

手書き文字認識を用いた講義動画のフレーム検索

重宗克啓* 岡村健史郎**

Flame Identification for Lecture Video Using Handwritten Character Recognition

Katsuhiko SHIGEMUNE and Kenshiro OKAMURA

Abstract

Recently, e-Learning systems which contain the lecturer video of face-to-face class are widely used in the field of education. For high-level educational effects, these contents should include the document images which are presented in synchronization with the frame of the lecturer video. However, creating this synchronization of contents takes so much cost. This paper proposes a method of making the synchronization between the string in the document and the frame of the lecture video using handwritten character recognition technology.

Key words: e-Learning, Lecture video, Document image, Character recognition

1. はじめに

現在、多くの教育機関がインターネットを用いて講義動画を配信している^①。動画内容は教室で行った講義をカメラで撮影したものである。この形式の e-Learning は、インターネットが利用できる環境があれば、いつでも、どこでも学習できるという利点がある。しかし、単に講義動画を流すだけでは、学習者が飽きやすく学習意欲を保つことが難しいという欠点もある。

学習意欲を保つための手法として、講義動画とプレゼンテーション資料の各スライドを同期させて画面に表示する同期再生コンテンツが開発されている^②。同期再生コンテンツとは、講義中の講師を撮影した動画と説明に用いたスライドを同期させて画面に表示するものである。しかし、このようなコンテンツを作成するには、スライドが切り替わるフレームを探し出して、メタデータを作成する必要があり、コンテンツ作成に大変なコストがかかる。

この欠点に対処するため、文字認識技術を利用して自動的にスライド同定を行う手法が提案されている^③。この手法は、プレゼンテーション資料から抽出したテキストデータと動画中のスライドを文字認識したデータからそれぞれの位置情報に信頼度を加えてマッチングすることで、切り替えのタイミングを検出するものである。しかしながらこの手法は、スライドに関する XML などを書かれたテキスト情報が必要で、黒板を使った通常の対面授業には利用できない。依然として多くの教育機関において黒板の板書を用いた授業が行われていることを考慮すると、これらの授業を録画し、コンテンツとして提供できる仕組みが必要となる。

そこでスライドを用いた授業だけでなく、黒板の手書

き板書を用いた授業を対象に、講義資料の代わりに、教室での授業において学習者が黒板を写した学習ノートを利用するよう考えた。これにより、教師が用意した手書きの資料やワープロなどで作成した講義資料にも対応できるようになるだけでなく、教師が用意した講義資料が無い場合でも短時間でコンテンツを作成できるようになる。

本研究が構築を考えているシステムについて説明する。コンテンツ提示画面の左側に視点固定型のカメラで撮影した講義動画を、右側に教室授業で黒板を手書きした学習ノートあるいは教師が作成した講義資料をスキャナで取り込んだものをそれぞれ配置する。コンテンツ提示画面例を図 1 に示す。講義動画には黒板を中心に教師と一緒に撮影したものをを用いる。右側の学習ノートに書かれたそれぞれの文字列には学習者が見たい画面を選んで視聴できるように、その文字列を説明している講義動画のフレームにリンクをはる。このリンクを学習者がブラウザ上でクリックした際に、その文字を説明しているフレームから再生を開始する。

学習ノート文字列と講義動画のフレームを関連づける仕組みについて説明する。まず、フレームおよび学習ノートのそれぞれの画像から、文字列の切出しと文字認識を行い、テキスト情報を抽出する。そのテキスト情報を元に、学習ノートの文字列とフレームの文字列がどれだけ似ているかを表す指標である類似度を計算する。この時、教師が雑に文字を書く、文字と図形が区別できない等の理由から、文字の切り出し誤りや文字の認識誤りが多く発生する。そこで、文字切り出し誤りに対処するため文字の挿入や削除を考慮した動的計画法を文字列の比較に用いた。更に、文字認識誤りに対処するために上位

