

1. はじめに

近年、e-learning において講義ビデオの配信が盛んに行われている。藤吉研究室では、ハイビジョンカメラで撮影した高解像度映像からのトリミングによる講義映像の自動生成に取り組んでいる [1]。講師の動きのみから仮想カメラワークを決定しており、板書など講義の他の要素については考慮していないため、視聴者が本当に見たい対象とずれが生じてしまうという問題がある。

そこで、本研究では講義ビデオ中の視聴者の注目点について調査を行い、その特徴を考慮した仮想カメラワーク生成法について提案する。

2. 講義ビデオにおける視聴者の注目点

20 人の被験者に講義映像 (各 10 分 × 4) を視聴してもらい、講義中の注目点を調査した。本研究の目的は、講義映像のトリミングの際に、この注目点をより多く捉えることである。

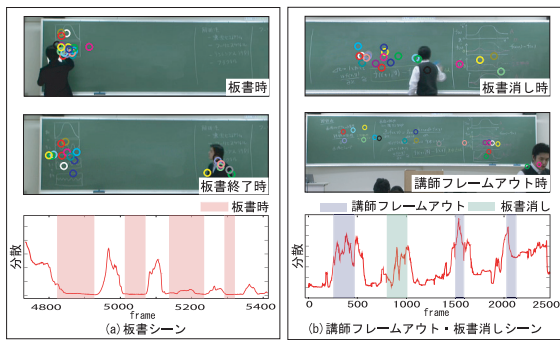


図 1：注目点の分散

図 1 における 4 シーンの実画像上にカラーのマークで示される点が、被験者から採取した注目点である。同図下に分散のグラフを示す。これらのシーンにて採取した注目点に、次の 2 つの特徴的な動きが確認された。

- 講師が板書時、視聴者の注目は板書領域に集中し、板書後には板書と講師の 2 領域に集中する (図 1(a))
 - 板書消し時や講師がフレームアウトしたとき、視聴者の注目は黑板上へ広く分散する (図 1(b))
- 前者の注目点の動きは、新規に書かれた板書を判読し終えるまでの時間に、個人差があるためだと考えられる。また、後者のシーンは講義の一区切りであるために、注目点の動きに板書を見直す動作が含まれていると考えられる。

3. 注目点を考慮したカメラワーク

採取した注目点データより得られた 2 つの視聴者の注目特徴から、仮想カメラワークへの応用を考える。

板書への注目固定時間の算出 板書後は板書領域に注目が固定されるため、その時間を決定する注目固定時間モデルが必要となる。そのため、採取した注目点データから板書後に注目が板書から離れるまでの時間を計測した (各 10 分 × 4、板書数計 42 回)。次に、注目固定時間を決める要因として板書の量が考えられるため、板書文字ピクセル数との関連を調査した。板書文字ピクセル数は、講師を消去した黑板画像間の差分によって検出したものを用いる。図 2 に注目固定時間モデルと板書文字ピクセル数との関係を示す。

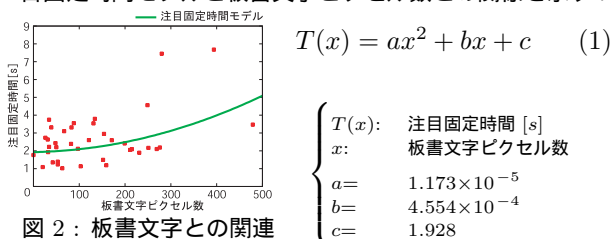


図 2：板書文字との関連

図 2 のプロット点に最小二乗法で 2 次関数をフィッティングした式 (1) を、注目固定時間モデルとして定義する。これを用いて、板書を考慮したカメラワークを生成する。板書を考慮したカメラワーク 仮想カメラワークを算出す

るための注目領域座標 $R(t)$ を、フレーム間差分により得られる講師位置 $O(t)$ 、板書位置 $B(t)$ 、注目固定時間モデル $T(x)$ により算出する。

$$R(t) = \begin{cases} B(t) : \text{板書検出時} < t < \text{板書検出時} + T(x) \\ O(t) : \text{otherwise} \end{cases} \quad (2)$$

式 (2) により算出された注目座標領域 $R(t)$ に対して、仮想パンニング生成を行う。以下にその手順を示す。

Step1 注目領域座標のぶれをバイラテラルフィルタにより抑制

Step2 零交差法による動作特徴点抽出

Step3 放送カメラマンの知見から得られた仮想パンニングモデルを用いて、トリミング位置を決定 (図 3(c))

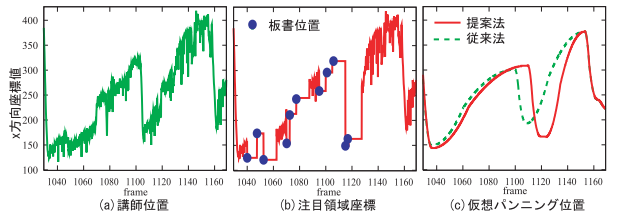


図 3：板書を考慮したカメラワーク生成

板書位置を付加することで、従来法と異なる動作特徴点が見られ、最終的なトリミング位置に変化が生じていることがわかる。

注目対象が不定となるシーン 講師が板書を消している時や、フレームアウトしてしまうシーンでは注目点の分散が大きく、黑板の広い範囲に分布しているため、通常のトリミングサイズ (512 × 384) では沢山の注目点を画面に捉えることが難しい。より視聴者の注目点に合う講義映像を生成するために、Yokoi 等による仮想ズームモデルに基づいてトリミングサイズを拡大してゆき、黑板全体を提示する。

4. 評価実験

提案手法により自動生成した講義ビデオの評価を行う。図 4 は板書後のシーンにおいて、注目点データをプロットし、画像中心との距離をヒストグラム化したものである。また、プレファレンススコアとして、提案法と従来法により生成された映像 (各 1 分 × 10) を 22 人の被験者に提示し、見易さの観点からどちらの講義ビデオが優れているかについてアンケートを行った。

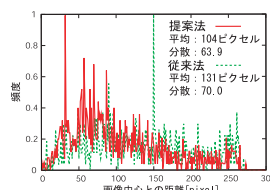


図 4：注目点との距離

表 1：アンケートによる調査結果

	得票率 [%]
本手法	56.1
従来法	32.1
その他	11.7

図 4 より、提案法による映像は平均距離および分散値が従来法に比べ小さく、視聴者の注目点により画像中心に近い位置に分布していることがわかる。また、アンケート調査から過半数の 56.1% の視聴者から従来法よりも良いという評価を得ることができ、 t 検定 (有意水準 5%) の結果、両者の間に有意差があることを確認した。

5. おわりに

本研究では、視聴者の注目を考慮した講義映像の生成法について提案した。今後はパワーポイントへの対応や、講師の姿勢を考慮したカメラワークの自動生成を目指す。

参考文献

[1] T.Yokoi et al.: "Generating a Time Shrunk Lecture Video by Event Detection", Proc. of ICME2006, 2006