† Reichman University

では、機械学習 [48] [5] [20] [52] [24] [23], コラボレーション [21], および空間オーディオ [16] による改善がなされている。しかし、AD はしばしば LV ユーザにとって過度に説明的に感じられる。

視覚的強化手法には、ポストプロセッシング (例: コントラスト, シャープネス) [44] や顕著性に基づく拡大 [2] が含まれるが, これらはビジュアルアクティビティを見逃したり, 費用がかかり, プレゼンテーションに最適化されていない [25]。

先行研究において, 我々は主に時間制御 (速度調整およびリプレイ機能) に重点を置き, 一般ユーザー向けのアクセシビリティの向上を探求した [45]。本研究では, 特に弱視者に対するアクセシビリティに注目し, 視覚的強化手法の探究を行い, デザインおよび実装の双方において改善を目指した。デザインプロセスにおいては, 参加型デザインアプローチを採用し, パーソナライゼーションの重要性を強調した。実装面においては, ビジュアルアクティビティを正確に検出および分類するためのフレーム間関係を定式化する方法を提示する。

2.2 学習動画におけるビジュアル・コンテンツのアクセシビリティ

今日では, YouTube [51], Khan Academy [27], Coursera [8] などのプラットフォームが広範なプレゼンテーション動画を提供している。制作過程において, 先行研究はアクセシブルなプレゼンテーションを簡単に作成する方法を提案している [42] [14] [28]。既存のプレゼンテーションに対して, 先行研究はスライド要素の意味的なハイライト [26] [50] と, 未対応のスライド要素の検出を探求している [43]。しかし, ビジュアルアクティビティはスライド要素とみなされず, 未対応のままである。

2.3 映像内オブジェクトとのインタラクション

ビデオ内オブジェクトはインタラクティブのために利用される。先行研究は, インタラクティブなナビゲーション [12] [32], ソフトウェアシミュレーション [49] [3], ノート作成 [6], 視線に基づくハイライト [37], テキスト操作 [10], またはマウスクリックを使用してスライド要素を拡大する手法を提案している [10]。本研究では, ビジュアルアクティビティをインタラクションのための基点として利用する。

3 ユーザ参加型デザイン

3人の低視力ユーザーと共にユーザー参加型デザインを行い, 以下の点について調査した: (RQ1) 低視力ユーザーがビジュアルアクティビティをどのように体験するか, (RQ2) ビジュアルアクティビティから生じるチャレンジは何か, (RQ3) アクティビティをより見やすく追いややすくするにはどのアプローチが効果的か, および (RQ4) 本ユーザー参加型デザインにおける経験はどのようなものであったか。

3.1 方法

3.1.1 参加者

家族や友人を通じて3人の低視力参加者を募集した。参加者はそれぞれ異なる視覚状態で, 異なるアクセシビリティのニーズを持っていた (表1を参照)。

3.1.2 手順

ユーザー参加型デザインはZoomを通じてオンラインで行われ, 各参加者に対して5回の60分セッションを行った。これらのセッションは5週間にわたって実施した。参加者は, 本研究全体に対して100ドルのギフトカードで時間に対する補償を受けた。セッション1では教育用プレゼンテーション動画を見る際の体験と課題について議論し (RQ1-RQ2), セッション2-4ではShowMeプロトタイプを反復設計し (RQ3), セッション5では協働設計の体験についての考察を行った (RQ4)。各セッションは分析のために記録され, 文字起こしされた。

3.2 調査結果

本セクションでは, 3名からそれぞれヒアリングした結果を記載する。その際, 我々に対して説明した参加者をカッコ書きで文末に記載することとする。

3.2.1 低視力の経験 (RQ1)

参加者全員は, 視覚情報が失われることは避けられないことが多いという考えを共有した。Johnはさらに, 見逃したコンテンツを視覚的に特定できないと, 何が見逃されているのかを把握するのが難しいと説明した。すべての参加者は, プレゼンテーション動画を視聴した経験があった。ビジュアルアクティビティに関して, 参加者全員が「何を見ればいいのか」あるいは「どこを見ればいいのか」についての不確実性を表明した。その結果, オンラインプレゼ

表 1. ユーザ参加型デザイン参加者情報。

仮名	年齢	性別	視覚状態	発症	教育	スクリーン拡大鏡
John	36	男性	アルビニズム, 乱視	先天性	高校卒	時々
Eva	30	女性	網膜色素変性症	先天性	大学卒	使用せず
Emily	33	女性	デュアン症候群	先天性	大学卒	使用せず