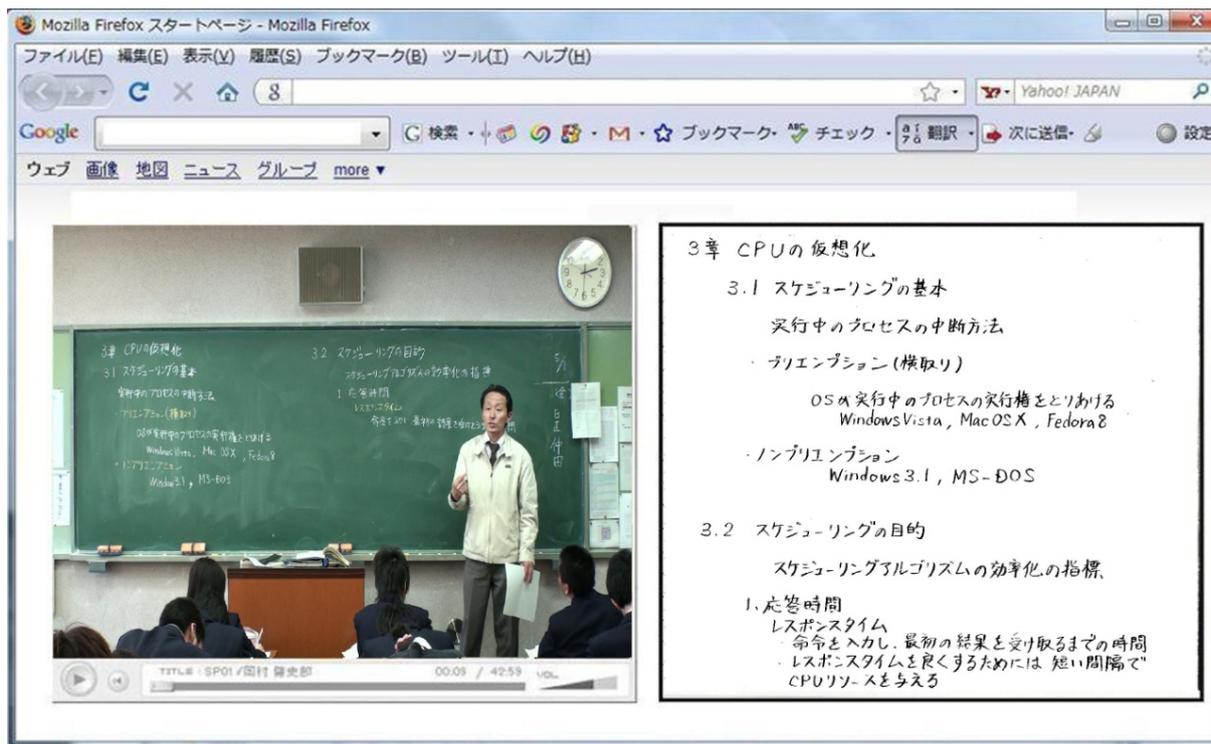


図1 コンテンツ提示画面の例

N位の認識結果までを用いて対応を求める。この二つの処理を用いて類似度を定義した。本論文では、この類似度を用いて資料の文字列とフレーム画面との対応表を作成し、フレーム検索が可能であることをいくつかの実験結果から説明する。

2. 提案するe-Learningシステム

2.1. e-Learningシステム概要

筆者らが構築を考えているe-Learningシステムについて説明する。コンテンツ作成のために講義の様子を視点固定型カメラで黒板を中心に撮影し、講義終了後、教師が板書するために用いた講義資料をスキャナで画像にする。この講義資料を事前に用意しない教師が存在することを考慮し、講義資料の代わりに学習者が黒板を写した学習ノートでも代替できるようになる。これにより講義資料がない講義でも容易にコンテンツ化することが

できるようになる。

講義を撮影した動画を見ながら学習するe-Learningシステムは、学習者が学習したいと思う場面を容易に選択して視聴する機能が重要となる。この仕組みとして学習ノートの文字列リンクを考えた。このリンクは、図1のコンテンツの提示画面において、右側の学習ノートの文字列をクリックすることで、図1の左側の講義動画がその文字列を説明している場面に移動する機能である。この機能は、学習ノートの文字列が講義動画のどのフレームに最初に表れるかを検索することにより実現する。この処理をフレーム検索と呼ぶ。

2.2. フレーム検索概要

フレーム検索の概念を図2に示す。フレーム検索を行うには、学習ノートの文字列がどのフレームに存在するかを調べる必要がある。この方法として、学習ノートにある文字列とフレームの黒板に書かれた全ての文字列との類似度をしきい値と比較し、類似度が大きい場合は、その文字列はフレームに存在するとし、類似度が小さい場合は存在しないとする。例えば、図2において、学習ノートの文字列「3章 CPUの仮想化」を、フレーム1~10に書かれた全文字列と比較した結果、フレーム4~9にある文字列との類似度がしきい値より大きい場合、「3章 CPUの仮想化」の内容は、フレーム4から説明されていると予想できるため、フレーム4にリンクをはる。

