Web教材の操作履歴を用いた講義動画像の自動インデクシング

田中 頼人 [1] 林 佑樹 [2] 萩野達也 [2] 千代倉弘明 [2]
[1] 慶應義塾大学 大学院政策・メディア研究科
[2] 慶應義塾大学 環境情報学部
〒252-8520 神奈川県藤沢市遠藤5322
e-mail: {yori,t01771yh,hagino,chiyo}@sfc.keio.ac.jp

概要

本稿では講義を録画して得られた動画像を自動的に単位化しインデクスを付与する手法を提案する。提案手法ではWebページとして作成した講義資料スライドや外部Webページを参照するタイミング情報をproxyサーバにより記録し、これを利用して動画像からインデクスを生成する。これにより人手による事後のインデクス付与作業を省き、効率的な講義動画像アーカイブの構築を支援する。

1 はじめに

広帯域ネットワーク環境の普及に伴い、教育現場での動画像メディアの活用に期待が高まっている。直感的に利用できる、容易に臨場感を得られるなどの利点を生かした講義の動画像を配信することにより学習者は受講における時間・場所の制約を回避することができる。しかしその反面、動画像は一覧性に乏しく目的の個所を探すことが容易ではない。事後利用の際には学習者が動画像全てを見ずに必要個所のみを見られることが望ましく、動画像を単位化しアクセスを支援するためのインデックスをいかに自動付与するかが課題となってきた。

動画像を有効利用するためのインデクシングの研究は近年多数行われているが、これらは移動物体認識、文字認識、音声認識等のパターン認識技術を用いたものが多くその精度は100%ではない。したがって自動付与されたインデクスを人手で修正するための作業コストが生じる。

このような背景から、我々は講師が用いる資料スライドを含むWebページを参照するタイミング情報をproxyサーバにより記録し、これを利用してインデクスを自動付与するシステムの開発を行った。Webブラウザ上でのページ切り替え操作を行った時刻に基づいてインデクスを生成することにより、事後の修正を必要とせずより実用に近い形での動画像アーカイブ構築を実現する。手間を伴わずにインデクスを生成することにより、学習者は講義終了後直ちに動画像を用いた復習を開始することができる。

本稿提案手法の目的は以下の通りである。

- 講師が単独で講義映像の録画からインデクス生成までを行うことができる
- 事後の編集作業を必要としない
- 学習者は興味に従って動画像中の見たい個所から再生を開始できる

2 関連研究

一覧性の乏しい動画像への効率的なアクセスを支援するために、動画像中に含まれる重要な情報を自動抽出する手法が提案されてきた。動画像を単位化するするためには動画像を時間軸上での区間に分割する必要がある。

動画像は複数のシーン(scene)から構成され、シーンは複数のショット(shot)から構成される。ショットは連続したフレーム(frame)から構成され、ショットの変化する個所はカット(cut)と呼ばれる[1]。各フレームは時間軸に従って並べられるため隣接したフレーム間では物理的特

HTTP-Proxy-Assisted Automatic Video Indexing for e-Learning
Yorihito Tanaka, Yuuki Hayashi, Tatsuya Hagino, Hiroaki Chiyokura
Graduate School of Media and Governance, Keio University
Fuculty of Environmental Information, Keio University

微量の類似度が高くなるが、ショットの変わり目であるカットにおいては類似度が低くなる。カットを検出するためのフレーム間の類似度を測定する方法として、対応する画素の輝度値の差分を用いる手法 [2]、ヒストグラムの差分を用いる手法 [3] が提案された。水平方向に幕が引かれるようにショットが切り替わるワイプ (wipe)、フレームが徐々に変化しながらある画像が他の画像に移行するディゾルブ (dissolve) 等の特殊なカットに対する検出手法も提案されている [4]。

動画像中の文字・音声情報も動画像の意味内容を端的に表している場合が多く、これらを認識して内容検索に役立てる研究も行われている。文字の含まれるフレームの検出や文字の切り出しを行う手法 [5][6]、音声認識によるテキストインデクス自動生成 [7]、動画像中の話者特定と追跡 [8] が提案されている。