ンテーション動画の視聴にはより高いレベルの注意が必要であると、参加者全員が述べた。また、JohnとEvaは時には挫折したり、視聴を避けたりすることもあると述べた。これは、低視力ユーザが微妙な視覚変化を検出するために多大な努力を投入しなければならないことを示している。

3.2.2 低視力のチャレンジ (RQ2)

異なるプレゼンテーション動画は異なるスタイルの選択を示し、多くの教師が小さなポインタや低コントラスト、細い電子ペンを使用する。これらの視覚スタイルによって反応時間が延び、参加者全員がビジュアルアクティビティの認識能力に影響が出ると感じた。Johnは特に「マーキング動作」が非常に微妙かつ短時間であるため難しいと感じている。Emilyにとっては、教師が線を一緒に描くことで理解しづらくすることがよくある。

最後に、拡大ソフトウェアを使用することは面倒で、絶え間ない手動調整を要する(John)。したがって、いくつかのユーザは詳細を見るのをあきらめ(John)、またはデジタルで拡大できるように、スクリーンショットを頻繁に一時停止して保存する(Eva)。

3.2.3 見やすくするアプローチ (RQ3)

ビジュアルアクティビティに対処するためのアプローチを洗練するために、我々は以前の研究に基づいたプロトタイプでの設計を行った [45]。

初期の印象: 初期のプロトタイプは赤い半透明円を使用し、フレーム間の最新の視覚的变化をハイライトした。ズーム機能は、連続して5秒以上の変化が発生し、空間的な閾値内にある場合にトリガーとして追加された。すべての参加者はこれらのハイライトを非常に助かると感じた。Johnは「それがないと迷子になる」と述べた。しかし、参加者全員が視覚スタイルを制御できないことを批判した。また、ズームは有用であったが、教師のスケッチ線は依然として細すぎて見えなかった(John, Emily)。参加者はすべてのアクティビティでズームのオプションも望んでいた。

反復設計の結果: 最終設計では、ハイライトスタイル・ズーム機能・視覚フィルタ(例:色反転)のパーソナライズを可能にした。ハイライト形状は、設定された半径の円またはアクティビティを囲むボックスの2つを提供した。丸を好む人もいるかもしれないが、すべての参加者は「正確にどこを見て何を見ればいいのか教えてくれる」としてボックスを好むと述べた(John)。アニメーションは視覚的な補助手段を強化する可能性があるため(Eva)、ハイライトをアニメーション化するオプションを追加した(サイズの変動)。しかし、JohnとEmilyはそれを

煩わしいと感じた。また、JohnとEvaはタイミングを合わせるためにインストラクターの活動の事前通知が必要であると表明した。

これに対処するために、透明性を高めるためにハイライトの開始時間を1.5秒だけ増やした。Emilyは、一貫性のある視覚化が見つけやすいと述べたため、カーソル(☞)や手(☞)を追加した。指し示すなどのさまざまな活動のために異なるビジュアルを試みたが、参加者は一貫性のあるスタイルを好んだ。

参加者の意見に基づき、すべてのアクティビティでズームを有効にし(Eva)、ズームがアクティブ時に自動で再生が一時停止するトグルを追加し、ズーム速度と拡大率を調整できるようにした(John, Eva)。これらの機能は有益と評価された。結果として、すべての参加者がShowMeを集中力向上と挫折軽減に役立つものと感じ、Johnは「それなしでは迷子になる」と述べ、Evaは「特別なメガネのようだ」と評価した。

3.2.4 本ユーザ参加型デザインにおける経験 (RQ4)

最終セッションでは、参加者に本ユーザ参加型デザインでの経験について尋ねた。すべての参加者はShowMeの設計に意味のある影響を与えたと感じた。Evaは、「それについて話した後にアニメーションが(実装されているのを)見てとても嬉しかった」と述べた。パーソナライズに関しては、すべての参加者が実装されたハイライトとズーム設定に満足感を表明した。JohnとEvaは将来的にShowMeを使用したいと述べたが、Emilyは新しいツールに適應することに関して懸念を示し、「アクセシビリティ機能のために以前iPadの使用を試みたがいつも使っているものとは非常に異なっており、あきらめてしまった」と述べた。

