

スライドとの同期による講義映像のデータベース化

柳沼 良知

インターネットやコンピュータの利用環境が整うにつれて、大学等で、講義映像のデータベース化やWeb配信が行われるようになってきた。このような講義映像データベースでは、映像を検索する場合にメタデータが必要となる。しかし、このようなメタデータをすべて手動で付与するのは困難であり、メタデータ付与の自動化が必要となる。このため、本稿では、画像検索技術を用いて講義映像とスライドの同期を行うことで、講義映像へのメタデータ付与の自動化を実現した。具体的には、講義映像に対応するスライドを画像検索により選び出し、そのスライド中の文字情報を、メタデータの種類であるRSSへと変換することで、講義映像の検索を実現した。また、評価実験を行い、講義映像とスライドの同期を99.7%の精度で行えることを明らかにした。

キーワード

スライド、講義映像、データベース、画像検索、同期

1. はじめに

近年、インターネットやコンピュータの利用環境が整うにつれて、日本オープンコースウェア・コンソーシアムJOCW (2008) やインターネット教育支援サービスTIES (2008) 等、講義映像のデータベース化やWeb配信が行われるようになってきた。このような講義映像データベースでは、映像を検索する場合にメタデータが必要となる。しかし、このようなメタデータをすべて手動で付与するのは困難であり、メタデータ付与の自動化が必要となる。

このような講義映像に対するメタデータ付与に関しては、映像と書き起こしテキストを同期させることによるキーワード付与手法 (森本ほか 2005) や映像中のテロップ認識によるキーワード抽出手法等が提案されている (柳沼ほか 2005)。一方、画像検索技術に関しては、従来から、映像シーンの分類 (Lienhart et al. 1998; 長坂ほか 1996; 杉山ほか 1998)、検索のための自動キーワード付与 (谷口ほか 2001; Wactlar et al. 1996; 柳沼ほか 1996) 等について研究が行われてきた。

本稿では、講義映像に対するメタデータ付与手法として、画像検索技術を用いて講義映像とスライドの同期を行うことで、講義映像へのメタデータ付与の自動化を実現した。具体的には、講義映像に対応するスライドを画

像検索により選び出し、そのスライド中の文字情報を、メタデータの種類であるRSS (Really Simple Syndication) へと変換することで、講義映像の検索を実現した。

具体的な枠組みは、図1のようになる。

- 1) まず、講義映像中に映っているスライド部分と、その講義に用いられたスライドとの同期を行う。これは、画像検索技術を用いることで実現することができる。
- 2) 次に、1枚1枚のスライドから、検索用のメタデータとして利用する文字情報の抽出を行う。
- 3) 抽出した文字情報を、メタデータの種類であるRSSへと変換することで、講義映像の検索を実現する。

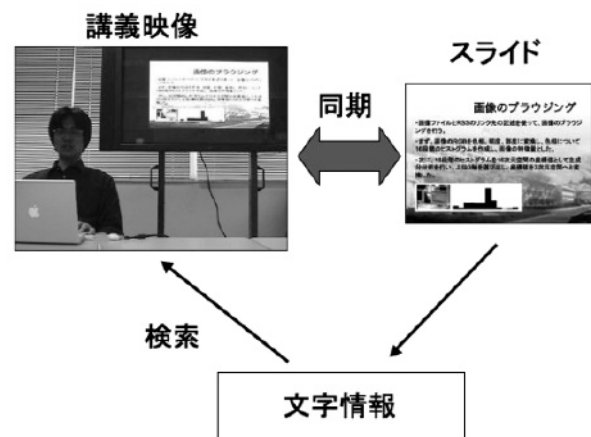


図1 メタデータ付与の枠組み

このRSSを用いたテキスト検索機能の実現については、2章で述べる。また、講義映像とスライドとの同期手法については、3章で述べる。

2. RSSを用いたテキスト検索機能の実現

Webサイトに関する見出しや要約などのメタデータをXML形式で記述し、Webサイトの更新情報などを効率的に公開する技術としてRSSがあり、ニュースの配信やBlogの更新情報の配信などに用いられている。最近では、Internet Explorer 7, Firefox 2, Safari 2等のWebブラウザがRSSに対応している。

スライドから抽出したテキスト情報は、図2のようなRSSファイルへと変換した。PowerPointのデータをWeb形式で保存した場合、1枚1枚のスライドがGIF形式の画像として保存されるとともに、1枚1枚のスライドの文字情報を含んだファイルが生成されるため、このファイルから自動的に1枚1枚のスライドの文字情報の抽出を行うとともに、その1行目をそのスライドのタイトルであるとして、RSSファイルに変換している。