これらの認識技術の教育への応用として吉田らは音声認識による講義動画像の自動インデクシングを提案した [9]。その他鳥山らによる遠隔講義を対象とした話者特定方式比較評価 [10]、小澤らによる講義動画像中のスライド同定 [11]、石塚らによる音声操作プロジェクタを用いた自動インデクシング [12] がある。

3 講義と録画のモデル

講義には様々な形態が考えられるが、本稿では撮影の対象とする講義を以下の形式に従うものとする。

1. 1つの教室内で講師1名と任意数名の学生によって行う
2. 講師は資料スライド提示用端末1台を用いる
3. 講師は資料スライドを HTML による Web ページとして事前に作成しておく
4. 黒板・ホワイトボードは用いず、文字や図形の記入はスライド提示用端末に接続したタブレットを用いて行う
5. スライド提示用端末に USB カメラが接続され、講師の表情や身振りはスライド提示用端末の画面上に表示される
6. 講師は必要に応じて資料スライド以外の Web ページ (以下「外部 Web ページ」とする) を資料スライド提示用端末上で参照することがある
7. スライド提示用端末上に表示された資料スライド、外部 Web ページ、タブレットによる記入内容、講師映像は合わせてプロジェクタを通じ教室内のスクリーンに投影される
8. 前述 (7) でのスクリーンに投影される動画像 (図 1) を録画し、講義終了後に非同期学習者に向けたストリーミング配信を行う

筆者らの研究グループは資料スライドと講師動画像を画面内画面 (Picture in Picture) により同時に録画するシステムの開発を行い、2002年春より運用を行ってきた [14]。この方式により図 2 に示すような 2 ウィンドウ構成のシステムに必要であった事後の同期作業が不要となり、講義終了とともに講義動画像が完成する。本稿で提案するシステムはこれを拡張し、録画された動画像に静止画像インデクスの自動付与を行うものである。

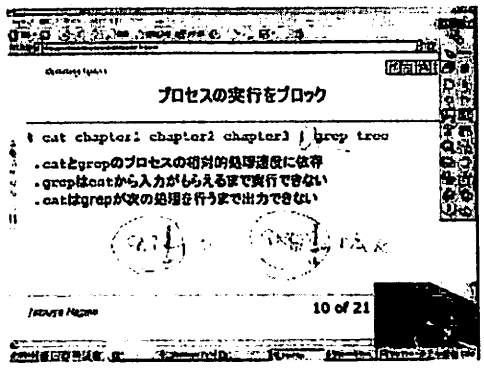

図 1: 録画する画面の例

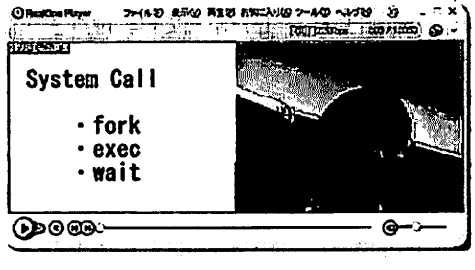

図 2: 2 ウィンドウ構成の画面例

4 システムの設計

4.1 構成要素

録画、インデクス作成、配信は3に示すように1台の端末と5台のサーバ、その他周辺機器から構成される。

スライド提示用端末 Webページとして用意された資料スライドを画面上に表示する。USBカメラ、タブレットからの入力も合わせて表示する。講師が持ち運び可能なラップトップ計算機が主に用いられる。

録画サーバ 動画像キャプチャ機能を備え、スライド提示用端末の出力画面を録画する。録画操作時には操作履歴取得proxyサーバに録画開始トリガを送信する。教卓内への常設を想定している。

Webスライド配信サーバ 事前に用意された資料スライドをHTTPにより配信する。

操作履歴取得proxyサーバ スライド提示用端末からWebページ(資料スライド及び外部Webページ)を参照する際のHTTP proxyとして機能する。資料スライド参照が生じた時刻と要求されたURLををデータベースサーバに送信する。