しかし、Emilyは研究への参加がアクセシビリティツールを試そうという姿勢にわずかな変化をもたらしたと述べた。研究中にいくつかの目の疲れや疲労を経験したが、すべての参加者は参加したことを嬉しく思っていた。今後目の疲労を軽減するために、より長い休憩や短くてもっとコースを空けたセッションが役立つと提案された。

4 ShowMe

この章では、我々のアプローチを詳細に説明し、ユーザインタラクションを示す。ShowMeは、アクティビティ認識を使用し、アクティビティビジュアルライゼーションを生成する(図2)。

4.1 アクティビティ認識

ビジュアルアクティビティは、ポイント・マーク・スケッチである(図1)。我々のアプローチは以前の研究に基づき[45]、既存のコンピュータビジョン(CV)動き検出アルゴリズムを使用している[40]。

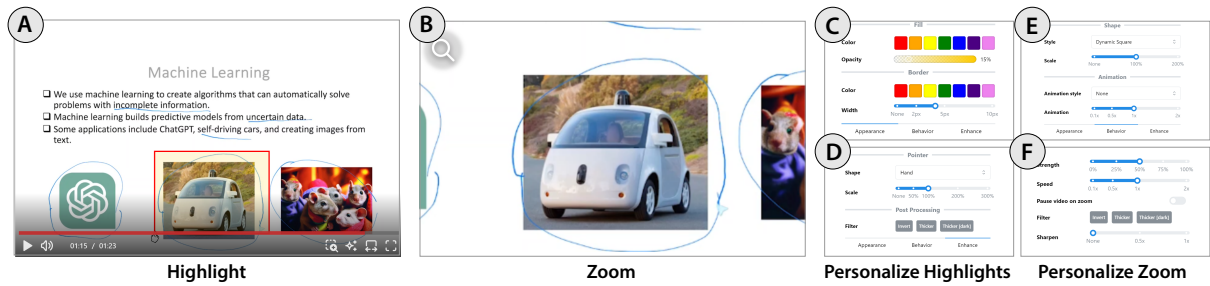


図 2. ShowMe のビデオプレーヤ. (A) 教師のアクションを教師が話すときにハイライト (赤の輪郭, 黄色の塗りつぶし), (B) 教師の活動にズームするオプションを提供. ShowMe はユーザに (C-E) 視覚的ハイライトと (F) ズーム動作をカスタマイズさせる.

従前の研究を時間領域に拡張し, アクティビティを時間的に認識しクラスタリングする. 前回の研究がフレーム間の局所的な違いを検査していたのに対し, ShowMe は違い間の関係を構造化して時間的にアクティビティを作成する.

4.2 アクティビティ視覚化

認識されたアクティビティは, ハイライトを通して可視化され, ズーム機能を備えている. ハイライトとズームはビデオ再生に追加され, 完全なユーザパーソナライゼーションが可能である. 図 2 はユーザインタフェースを示している.

4.2.1 アクティビティハイライト

3章で述べたように, 低視力ユーザは幅広いニーズを持っており, パーソナライゼーションが重要な要素である. ユーザは追加ボタン を使用して設定パネルを開くことができる (図 2 の C-E を参照). ハイライトの設定には, 背景色と不透明度, 境界色と厚み, ポストプロセッシングフィルタ, アニメーションなどが含まれる (表 2 を参照).

表 2. ハイライトのデフォルト設定.

カテゴリー	設定	デフォルト
ビジュアル	塗り色	黄色
	塗り不透明度	15%
	境界色	赤
	境界幅	4px
動作	形状スタイル	ボックス
	スケール	100%
	アニメーション種類	なし
	アニメーション速度	1.0x
改善	ポインタースタイル	なし
	ポインタースケール	100%
	フィルタ	なし

4.2.2 アクティビティ拡大

拡大は Z キーボードキーを押すことによって行われる. 押すと, ビデオプレーヤは現在ハイライトされているアクティビティにスムーズにズームする. アクティビティの位置が更新されると, ズームエリアもスムーズな遷移で更新される. ユーザは追加ボタン を使用して設定パネルを開くことができる (図 2 の F を参照). ビデオプレーヤは, 左上に拡大の視覚的インジケーションを含んでいる (図 2 の B 参照). ユーザはズーム遷移速度を個別化し, 拡大時に一時停止するトグルなどを利用できる (表 3 を参照).

表 3. ズームのデフォルト設定.