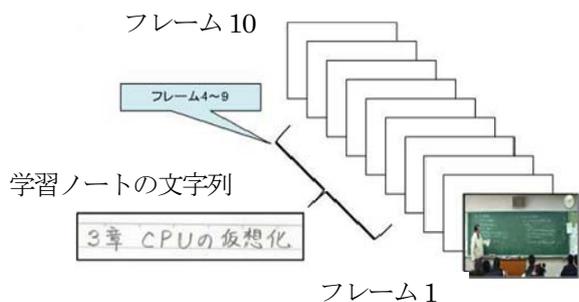


図2 フレーム検索の概念

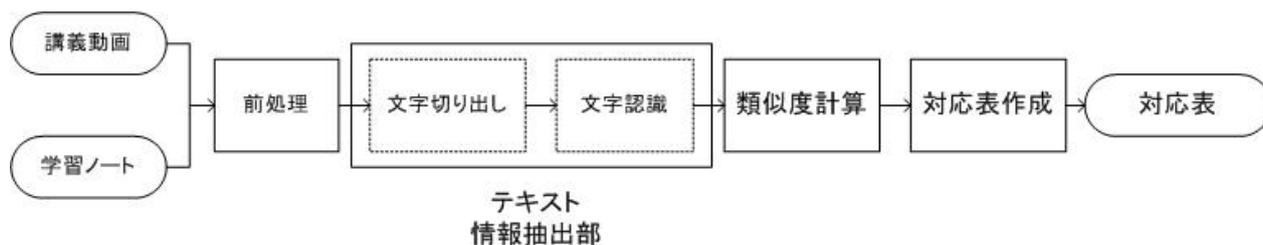


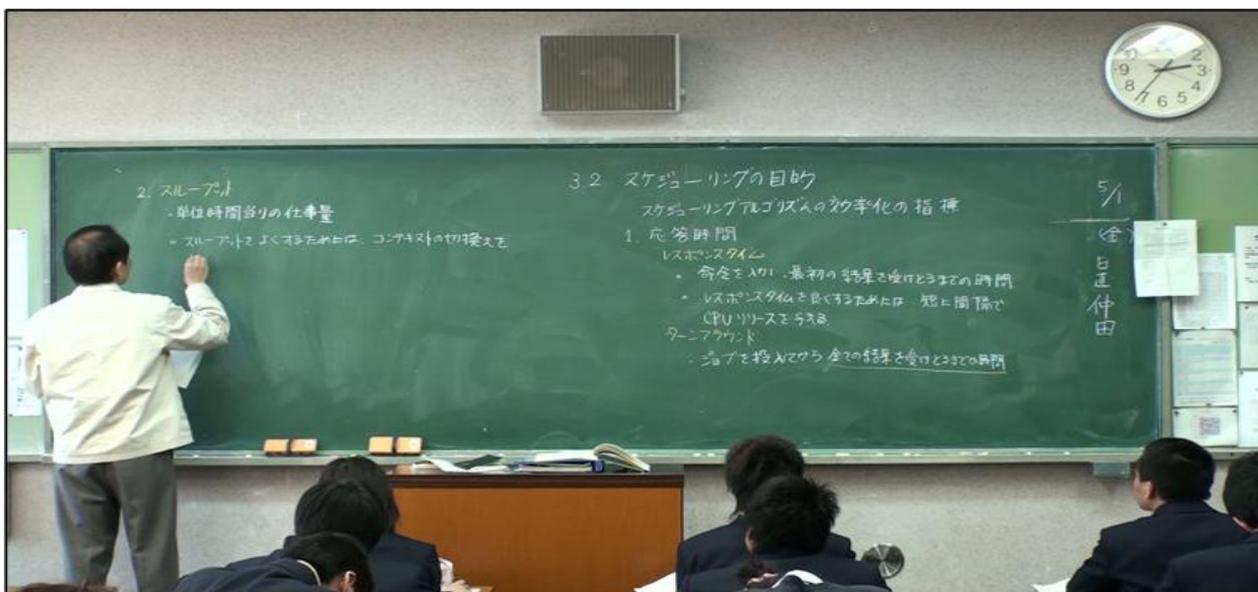
図 3 全体構成

フレーム検索を行う前に、まず、講義動画、学習ノートに対して、黑板領域の抽出や学習ノートの罫線を削除する前処理を行う。次に、前処理されたデータをテキスト情報抽出部に渡し、文字認識結果を得る。最後に、学習ノートの文字列とフレームとの対応表の作成を行う。全体構成を図3に示す。以下の項で、その詳細を述べる。