データベースサーバ 操作履歴取得proxyサーバから受信した時刻、URL情報を蓄積する。録画終了後には講義動画像配信サーバからの要求に従いそれらの情報を引き渡す。

講義動画像配信サーバ 録画サーバから転送された動画像ファイルとデータベースサーバに蓄積された時刻及びURL情報を用いてインデクスを生成し、ストリーミング配信を行う。インデクスは動画像よりフレームとして切り出した静止画像(図4)とし、学習者はこれらをWebページ上から選択することで目的の位置から動画像の視聴を開始できる。それを補助するものとしてWebスライド配信サーバ上の資料スライドから抽出した見出し文字列、外部Webページへのハイパーリンク、動画像先頭からの時間情報を併せて付与する。

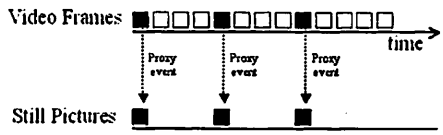

図 4: 静止画像の切り出し

4.2 通信遅延への対処

スライド提示用端末、操作履歴取得proxyサーバ、Webスライド配信サーバのHTTPによる通信は図5のように行われるが、図中①の通信が発生した時刻をWebスライドが参照されたタイミングとして記録すると②及び③での遅延が起こった場合にタイミング情報発生とスライド提示用端末上での画面変化の同期が取れなくなる。この問題を回避するために、本システムでは