設定	デフォルト
ズーム強度	50%
ズーム速度	1.0x
ズーム時の一時停止	いいえ
シャープネス	1.0x
フィルタ	なし

4.3 インタラクションの概要

低視力ユーザが学習動画を視聴すると, アクティビティのハイライトが表示される. ユーザは再生中でも設定をリアルタイムで変更可能で, Z キーを押すと拡大表示に切り替わる. 拡大後は矢印キーで拡大レベルを調整できる. タイムラインナビゲーションや時間スキップ, キーボードショートカット, フルスクリーンなどの通常機能もサポートされ, ユーザは引き続きスムーズな操作が可能である. また, Z キーで拡大を解除したり, 他の設定を調整することもできる.

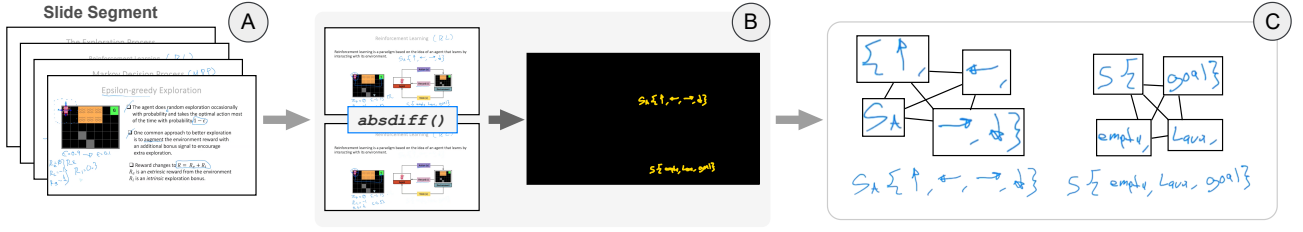


図 3. ShowMe のアクティビティ認識パイプライン. (A) スライドセグメントから始まり、各フレームウィンドウに対して (B) 視覚的な差異を認識し、最後に (C) 接続されたコンポーネント解析に基づいてアクティビティを分類するグラフとして差異を表現している.

5 実装

5.1 アクティビティ認識

認識パイプラインは3つの主要なコンポーネントから成り立っている：(1) スライドセグメント抽出、(2) アクティビティ検出、(3) アクティビティ分類である (図 3 参照). このパイプラインは既存の研究 [45] を拡張したものであり、アクティビティ認識の基盤として安価な動き検出手法 [40] を使用している. Python [17] で開発し、OpenCV [30] と NetworkX [11] ライブラリを使用している.

5.1.1 スライドセグメント抽出

最初のステップはスライドセグメントの抽出である. 以前の研究 [43] [45] と同様に、PySceneDetect [9] を使用してスライド遷移を検出している. これは Python 用の既製のショット検出ライブラリで、Content Detector を使用している. スライドセグメントはそのセグメント内でアクティビティの検出を境界づけるのに使用される.

5.1.2 アクティビティ検出

アクティビティ検出は、2つの近接したフレーム間の差異を検出することから始まる [40]. アルゴリズムはフレームのスライディングウィンドウで各スライドセグメントに適用される. デフォルトのウィンドウサイズ W_{size} は $0.33 \cdot V_{fps}$ に設定され、ここで V_{fps} は動画のフレーム毎秒である. それぞれの結果として得られる輪郭 [47] は、圧縮アーティファクトや見逃せない大きな変化 (例: フレームの60%を超える) を無視するためにサイズでフィルタリングされる. その結果として得られるのは変化領域 (RoC) である. 次に、RoC からグラフ表現を構築する. すべての RoC について、表しているノードは

$$n_i = (T, X, Y, W, H) \quad (1)$$

である. ここで T は開始タイムスタンプ、 X と Y は左上の座標、 W と H は RoC の幅と高さである. ノード n_i と n_j の間にエッジが作成されるのは、2つの RoC が次の条件を満たす場合である:

- 時間的に近い: $\sqrt{T_i^2 - T_j^2} \leq \Theta_T$ ここで Θ_T は設定された閾値である.

- 空間的に近い: $\sqrt{\Delta X^2 + \Delta Y^2} \leq \Theta_S$ ここで ΔX と ΔY はそれぞれ最小の水平または垂直距離であり、 Θ_S は設定された閾値である.

エッジの重みは、2つの RoC 間の視覚的類似度として計算される. 視覚的類似度は hu モーメント [22] を使用して基礎となる輪郭形状をマッチングすることによって計算される. [10] に触発され、2つの RoC をマージする条件は以下である:

- 時間的に隣接: $\sqrt{T_i^2 - T_j^2} \leq \frac{W_{size}}{V_{fps}}$.