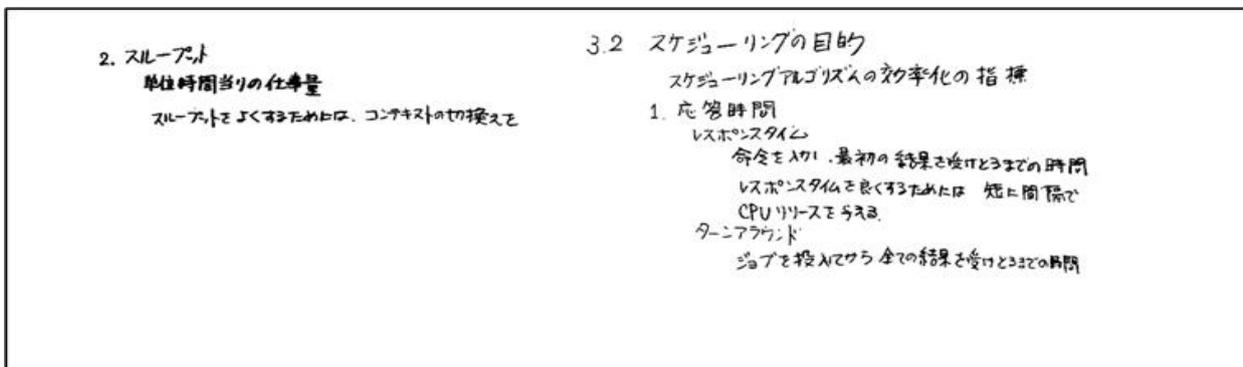
2.3. 前処理

前処理では、フレームと学習ノートから教師や図を除

去した文字情報のみの画像を作成する。まず、撮影した講義動画をフレームごとに画像に変換する作業を行う。フレーム画像の一例を図4(a)に示す。フレームの画像は、黑板領域を抽出した後、文字認識を行うために教師や図を削除した後に二値化する。処理結果を図4(b)に示す。次に、黑板を手書きした学習ノートあるいは教師が作成した講義資料をスキャナで画像にする。ノート画像には罫線や図が含まれるため二値化した後、これらを削除する。処理結果を図5に示す。



(a) 講義動画の原画像



(b) 講義動画の前処理結果

図 4 講義動画に対する前処理

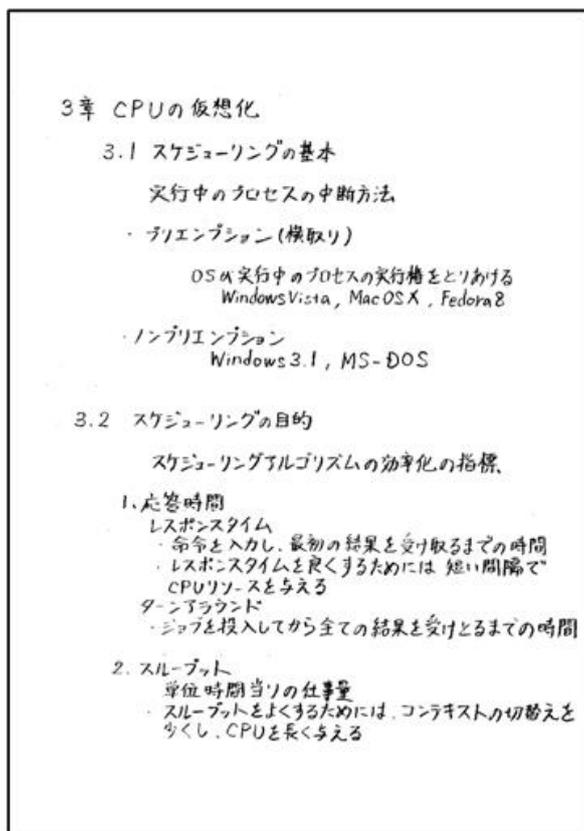


図 5 学習ノートに対する前処理結果

現状では、この前処理部分の構築が未完成であるため汎用の処理画像ソフトを用いて手作業で処理を行っている。

2.4. テキスト情報抽出部

テキスト情報抽出部では文字切り出しと文字認識を行う。文字切り出しでは、二値化した対象画像に対して文字切り出しを行った後、各切り出し文字をまとめて文字列を抽出する。この処理には文献4にある画素密度を用いた文字切り出し手法を用いた。図6に文字と文字列の切り出し結果を示す。内枠が切り出した個々の文字を、外枠が文字をまとめた文字列をそれぞれ表す。図6において、(a)はそれぞれの文字を正しく切り出しているが、(b)では複数の文字を誤って一つの文字として切り出している。文字切り出しではこのような切り出し誤りが発生する。