- スライド提示用端末からの要求が生じた時刻①
- 操作履歴取得proxyサーバとWebサーバ間の通信に要した時間②+③

を合わせて記録することとした。

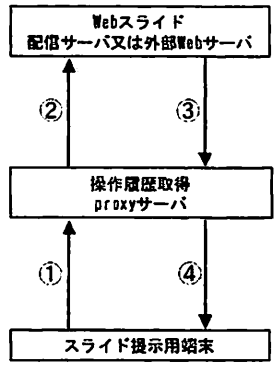

図 5: 通信遅延への対処

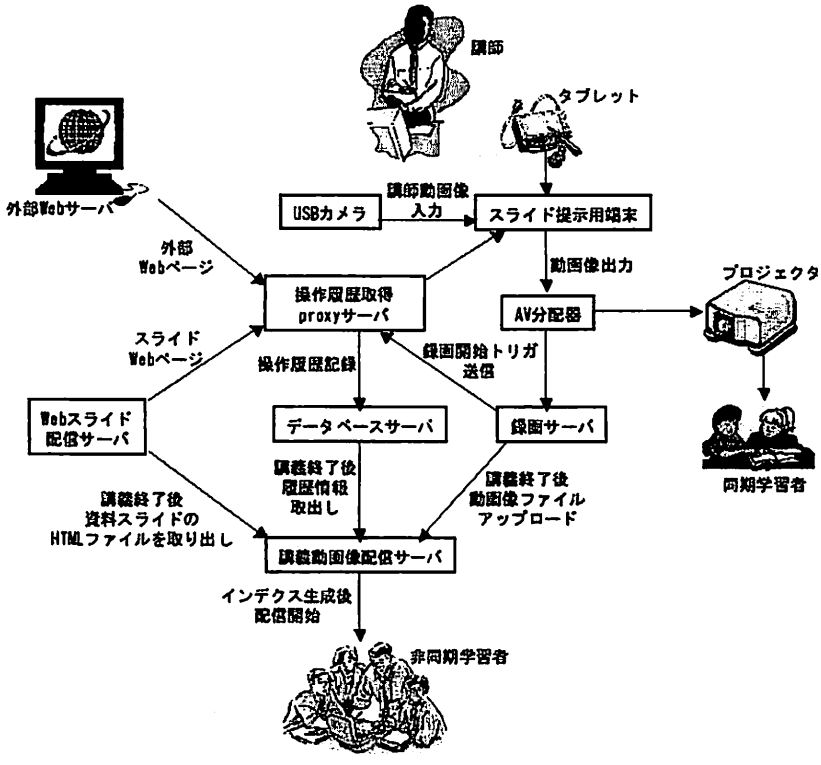

図 3: システム構成

4.3 システムの利用手順

講義開始から動画像配信開始に至るまでの処理の流れは以下のようになる。

1. 講師は資料スライドを画面上に提示して講義を開始し、同時に録画サーバを操作して録画を開始する。録画サーバは録画開始トリガを操作履歴取得 proxy サーバに送信する。
2. 録画開始トリガを受信した操作履歴取得 proxy サーバはその時刻をデータベースサーバに送信する。
3. 講師は Web ブラウザを操作して資料スライドを切り替え、タブレットによる入力を交えながら講義を行う。操作履歴取得 proxy サーバ側では資料スライドが参照された各時刻とその URL をデータベースサーバに送信する。
4. 講師は講義中の必要に応じて外部 Web ページを参照する。この場合も操作履歴取得 proxy サーバはそのページが要求された時刻と URL をデータベースサーバに送信する。
5. 講義終了後、講師は録画サーバ上に作成された動画像ファイルと Web スライド配信サーバ上の資料スライドを講義動画配信サーバ上に転送し、インデクス生成ソフトウェアで処理を行う。
6. インデクス生成ソフトウェアは録画開始トリガの発生した時刻と各 Web ページが配信された時刻との差分を用いて動画像から静止画像を切出し、これをインデクスとする。

4.4 学習者用インタフェース

インデクス生成ソフトウェアは

- 静止画像
- 動画像の先頭からの位置を示す矩形
- 資料スライド(HTML ファイル)から抽出した見出し文字列
- 外部 Web サーバ上のページへのハイパーリンク

からなるユーザインタフェース画面(図 6)を生成する。学習者は Web ブラウザ上で任意の静止画像を選択することにより、該当するショットからの動画像を再生することができる。静止画像の下部に付加された矩形は動画像の先頭からの経過時間を視覚化したものであり、学習者が講義時間中におけるスライドの相対的な位置を把握できるようになっている。見出し文字列の抽出は資料スライドの HTML ファイルから<H1>要素を抜き出すことによって行う。同一ファイル中に<H1>要素が複数存在する場合にはファイルの先頭に近いものを採用する。

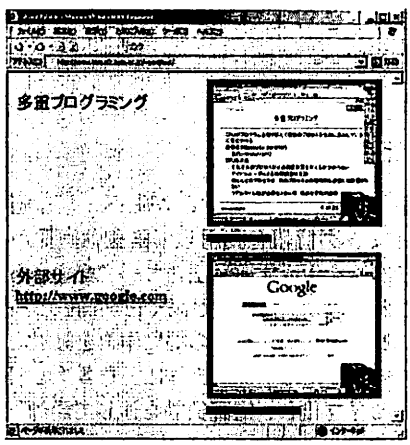

図 6: 生成されたユーザインタフェース

4.5 インデクス生成用インタフェース

講義終了後のインデクス生成作業は全て Web によるインタフェースを通じて行われ、

1. 講義動画像配信サーバへの動画像ファイルの転送
2. インデクス生成処理の開始
3. 生成されたインデクス用静止画像の確認

の 3 段階で完了する。動画像ファイルの転送に用いるインタフェース画面の例を図 7 に示す。

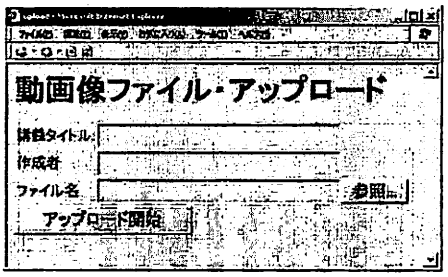

図 7: 動画像ファイル転送用インタフェース

5 実装と実験

4 節で述べた録画・インデクス生成システムを実装し、2003 年 7 月 8 日に本学で行われた学部講義を対象に実験を行った。

操作履歴取得 proxy サーバ	Vine Linux2.6,Apache
データベースサーバ	Vine Linux2.6 PostgreSQL
録画サーバ	WindowsXP Professional
講義動画像配信サーバ	Windows2000 Server