- 空間的に同一: $\sqrt{\Delta X^2 + \Delta Y^2} \leq \epsilon$ ここで ϵ は小さな値である.

- 視覚的に類似: $Sim(n_i, n_j) \geq \Gamma$ ここで Γ は設定された閾値である.

これによりオブジェクトの動き (例: ポインタ) が見逃されるのを防ぐ.

5.1.3 アクティビティ分類

RoC グラフの連結成分 (CC) 解析を使用する. 各 CC について、アクティビティ

$$a_i = (CC_i, X, Y, W, H, T_{start}, T_{end}) \quad (2)$$

を定義する. ここで CC_i は連結成分を表しており、 X, Y, W, H はバウンディングボックス (つまり、関連するすべての RoC を含む最小の長方形) を表している. タイムスタンプ T_{start} と T_{end} はアクティビティの開始と終了を表す. 次に、 W と H がそれぞれビデオの幅と高さの1%から70%の間にある場合にアクティビティを有効としてマークする. 最後に、各アクティビティのタイプをヒューリスティックなルールベースのアプローチを使用して分類する. 例えば、平均エッジ類似度がある閾値を超える CC は pointing として分類できる. ユーザ参加型デザイン参加者は一貫したスタイルを好んでいたが、将来の研究では、より頑健な分類アプローチを用いてスタイルの多様性をさらに探る必要がある.

5.2 アクティビティ視覚化

ユーザ参加型デザインの結果、低視力ユーザの多様な視覚ニーズに対応するためのパーソナライズの重要性を強調している。そこで、ビデオ再生の拡張オーバーレイとして視覚化を実装した。この実装は TypeScript [7] を用いて React [15] 上で行い、ユーザ設定やデバイス（例：画面解像度、アスペクト比）に適応するリアクティブな動作を実現した。

5.2.1 アクティビティハイライト

再生中、各 `ontimeupdate` ビデオプレーヤーイベントは現在のアクティビティを収集する。ShowMe は、アクティビティの開始時刻に基づくヒューリスティックを用いて最も関連性のあるアクティビティを選択する。この選択されたアクティビティはユーザの設定に基づいて視覚化される。以前の研究 [45] が静的スタイルを使用していたのとは異なり、ShowMe のキーとなる機能はリアルタイムパーソナライゼーションとリアクティブデザインである。

5.2.2 アクティビティ拡大

ズームインビューは、動画フレームを反映する `canvas` エレメントを使用して実装されている。拡大は (C_x, C_y, S_x, S_y, Z) で定義され、ここで C_x, C_y はアクティビティの中心、 S_x, S_y は左右のシフト量、 Z はズームファクターである。アクティビティの中心にズームファクターでズームしながら、ズームしたビューをビデオパネルに完全に収めるようにシフトする。

6 議論

バリアフリー映像。 低視力ユーザは残存視力に頼ってタスクを行うことが多いため、視覚コンテンツへのアクセスの必要性は重要である。このアクセスを促進するためには、ユーザはコンテンツを個々のニーズに合わせて適応させる必要がある。視覚アクセスが利用できない場合、多くの弱視ユーザーは疲れ、フラストレーションを経験し、中には作業を放棄する人もいる。弱視者の視覚状態は千差万別であるため、万能なものは存在しない。例えば、アニメーションは視野の狭い人には役立つが、視力の弱い人には非常に邪魔になる。したがって、パーソナライゼーションと適応性はアクセシビリティの重要な要素である。我々のユーザ参加型デザインはプレゼンテーション動画に焦点を当てたが、他の領域でも同様の課題が経験されている [53]。

ユーザ参加型デザイン。 アクセシビリティにおいて可視性認識は重要な考慮点である。一般的な信念に反して、最も重要な可視性認識は、すべてを認識することはできないという認識である。我々のアプローチはこの原則を反映するためにユーザ参加型デザイ

ンの結果を反映している。参加者の数は限られているが、彼らの独自の経験は、効果的なアクセシビリティツールを作成する上でユーザ主導の設計の必要性を強調している。ShowMe は馴染みのある視覚的慣例に基づいて作られ、さらに適応性とパーソナライゼーションが追加されている。しかし、将来の研究では ShowMe をより大規模な低視力ユーザグループでテストし、利用パターンや嗜好をより明確にする必要がある。