<channel>タグは、1つの内容のまとまりを表すタグであり、<channel>のタイトルを表す<title>および概要を表す<description>には、人手により「moviedb by YGN」を記述した。<link>には、このRSSに関連するWebページとして、講義映像をリスト表示するWebページへのリンクを人手により記述している。<channel>タグの中の<item>タグの中には、1枚1枚のスライドごとに切り出した映像の内容を記述する。例えば、1つ目の<item>タグには、最初のスライドのタイトル(<title>)、および、そのスライドに含まれるテキスト(<description>)を記述している。この<item>タグ中の<title>と<description>については、1枚1枚のスライドの文字情報から自動的に抽出している。<pubDate>は、RSSの内容の更新日時であり、RSSの作成時の時間を自動で記述している。<link>には、このスライドに対応する映像ファイルへのリンクを記述する。映像は、スライド名がmanual.pptの場合、1枚目のスライドに対応する映像部分をmanual0.mov、2枚目のスライドに対応する映像部分をmanual1.movのように名前をつけて切り出しを行っている。<link>の部分で、映像を載せるWebサーバのURL部分については、人手にて入力している。

講義映像の切り出しについては、3章以降で述べる講義映像とスライドの同期結果により、自動的に切り出しを行った。具体的には、講義映像とスライドの同期結果により、1枚1枚のスライドに対応する映像部分が、もとの講義映像の何秒から始まり、何秒で終わるかが分かるため、その部分のみを切り出した映像をプログラムにより自動で作成している。(ただし、この切り出された映像は、映像ストリーミングに対応していないため、別

途、市販の映像編集ソフト等を使って、人手にてストリーミング形式へとフォーマットを変換する必要がある。)

Safari 2で図2のRSSを読み込んだ画面が図3である。ページの一番上には、<channel>タグの中の<title>が表示される。その下に、<item>タグの<title>および、<description>が表示されている。それぞれのタイトルをクリックすることで、対応する映像ファイルをダウンロードして表示することができる。また、画面右上の検索窓にキーワードを入力することで、<title>タグおよび<description>タグの内容を検索し、必要な映像を探し出すことができる。