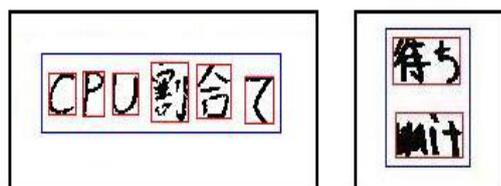
文字認識部では、文字列内にある文字を順に文字認識する。この処理には文献5にある手法を用いた。表1に図6(a)にある6個の切り出し文字画像に対する上位第1位から第5位までの文字認識結果を示す。表1において、○印は認識として正しい文字クラスを表す。6文字中、第1位の認識率は50%と低いが、上位5位までの累積認識率は100%まで高くなる。ここで、文字サイズは64×

64に正規化し、文字認識の特徴には加重方向ヒストグラム、識別にはマハラノビス距離をそれぞれ用いる。

ここで、表2に、講義動画図4(b)と学習ノート図5(b)を対象にした文字切り出しと文字認識の成功率をそれぞれ示す。文字切り出しの成功率は、複数の文字を誤って一つの文字に統合してしまう誤りや、一つの文字を分割してしまう誤りが発生したため平均85.6%になった。次に、正しく切り出された文字に対する文字認識率は、全体として55.1%になった。特に講義動画の文字認識率は43.4%と低くなっている。これは、講義動画の解像度が低く一文字の文字サイズが25×25程度と小さくなったこと、学習ノートに比べ雑に書かれたこと、などが原因である。文字切り出しと文字認識が両方ともに正しい割合は全体として48.0%と非常に低くなっている。このよ

表 1 切り出し文字に対する認識結果

		文字番号					
		1	2	3	4	5	6
認識順位	1	◎	戸	し	⊙	台	㊦
	2	こ	Ⓟ	り	筈	含	で
	3	密	‘	じ	罰	Ⓢ	察
	4	[胃	Ⓣ	謝	舎	マ
	5	亡	F	圃	訊	舌	こ



(a)切り出し 成功例 (b) 失敗例

図6 切り出し文字と文字列

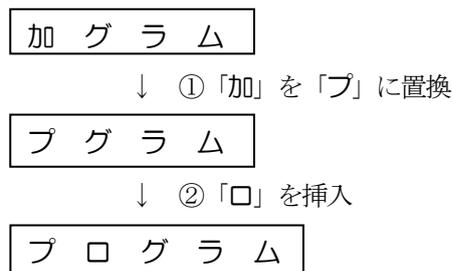


図7 文字の削除・挿入の例

うな文字切り出し誤りや文字認識誤りに対処するため、動的計画法と複数の認識結果を用いて文字列の類似度を定義し、フレーム検索することを考えた。

3. 類似度計算と対応表の作成

3.1. 動的計画法による距離計算

3.1.1. 二つの文字列間の距離計算

フレーム検索では、学習ノートの文字列の認識結果と講義動画フレームの文字列内にある文字の認識結果から文字列間の近似照合を行うことによって、学習ノートの文字列がどの講義動画フレームに存在するかを判定する。この近似照合において、文字切り出しにおける切り出し誤りに対処するため、文字を挿入・削除しながら対応を求める動的計画法を用いる⁶⁾。

図7に二つの文字列間の距離計算の例を示す。図7において、一方の文字列の認識結果が「加グラム」、もう一方の文字列の認識結果が「プログラム」とする。このとき二つの文字列が一致するように、挿入、削除および置換処理を行う。これら三つの処理のコストは全て1とする。図7では、まず、文字列1の「加」を「プ」に置換する。次に、文字列に「ロ」を挿入する。これにより二つの文字列が一致する。図7の例では、置換が1か所、挿入1か所ある。よって二つの文字列間の距離は2となる。二つの文字列A, B間の距離は動的計画法により、式(1)~(5)で求める。

$$M(i, j) = \min \begin{cases} M(i, j-1) + 1 & \dots (1) \\ M(i-1, j-1) + P(i, j) & \dots (2) \\ M(i-1, j) + 1 & \dots (3) \end{cases}$$

$$P(i, j) = \begin{cases} 0 & \text{if } A[i] = B[j] \\ 1 & \text{if } A[i] \neq B[j] \end{cases} \dots (4)$$

$$d(A, B) = M(L_A, L_B) \dots (5)$$

A[i]は文字列Aのi番目の切り出し文字に対する第1位の認識結果、B[j]は文字列Bのj番目の切り出し文字に対する第1位の認識結果である。また、文字列A, Bの文字列長をそれぞれL_A, L_Bとする。図8に行列Mの例を示す。行列MのM(i, 0), i = 0~L_A, M(0, j), j = 0~L_B、には初期値としてi, jの値を用いる。行列Mの他の要素は式(1)~(4)を用いて繰り返し計算し、最終的にM(L_A, L_B)が二つの文字列A, Bの距離d(A, B)となる。ここで、式(1)は文字の挿入に、式(2)は文字の置換に、式(3)は文字の削除に、それぞれ対応している。また、式(2)のP(i, j)は置換処理に対するコストを表し、A[i]とB[j]が一致している場合には0、不一致の場合は1となる。