技術的境界とユーザ満足度。 我々のアクティビティ認識アプローチにはいくつか重要な境界がある。例えば、ShowMe はアクティビティの意味的な重要性を考慮しておらず、複数のアクティビティが発生した場合に重要でないものをハイライトする可能性がある。ハイライトは参照している要素ではなく、変更そのものに焦点を当てる；例えば、テキストをマークする際、マーク自体をハイライトし、関連するテキストではない。将来の研究では、アクティビティ検出と視覚ハイライトに対する意味的アプローチを検討すべきである。それにもかかわらず、ユーザ参加型デザインの結果は、不完全な解決策でも低視力ユーザの体験を大幅に向上させることができることを示している。

新ドメインへの拡張。 視覚的探索問題は低視力ユーザの生活に一般的にみられる問題であるが [53] [45]、我々のアクティビティベースのアプローチはプレゼンテーション動画に特化されている。しかし、我々のアプローチは機械学習に頼らないため、ソフトウェアチュートリアルのような他のビデオ領域にも、さらには VR/AR のような他のメディアにもシームレスに適応することができる。ただし、ライブ放送には追加の考慮が必要である。ShowMe は完全に利用可能な時間情報に依存しているため、ライブシナリオではアクティビティの境界（空間と時間の両方）をよりよく特定するために人工バッファを必要とする。将来の研究では、ライブコンテキストにおける低視力ユーザの好みを調査すべきである。

7 結論

プレゼンテーション動画で、教師は学習者を視覚的に導くために電子ポインタあるいはペンを使用することが多いが、低視力 (LV) ユーザはこれらの微細な手がかりを頻繁に見逃す。我々のユーザ参加型デザインは、LV 学習者が小さな視覚変化を常に探す「視覚的探索問題」を特定するのに役立つ。それに応じて、これらの活動を検出し、ハイライトと拡大で再生を向上させるアクセシブルツールである ShowMe を開発した。パーソナライゼーションを念頭に置いて設計された ShowMe は、LV 参加者の挫折を軽減し、学習の容易さを改善する。