表 1: 実装システムの構成

表 1 の構成でシステムを構築し、Windows Media Video 形式による録画と jpeg 形式によるインデクス画像生成処理、ユーザインタフェース画面生成までを行った。動画像の画面解像度は 640 x 480pixels、フレームレートは 10fps でスライドの文字やタブレットからの入力文字を十分に判読できる品質である。動画像ファイルの容量は講義 1 コマ 90 分あたり約 250MByte である。

インデクス生成を行った講義動画像配信サーバは CPU: PentiumIV 3GHz、Memory: 1GByte の構成で、インデクス画像 1 枚の生成に要する平均時間は 534 ミリ秒であった。静止画像の切り出しには Windows Media Video のタイムコードインデクスを用いているため動画像の先頭からの時間が増加しても切り出し処理に要する時間は増加しない(図 8)。講義中に行われた

スライド参照20回分のインデクス画像生成とユーザインタフェース画面生成の合計処理時間は15秒程度であり、これは講師が単独で行うことが容易な範囲内であるといえる。録画サーバから講義動画像配信サーバへの動画像ファイル転送時間も含めて5分程度で全ての作業が完了し、学生は直ちにストリーミング配信による復習が可能となる。

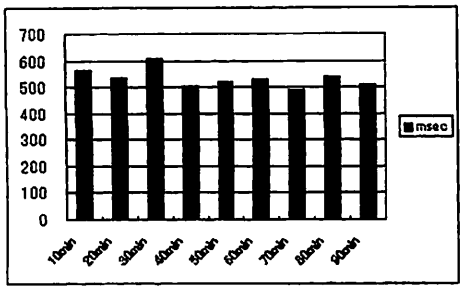

図 8: 静止画像切り出しの処理時間

6 考察

6.1 インデクス生成手法

小林ら[13]は講義動画像に対して時間を等分割した任意枚数の静止画像インデクスを生成する機能を実現したが、この方式では学習者が講義の内容に基づいて目的の個所を素早く探す操作を支援できていない。

パターン認識技術を用いて動画像へのインデクシングを行う方法は多数提案されているが、動画像の内容を認識し処理することは一般的に難しい。動画像中の各フレームの特徴量を用いたカット検出[2][3]や移動物体追跡による手法[15]はカメラ操作や照明条件の影響を受ける。隣接フレーム間の非類似度を用いた場合、資料スライド画面切り替えのような文字のみの変化は類似度が高いため検出されにくい。

動画像中の文字情報認識では文字の解像度が低くノイズが多いことから認識精度が低くなることが課題とされる。吉田ら[9]は音声認識によるテキストインデクス自動生成を講義動画像に適用したが、この方式では講義分野の専門用語・キーワードを予め講師が選択して登録しておく必要がありインデクス生成の性能はキーワードの選択に依存する。伊福部ら[16]は音声認識により講義動画像に字幕を付与する手法を提案したが、この方式では音声認識のための訓練を受けた補助者が講師の発話内容を復唱することを前提としており処理が全て自動化されているわけではない。

これらに対して本稿提案手法は講師の画面切り替えタイミングのみを用いておりパターン認識処理の難しさに起因する誤検出を避けることができる。画面切り替えは講師の明示的な意思によって行われるため、そのタイミング情報は画像や音声から抽出された特徴量よりも重要な要素である。

6.2 資料スライド作成環境

本稿提案手法ではWebページとして作成された資料スライドを用いる。Webページの作成方法は問われないため、資料スライドを作成する講師は特定のスライド作成用ソフトウェアの操作を本システムのために習得する必要はない。Microsoft PowerPoint、MagicPoint等の代表的なプレゼンテーション用ソフトウェアはスライドをWebページとして出力する機能を有している。

7 おわりに

本稿では講師が端末上でWebページ参照するタイミング情報を利用して講義動画像からインデクスを生成する手法について提案を行った。これにより、利便性が高く付加価値の高い講義の動画像配信が事後の編集作業を経ずに可能となる。今後は本システムにより多くの講義をアーカイブ化し、実証実験を通じて講師・学習者双方による評価を行う必要がある。