3.1.2. 第N位までの認識結果を用いた二つの文字列間の距離計算

講義動画や学習ノートに対する文字認識においては第1位の認識結果に対する認識率は非常に低い。そこで、上位N位までの認識結果を用いることで認識誤りに対処する方法を考えた。具体的には式(4)の置換コストを式(6), (7)に示すように変更した。

		加 グ ラ ム					
		i	0	1	2	3	4
j	0	0	1	2	3	4	
	プ	1	1	2	3	4	
	ロ	2	2	2	3	4	
	グ	3	3	2	3	4	
	ラ	4	4	3	2	3	
	△	5	5	4	3	2	

図8 行列Mの例

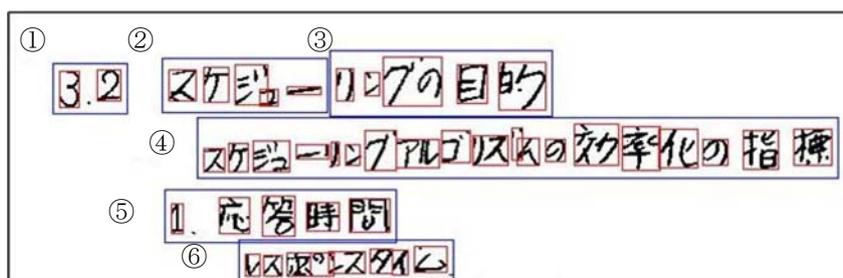


図 10 講義動画フレームからの抽出文字列



文字列 1

文字列 2

図 11 学習ノートからの抽出文字列

4. 評価実験

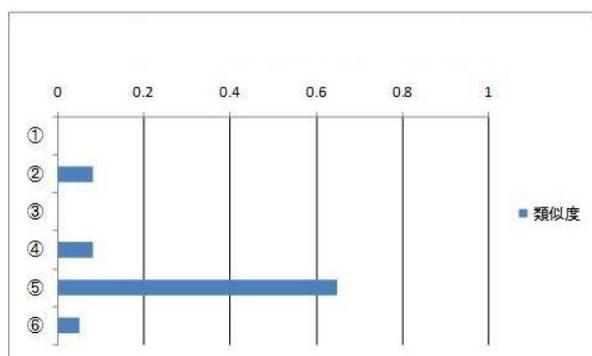
4.1. 予備実験概要

フレーム検索において、文字列がフレームに存在することを判定する式(9)のしきい値 α を決定するために予備実験を行った。大島商船高等専門学校で行われた「オペレーティングシステム」の講義から、比較的図が少なく文字が多く書かれた黒板領域の一部を講義動画フレームの入力画像とした。動画は SONY 製デジタルビデオカメラ HDR-SR12 を用い録画画質は 1920×1080 画素で撮影した。画像から丁寧に書かれた文字領域を手作業で抜き出し、二値化した後、文字切り出しと文字列抽出を行った。抽出した 6 個の文字列を図 10 に示す。図 10 の 6 個の文字列には①～⑥の番号を付けた。学習ノートは、講義で学生が作成した B5 サイズのノートをスキャナで画像にしたものを使用した。画像サイズは 1000×1400 画素である。この画像から罫線を削除し、講義画像

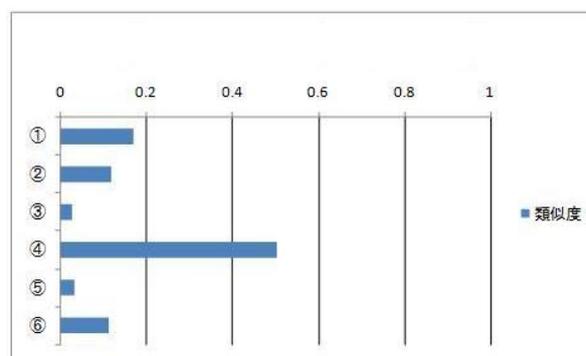
と同様の処理によって文字列抽出を行い、その中から丁寧に書かれた短い文字列と長い文字列をそれぞれ検索する文字列とした。使用した二つの文字列の画像を図 11 に示す。なお、文字認識結果は上位 5 位までを用いた。