謝辞

本研究は Google, および JST AdCROP (JP-MJKB23) の一部支援を受けて行われた。その経済的なサポートに感謝の意を表す。

参考文献

- [1] W. Aljedaani, M. Alkahtani, S. Ludi, M. W. Mkaouer, M. M. Eler, M. Kessentini, and A. Ouni. The State of Accessibility in Blackboard: Survey and User Reviews Case Study. In *Proceedings of the 20th International Web for All Conference, W4A '23*, p. 84–95, New York, NY, USA, 2023. Association for Computing Machinery.
- [2] A. S. Aydin, S. Feiz, V. Ashok, and I. Ramakrishnan. Towards making videos accessible for low vision screen magnifier users. In *Proceedings of the 25th International Conference on Intelligent User Interfaces, IUI '20*, p. 10–21, New York, NY, USA, 2020. Association for Computing Machinery.
- [3] N. Banovic, T. Grossman, J. Matejka, and G. Fitzmaurice. Waken: reverse engineering usage information and interface structure from software videos. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pp. 83–92, 2012.
- [4] S. E. Burgstahler and R. C. Cory. *Universal design in higher education: From principles to practice*. Harvard Education Press, 2010.
- [5] V. P. Campos, L. M. Gonçalves, W. L. Ribeiro, T. M. Araújo, T. G. Do Rego, P. H. Figueiredo, S. F. Vieira, T. F. Costa, C. C. Moraes, A. C. Cruz, et al. Machine generation of audio description for blind and visually impaired people. *ACM Transactions on Accessible Computing*, 16(2):1–28, 2023.
- [6] Y. Cao, H. Subramonyam, and E. Adar. VideoSticker: A tool for active viewing and visual note-taking from videos. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*, pp. 672–690, 2022.
- [7] M. Corporation. TypeScript: JavaScript With Syntax for Types. <https://www.typescriptlang.org/>, 2012. Accessed: 2024-10-21.
- [8] Coursera. Coursera. <https://www.coursera.org/>. Accessed: 2023-10-05.
- [9] R. D. and A. D. PySceneDetect, 2024.
- [10] L. Denoue, S. Carter, M. Cooper, and J. Adcock. Real-time direct manipulation of screen-based videos. In *Proceedings of the companion publication of the 2013 international conference on Intelligent user interfaces companion*, pp. 43–44, 2013.
- [11] N. Developers. NetworkX, 2024.
- [12] P. Dragicevic, G. Ramos, J. Bibliowicz, D. Nowrouzezahrai, R. Balakrishnan, and K. Singh. Video browsing by direct manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 237–246, 2008.
- [13] Alfred T D ' Agostino. Accessible teaching and learning in the undergraduate chemistry course and laboratory for blind and low-vision students. *Journal of Chemical Education*, 99(1):140–147, 2021.
- [14] M. Elias, A. James, E. Ruckhaus, M. C. Suárez-Figueroa, K. A. De Graaf, A. Khalili, B. Wulff, S. Lohmann, and S. Auer. SlideWiki-Towards a Collaborative and Accessible Platform for Slide Presentations. In *EC-TEL (Practitioner Proceedings)*, pp. 1–3, 2018.
- [15] I. Facebook. React: A JavaScript Library for Building User Interfaces. <https://reactjs.org/>, 2013. Accessed: 2024-10-21.
- [16] D. Fan, S. Junuzovic, J. C. Tang, and T. Jaeger. Improving the Accessibility of Screen-Shared Presentations by Enabling Concurrent Exploration. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2023, New York, NY, USA, October 22-25, 2023*, pp. 44:1–44:16. ACM, 2023.
- [17] P. S. Foundation. Python Programming Language, 2024.
- [18] S. Ghosh and A. Figueroa. Establishing TikTok as a Platform for Informal Learning: Evidence from Mixed-Methods Analysis of Creators and Viewers. In T. X. Bui ed., *56th Hawaii International Conference on System Sciences, HICSS 2023, Maui, Hawaii, USA, January 3-6, 2023*, pp. 2431–2440. ScholarSpace, 2023.
- [19] S. Grice and J. Hughes. Can music and animation improve the flow and attainment in online learning? *Journal of Educational Multimedia and Hypermedia*, 18(4):385–403, 2009.
- [20] T. Han, M. Bain, A. Nagrani, G. Varol, W. Xie, and A. Zisserman. AutoAD: Movie description in context. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18930–18940, 2023.
- [21] M. Hirvonen, M. Hakola, and M. Klade. Co-translation, consultancy and joint authorship: User-centred translation and editing in collaborative audio description. *Journal of Specialised Translation*, (39):26–51, 2023.
- [22] M.-K. Hu. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2):179–187, 1962.
- [23] G. Jain, B. Hindi, C. Courtien, C. Wyrick, X. Y. T. Xu, M. C. Malcolm, and B. A. Smith. Towards accessible sports broadcasts for blind and low-vision viewers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2023.
- [24] L. Jiang, M. Phutane, and S. Azenkot. Beyond audio description: Exploring 360 video accessibility with blind and low vision users through