マルチメディア情報のメタデータ記述の枠組であるMPEG-7の標準化が行われており[17]、コンテンツの相互運用性の向上が期待されている。MPEG-7との融合によるコンテンツ流通の効率化は今後の課題である。

また、筆者らの所属する慶應義塾大学では講義情報の検索やオンラインレポート提出などをシームレスに行う支援システム"Web Learning

System" を運用中であり、このシステムとの連携も次の課題である。連携が適切に行えれば、学習者による講義動画像の視聴、必要な外部資料のダウンロード、レポートの提出、講義内容へのフィードバック送信という流れが一貫した操作体系の中で行えるようになる。

謝辞

本研究は文部科学省 ·21 世紀 COE プログラム ·研究拠点形成費補助金 「次世代メディア · 知的社会基盤」の援助を受けている。

参考文献

[1] 西尾章治郎, 田中克己, 上原邦昭, 有木康雄, 加藤俊一, 河野浩之,"情報の構造化と検索", 岩波書店,2000

[2] 大辻清太, 外村佳伸, 大庭有二,"動画カット検出", 電子情報通信学会画像工学研究会,IE91-116,pp.25-31(1991)

[3] Yoshinobu Tonomura,"Video handling based on structured information for hypermedia systems",ACM Conference on Multimedia Information Systems,pp.333-344,1991

[4] Zhang H., Kankanhalli A., Smoliar S., "Automatic partitioning of full-motion video,",ACM Multimedia Systems, 1, pp.10-28, 1993

[5] 新井啓之, 桑野秀豪, 倉掛正治, 杉村利明,"映像中のテロップ表示フレーム検出方法", 電子情報通信学会論文誌 Vol.J83-D-II no.6,pp.1477-1486,2000

[6] 桑野秀豪, 倉掛正治, 小高和己,"映像データ検索のためのテロップ文字抽出", 電子情報通信学会研究報告 PRMU96-98,pp.39-46,1996

[7] 南憲一, 阿久津明人, 浜田洋, 外村佳伸,"音情報を用いた映像インデクシングとその応用", 電子情報通信学会論文誌 Vol.J81-D-II no.3,pp.529-537,1998

[8] Lie Lu, Hong-Jiang Zhang,"Speaker change detection and tracking in real-time news broadcasting analysis",Proceedings of the tenth ACM international conference on Multimedia,602-610,2002

[9] 吉田孝博, 多田一基, 半谷精一郎,"講義ビデオのアクセシビリティを改善したブラウザとその評価, 信学技報 ET2002-47,pp.19-24,2002

[10] 鳥山朋二, 小林和則, 古家賢一, 中村智明, 西原功, 中野慎夫," 遠隔講義における話者特定方式の比較", 信学技報 ET2002-34,2002

[11] 小澤憲秋, 武部浩明, 勝山裕, 直井聡, 横田治夫,"文字認識を利用した講義動画中のスライド同定", 情報科学技術フォーラム FIT2002LI-5,pp.133-134

[12] 石塚健太郎, 亀田能成, 美濃導彦,"講義の自動撮影系における音声 ·映像インデキシング", 信学技報 PRMU99-258,pp.91-98,2000

[13] 小林裕之, 関秀行, 千代倉弘明,"簡易利用可能な授業の自動アーカイビング/配信システムの開発", 情報処理学会コンピュータと教育研究会 66-5,pp.31-36,2002

[14] 板宮朋基, 林佑樹, 千代倉弘明,"ワンマン録画可能な講義ビデオ作成システム", 情報処理学会研究報告 2003-CE-70,pp.17-20,2003

[15] 宮内新, 笹代孝次, 渡邊露文, 石川知雄,"移動オブジェクトの追跡による動画像インデクシングシステム", 映像情報メディア学会誌 Vol.54 no.3,pp.459-462,2000

[16] 株式会社ビー ·ユー ·ジー," 音声同時字幕システム", http://www.bug.co.jp/topics/todai.html

[17] 柴田正啓, "コンテンツ記述の標準化 MPEG-7", 情報処理,Vol.41,no.2,pp.176-182,2000