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
<channel>
<title>moviedb by YGN</title>
<link>http://localhost/moviedb/</link>
<description>moviedb by YGN</description>
<item>
<pubDate>Mon, 21 Jan 2008 15:59:49 +0900</pubDate>
<title>映像/画像コンテンツの検索</title>
<description>
映像/画像コンテンツの検索
柳沼良知
独立行政法人メディア教育開発センター
</description>
<link>http://localhost/moviedb/manual0.mov</link>
</item>
...
</channel>
</rss>
```

図2 RSSファイルの例



図3 RSSファイルの表示

3. 講義映像とスライドの同期

3.1 スライド部分の抽出

講義映像中のスライド部分は、ディスプレイの縦横比

やカメラ位置等により、形が歪んでいる。例えば、図1の講義映像では、モニタがハイビジョン対応であり、スライドが横長に表示されている。また、カメラの位置により、スライドの下側が若干広く撮影されている。

このスライド部分の抽出を行うため、マウスでスライド部分の大きめの位置を指定し、その4隅の位置を少しずつ変えながら、スライド部分と、対応するスライド画像の差分が最も小さくなる位置を探し出すことで、講義映像中のスライド部分の抽出を行った。なお、以下の説明では、「スライド画像」をスライドの1枚1枚を画像として保存したもの、「フレーム画像」を講義映像から一定間隔で抜き出した画像、「スライド部分」をフレーム画像中のスライドに対応する部分と定義する。また、講義映像中のスライド部分の抽出では、講義映像の最初のフレーム画像には、最初のスライド、最後のフレーム画像には、最後のスライドが映っていることを仮定している。

具体的には、以下の手順でスライド部分の抽出を行った。

- 1) マウスをドラッグし、スライド部分の左上点、右下点を指定することで、講義映像中のスライド部分の大きめの位置を指定する。
- 2) (最初のフレーム画像のスライド部分と最初のスライド画像の差分+最後のフレーム画像のスライド部分と最後のスライド画像の差分)を、図4のように、左上点のx座標、y座標を±1の範囲で変えた9点について計算し、値が最も小さくなる点へ、左上点を移動する。なお、差分の計算方法については、3.2で述べる。
- 3) 上記の処理を、左上点→右上点→右下点→左下点→左上点の順に繰り返す。
- 4) スライド部分の4隅の位置が変わらなくなったら終了。

以上の処理により、マウスで大きめに指定されたスライド部分に比べて、より正確なスライド部分の抽出を行うことができる。

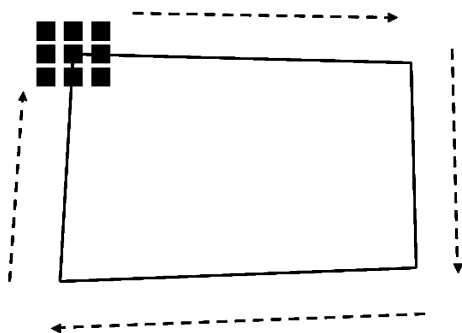


図4 スライド部分の抽出

3.2 差分の計算

3.2.1 差分の計算手順

3.1の処理に必要な差分の計算は、以下の手順で行なう。

- 1) まず、スライド部分の歪み補正を行なう。補正を行ったスライド部分とスライド画像は、後の処理のため、一定の大きさに規格化する。
- 2) スライド部分およびスライド画像について、白黒の2色に変換する2値化処理を行なう。2値化処理は、明るさ情報のみを利用することから、色味の影響を受けにくくなる。また、2値化した画像に対してぼかし処理を行なうことで、文字の位置ずれによる影響を少なくする。
- 3) 最後に、スライド部分とスライド画像の差分画像を作成し、この差分画像の明るさの平均値をスライド部分とスライド画像の差分とする。

以上の処理の詳細については、3.2.2から3.2.4で述べる。

3.2.2 歪み補正

歪みの補正手法を示したのが図5である。図5(a)が変換前のスライド部分とする。左上点を原点Oとし、右上、右下、左下へのベクトルをそれぞれ、 X_1 , X_2 , X_3 とする。スライド部分の4隅の座標を (x_0, y_0) , (x_1, y_1) , (x_2, y_2) , (x_3, y_3) とすると、 $X_1 = (x_1, y_1) - (x_0, y_0)$, $X_2 = (x_2, y_2) - (x_0, y_0)$, $X_3 = (x_3, y_3) - (x_0, y_0)$ である。以下では、スライドの右上部分を例に変換手法の説明を行なうが、スライドの左下部分についても、同様の処理により歪み補正を行なうことができる。

図5(b)が変換後の画像であり、横幅がW、縦幅がHの長方形とする。変換後の点の座標を (m, n) とした場合、この点に対応する変換前の画像中の点の座標 (m', n') が分かれば、その点の色情報を変換後の点にコピーすることで、歪みを補正した画像を作成することができる。

変換後の画像の原点から $(W, 0)$, および、 (W, H) に向かうベクトルを、それぞれ $p: (1-p)$ に内分する点を結んだ直線上に (m, n) があるとすると、また、 (m, n) は、この直線を、 $q: (1-q)$ に内分するとする。この時、

$$p = m/W$$

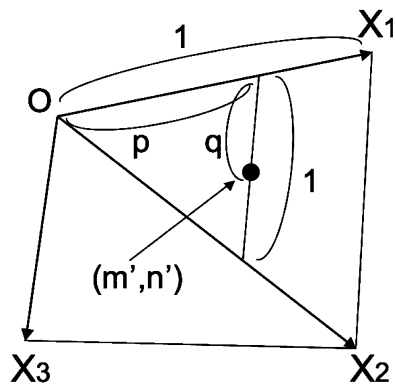
$$q = n/(p \cdot H) = (n \cdot W)/(m \cdot H)$$

である。

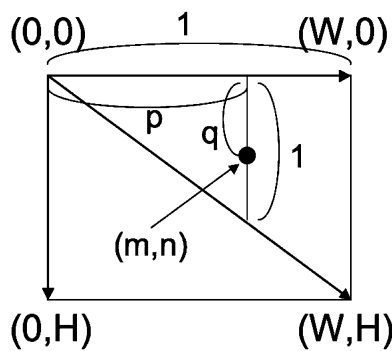