- collaborative creation. In *Proceedings of the 25th international ACM SIGACCESS conference on computers and accessibility*, pp. 1–17, 2023.
- [25] L. Jiang, M. Xu, T. Liu, M. Qiao, and Z. Wang. Deepvs: A deep learning based video saliency prediction approach. In *Proceedings of the european conference on computer vision (eccv)*, pp. 602–617, 2018.
- [26] H. Jung, H. V. Shin, and J. Kim. Dynamic-slide: Exploring the design space of reference-based interaction techniques for slide-based lecture videos. In *Proceedings of the 2018 Workshop on Multimedia for Accessible Human Computer Interface*, pp. 33–41, New York, NY, USA, 2018. ACM.
- [27] Khan Academy. Khan Academy. <https://www.khanacademy.org/>. Accessed: 2024-08-10.
- [28] J. Kong, D. Sabha, J. P. Bigham, A. Pavel, and A. Guo. TutorialLens: authoring Interactive augmented reality tutorials through narration and demonstration. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, pp. 1–11, 2021.
- [29] R. E. Ladner and K. Rector. Making your presentation accessible. *Interactions*, 24(4):56–59, 2017.
- [30] O. S. C. V. Library. OpenCV, 2024.
- [31] I. A. Mastan, D. I. Sensuse, R. R. Suryono, and K. Kautsarina. Evaluation of distance learning system (e-learning): a systematic literature review. *Jurnal Teknoinfo*, 16(1):132–137, 2022.
- [32] T.-J. K. P. Monserrat, S. Zhao, K. McGee, and A. V. Pandey. Notevideo: Facilitating navigation of blackboard-style lecture videos. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1139–1148, 2013.
- [33] R. Natalie, J. Loh, H. S. Tan, J. Tseng, I. L. Y.-R. Chan, E. H. Jarjue, H. Kacorri, and K. Hara. The efficacy of collaborative authoring of video scene descriptions. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–15, 2021.
- [34] E. Navarrete, A. Nehring, S. Schanze, R. Ewerth, and A. Hoppe. A Closer Look into Recent Video-based Learning Research: A Comprehensive Review of Video Characteristics, Tools, Technologies, and Learning Effectiveness. *CoRR*, abs/2301.13617, 2023.
- [35] E. Navarrete, A. Nehring, S. Schanze, R. Ewerth, and A. Hoppe. A Closer Look into Recent Video-based Learning Research: A Comprehensive Review of Video Characteristics, Tools, Technologies, and Learning Effectiveness. *arXiv preprint arXiv:2301.13617*, 2023.
- [36] A. Nevsky, T. Neate, E. Simperl, and R. Vatavu. Accessibility Research in Digital Audiovisual Media: What Has Been Achieved and What Should Be Done Next? In P. L. Callet, M. P. D. Silva, T. Vigier, K. Tahiroglu, N. Murray, G. Valenzise, and M. Wang eds., *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences, IMX 2023, Nantes, France, June 12-15, 2023*, pp. 94–114, Nantes, France, 2023. ACM.
- [37] C. Nguyen and F. Liu. Gaze-based notetaking for learning from lecture videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2093–2097, 2016.
- [38] A. C. of the Blind. Audio Description Project, Guidelines for Audio Describers. <https://www.acb.org/adp/guidelines.html>, n.d. Accessed: 2024-08-10.
- [39] J. Packer, K. Vizenor, and J. A. Miele. An overview of video description: history, benefits, and guidelines. *Journal of Visual Impairment & Blindness*, 109(2):83–93, 2015.
- [40] S. Parveen and J. Shah. A motion detection system in python and opencv. In *2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV)*, pp. 1378–1382, Virtual Conference, 2021. IEEE, IEEE.
- [41] A. Pavel, G. Reyes, and J. P. Bigham. Rescribe: Authoring and automatically editing audio descriptions. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 747–759, 2020.
- [42] Y.-H. Peng, J. Jang, J. P. Bigham, and A. Pavel. Say it all: Feedback for improving non-visual presentation accessibility. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.
- [43] Y. Peng, J. P. Bigham, and A. Pavel. Slidecho: Flexible Non-Visual Exploration of Presentation Videos. In J. Lazar, J. H. Feng, and F. Hwang eds., *ASSETS '21: The 23rd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, USA, October 18-22, 2021*, pp. 24:1–24:12. ACM, 2021.
- [44] A. Sackl, F. Graf, R. Schatz, and M. Tschelegi. Ensuring accessibility: Individual video playback enhancements for low vision users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–4, 2020.
- [45] Y. Sechayk, A. Shamir, and T. Igarashi. Smart Replay: eラーニング動画における視覚的・時間的アクセシビリティの向上. In *Proceedings of the Workshop on Interactive Systems and Software (WISS)*, Japan, 2023.
- [46] J. Snyder. Audio description: The visual made verbal. In *International congress series*, Vol. 1282, pp. 935–939. Elsevier, 2005.
- [47] S. Suzuki, et al. Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, 30(1):32–46, 1985.

- [48] Y. Wang, W. Liang, H. Huang, Y. Zhang, D. Li, and L.-F. Yu. Toward automatic audio description generation for accessible videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.
- [49] S. Yang, J. Vermeulen, G. Fitzmaurice, and J. Matejka. AQUA: Automated Question-Answering in Software Tutorial Videos with Visual Anchors. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2024.
- [50] C. Yip, J. M. Chong, S. Y. Kwek, Y. Wang, and K. Hara. Visionary Caption: Improving the Accessibility of Presentation Slides through Highlighting Visualization. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–4, 2021.
- [51] YouTube. YouTube. <https://www.youtube.com/>. Accessed: 2024-08-10.
- [52] B. F. Yuksel, P. Fazli, U. Mathur, V. Bisht, S. J. Kim, J. J. Lee, S. J. Jin, Y.-T. Siu, J. A. Miele, and I. Yoon. Human-in-the-loop machine learning to increase video accessibility for visually impaired and blind users. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pp. 47–60, 2020.
- [53] Y. Zhao, S. Szpiro, J. Knighten, and S. Azenkot. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16*, p. 73–84, New York, NY, USA, 2016. Association for Computing Machinery.