4.2. 予備実験の結果

図 10 のフレーム画像内の 6 個の文字列①～⑥と学習ノートの二つの文字列間の類似度をそれぞれ調べた。結果を図 12 に示す。学習ノートの文字列 1 をフレームの文字列①～⑥と比較して類似度を求めた結果が図 12(a)である。学習ノートの文字列 1 はフレームの文字列⑤に対し最も高い類似度 0.65 を示した。一方、他の文字列①～④および⑥に対する類似度は 0.1 未満となった。次に、学習ノートの文字列 2 とフレームの文字列①～⑥とを比較して類似度を求めた結果が図 12(b)である。学習ノートの文字列 2 はフレームの文字列④に対し最も高い類似度 0.51 を示した。その他の文字列に対する類似度は 0.2



(a) 文字列 1 との類似度



(b) 文字列 2 との類似度

図 12 類似度の計算結果

表 4 実験結果
(a) 学習ノートの文字列(1~15)の対応表

		学習ノートの文字列														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
フレーム	1	・	・	・	・	・	・	・	・	・	・	・	・	・	・	・
	2	○	・	・	・	・	・	・	・	・	・	・	・	・	・	・

(b)学習ノートの文字列(16~36)の対応表

		学習ノートの文字列																				
		16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
フレーム	1	・	・	・	・	・	・	・	・	・	・	・	・	・	○	・	・	○	・	・	・	・
	2	○	・	○	○	○	○	○	・	○	・	・	・	・	・	・	・	・	・	○	・	・

未満となった。この結果から評価実験では0.3を式(9)のしきい値に用いることにした。

4.3. 評価実験

評価実験では、講義動画を使って、学習ノートの文字列がどのフレームに存在するかを判定するフレーム検索が可能であることを確認する。講義動画フレームには、予備実験で使用した講義動画から2フレームを検索対象画像として用いる。黒板領域を抽出し、二値化した後、手作業にて教師、図形部分を削除した画像を図13に示す。図13(a)は1692×550画素、図13(b)は1684×552画素である。学習ノートには、講義の黒板内容を写した学習ノートを用いる。学習ノートの画像サイズは985×1415画素で、文字列抽出処理を行い36個の文字列を抽出した。抽出結果を図14に示す。図14にある文字列1~15は図13のフレーム1, 2のいずれにも存在しない文字列である。一方、残りの文字列16~36の文字列はフレーム1には存在せず、フレーム2にのみ存在する文字列である。したがって、文字列1~15はいずれのフレームにも存在しないという結果が、文字列16~36はフレーム2にのみ存在するという検索結果が出ることを実験で確認する。実験において、類似度計算の文字認識結果には上位5位までを用い、存在判定の式(9)におけるしきい値 α は予備実験の結果から0.3とした。

4.4. 評価実験の結果と考察

実験で得られたフレーム検索の結果を表4に示し、それらの正解率をまとめたものを表5に示す。表4(a)は、学習ノートの文字列1~15をフレーム1, 2に対して検索した結果である。表4(b)は、学習ノートの文字列16~36をフレーム1, 2に対して検索した結果である。対応表の○印は学習ノートの文字列がフレームに存在すると判定

したことを、・印は存在しないと判定したことを表す。

文字列1~15はフレーム1, 2に存在しないため、表4(a)はフレーム1, 2ともにすべて・印となるべきである。表4(a)を見ると15文字列中14文字列が両フレームとも・印であり、正解率は93.3%である。一方、学習ノートの文字列16~36はフレーム2にのみ存在する文字列であるため、表4(b)の上段はすべて・印、下段はすべて○印になるべきである。表4(b)を見ると21文字列中8文字列を正しく判定している。正解率は38.1%である。これらを平均すると、学習ノートの文字列1~36に対するフレーム1, 2の文字列のフレーム検索の正解率は61.1%となる。

2.4の表2で示したとおり、テキスト情報抽出部の総合正解率の平均は48.0%であった。このようなテキスト抽出部の正解率が低い画像に対して、動的計画法を用いた類似度計算が寄与した点について具体的な例を挙げて説明する。図15に示した学習ノートの文字列20は「リ・ン・グ・ア・ル・ゴ・リ・ズ・ム・の・効・率・化・の・13指・標」と16文字で切り出されている。一方、フレーム2に存在する文字列は「ス・ケ・ジュ・ー・リ・ン・グ・ア・ル・ゴ・リス・“・ム・の・効・率・化・の・指・標」の19文字で切り出されている。なお、学習ノートの文字列20は文字列抽出の誤りにより、前半部の「スケジュール」が別の文字列として抽出されている。このような文字の切り出し誤りおよび文字列抽出の誤りを持つ

表 5 対応表の正解率

学習ノート	文字列数	正解文字列数	正解率 [%]
文字列1~15	15	14	93.3
文字列16~36	21	8	38.1
合計	36	22	61.1



図 15 存在判定の成功例

文字列に対し、動的計画法を用いた距離計算を行った結果、距離は 12.0 となった。類似度は 0.37 となり、表 4(b) に示すように、学習ノートの文字列 20 はフレーム 2 に存在すると判定した。

続いて、式(6)、(7)を用いて上位 5 位までの認識結果を使った効果を考察する。黒板文字は雑であることが多く、表 2 に示すように講義動画に対する文字認識率は 43.4% と低い。そのため、第 1 位の認識結果だけでは二つの文字列が一致する確率は低くなる。例えば、図 15 のフレームの文字列では、第 1 位の認識結果のみの場合の類似度は 0.16 であるが、上位 5 位の認識結果を用いると類似度は 0.37 になる。この結果、上位 5 位の認識結果を用いた場合は、しきい値 0.3 を上回るため、存在すると判定できた。

次に判定誤りの原因について述べる。シミュレーション実験において判定の誤りの主な原因は、テキスト情報抽出部の誤りである。これは、フレームの文字列抽出と学習ノートの文字列抽出の結果が異なっている場合、類似度が大きく異なるためである。現在、テキスト情報抽出部の処理は、文献 4、文献 5 で用いた手法を適応している。この手法は郵便宛名住所画像を対象としたものであり、学習ノートや講義動画の画像には最適化されていない。これらの講義動画や手書きノートに対応するために、文字サイズや教師の書き方などに関する事前知識を導入した文字列抽出が必要である。

検索率向上のための手段として、対応表の作成後に検索結果を修正することが可能と考えられる。隣接する講義フレームの内容は似通っている。そのため学習ノートにある文字列の検索結果が連続する画像フレームの途中で違う場合は、同じようになるように修正する方法が考えられる。例えば、表 4(b) の S₁₇ や S₂₃ は前後の関係から判定誤りであると推測でき、●印を○印に変更することで正解率を向上できると考えられる。

5. おわりに

本論文では、講義動画を配信する e-Learning システムにおいて、講義動画フレームと学習ノートの文字列との対応を求める方法を提案した。まず、二つの文字列の近似照合において類似度を定義した。この類似度は、文字列サイズに依存することなく 0~1 の範囲に入るようにした。この類似度を用いてフレーム検索処理の一部である対応表を作成し、文字列に対応するフレームを検出した。類似度を求めるにあたり、動的計画法を用いたこ

とで、文字の切り出し誤りおよび文字列抽出の誤りを持つ文字列の検索が可能となった。さらに、文字認識結果に上位 N 位までの認識候補を用いることで認識率の低い文字にも対応が可能になった。

本手法が可能であることを示すため、講義動画を使って、学習ノートの文字列がどのフレームに存在するかを判定するフレーム検索を行った。その結果、61.1%の正解率を得た。テキスト情報抽出部の総合正解率が 48.0%であることを考えると比較的良好な結果が得られたと思われる。

今回のシミュレーション実験では 2 つのフレームに対して検索を行ったが、今後、多くのフレームを対象にした実験を行う予定である。また、現在、マニュアルで行っている文字領域の抽出を自動化する必要がある。

参考文献

- [1] 独立行政法人メディア教育開発センター：eラーニング等の ICT を活用した教育に関する調査報告書(2007 年度)、<http://www.nime.ac.jp/reports/001/>
- [2] 八重樫理人 他：講義コンテンツ自動作成システムの開発、電子情報通信学会論文誌 D, Vol.91-D No.9, pp.2280-2292, 2008.
- [3] 武部浩明, 小澤憲秋, 勝山裕, 横田治夫, 直井聡. : 文字認識技術を利用した講義動画のスライド同定、電子情報通信学会論文誌 D, Vol.91-D No.12, pp.2819-2832, 2008.
- [4] 岡村健史郎, ユジンゴンザレスクルズ, 佐長康久, 浜本義彦. : 画素密度検出エージェントによる文字列の抽出と文字切り出し、大島商船高等専門学校紀要, 第 36 号, pp.43-54, 2003.
- [5] 田中裕貴, 末弘光次郎, 岡村健史郎. : 画素密度による文字切り出しを用いた手書きあて名認識システム、日本機械学会中国四国学生会第 35 回学生員卒業研究発表講演会公演前刷集, pp. 211, 2005.
- [6] 森本隼人, 岡村健史郎, 田中裕貴, 斉藤誠. : 認識率の低い文字を含む手書き宛名画像を対象とした文書理解システム：電気・情報関連学会中国支部第 57 回連合大会講演論文集, pp.111-112, 2006.