ベクトル X_1 , および、 X_2 を、それぞれ $p: (1-p)$ に内分する点は、 $p \cdot X_1$, $p \cdot X_2$ であり、この座標を $q: (1-q)$ に内分する点が (m', n') となる。よって、

$$(m', n') = (1-q) \cdot (p \cdot X_1) + q \cdot (p \cdot X_2)$$

となる。ここに、前述の p , q の値を代入することで、変換後の点 (m, n) に対応する変換前の点 (m', n') が求まり、この点の色を変換後の画像にコピーすることで、歪みを補正した画像を作成することができる。



(a) 変換前



(b) 変換後

図5 歪み補正

3.2.3 2値化/ぼかし処理

歪み補正をしたスライド部分およびスライド画像に対して、画像の2値化処理を行なった。これは、講義映像を撮影する際のディスプレイやカメラの特性による色味や明るさの変化の影響を受けにくくするためである。また、文字の位置ずれによる影響を少なくするために、2値化した画像に対してぼかし処理を行った。

図6にスライド部分に対して2値化、ぼかし処理を行なった結果の例を示す。スライド部分の歪みが補正され、文字領域、および、色が濃く特徴的な部分が黒く表示さ



図6 2値化/ぼかし処理結果

れていることが分かる。スライド画像についても同様に、2値化/ぼかし処理を行う。なお、暗い背景に明るい文字が表示されている場合については、スライド部分およびスライド画像の白黒を反転させてから2値化、ぼかし処理を行う。暗い背景に明るい文字か否かについては、講義映像とスライドの同期ツールで、「白地に黒」または「黒地に白」のラジオボタンをマウスでクリックすることで指定している。

3.2.4 差分の計算

最後に、スライド部分とスライド画像の差分画像を作成して、スライド部分とスライド画像の差分の大きさを計算する。差分画像 $C(i, j)$ は、2枚の画像の (i, j) での明るさを $A(i, j)$ 、 $B(i, j)$ として、

$$C(i, j) = |A(i, j) - B(i, j)|$$

として明るさを求めた画像である。差分画像では、2枚の画像の明るさの差が大きい部分が明るく、明るさの差が小さい部分が暗く表示される。この差分画像の明るさの平均値をスライド部分とスライド画像の差分とした。

3.3 講義映像とスライドの同期

3.1の処理により、マウスで大まかに指定されたスライド部分に比べて、より正確なスライド部分の抽出を行うことができた。この結果を使い、すべてのフレーム画像およびスライド画像に対して、以下の処理を行うことで講義映像とスライドの同期を行った。

- 1) 3.2.2の処理により、スライド部分の歪み補正を行なう。補正を行ったスライド部分とスライド画像は、後の処理のため、一定の大きさに規格化する。
- 2) 3.2.3の処理により、スライド部分およびスライド画像について、白黒の2色に変換する2値化処理を行なう。2値化処理は、明るさ情報のみを利用することから、色味の影響を受けにくくなる。また、2値化した画像に対してぼかし処理を行なうことで、文字の位置ずれによる影響が少なくなる。
- 3) 3.2.4の処理により、スライド部分とスライド画像の差分画像を作成し、スライド部分とスライド画像が対応するかの判定を行う。差分画像は、2枚の画像の明るさの差が小さい部分が暗く表示される画像であり、作成した差分画像が最も暗くなるスライド画像を、そのスライド部分に対応するスライド画像とした。

4. 評価

4.1 対象としたデータ

講義映像とスライドの同期精度を評価するための対象としたデータは、表1に示すような5つのスライドであ

る。それぞれ、15分から25分の発表で用いられたスライドであり、それぞれのスライドの枚数は、16-25枚となっている。配色については、スライドA-Dは、白黒画像に変換した場合に、白地に黒で文字が表示され、スライドEは、黒地に白で文字が表示される。

スライドAについては、スライドの表示を実際に手動で行ないながら映像の撮影を行なった。撮影した映像から1秒ごとに書き出したフレーム画像の枚数は528枚であった。スライドB-Eは、スライドAと同様の配置で、1枚1枚のスライド画像を自動で10秒ごとに表示し、映像の撮影を行なった。このため、フレーム数は、(スライド数×10)となっている。

表1 対象としたデータ

	スライド数	配色	フレーム数
スライドA	21	白地に黒	528
スライドB	25	白地に黒	250
スライドC	16	白地に黒	160
スライドD	16	白地に黒	160
スライドE	24	黒地に白	240

4.2 同期精度

講義映像とスライドの同期精度は表2のようになった。同期精度は、フレーム数を、講義映像から1秒ごとに抜き出したフレームの数、正解数を、スライドとの対応付けが正しく行なわれたフレーム数として、(正解数/フレーム数)で計算している。

マウスでスライド部分の左上と右下の2点を指定して、講義映像とスライドとの同期を行った場合の精度は、平均で90.7%であった。一方、3.1の処理により、スライド部分をより正確に抽出した後に同期処理を行った場合の精度は平均で99.7%となり、本手法によって高い精度で講義映像とスライドの対応付けを行うことができた。正解率が高いのは、限られた数のスライドの中から正解を選び出せば良いためと考えられる。

表2 同期精度

	2点指定 (%)	位置推定後 (%)
スライドA	100	100
スライドB	100	100
スライドC	98.8	98.8
スライドD	95.6	100
スライドE	59.2	99.6
平均	90.7	99.7

4.3 処理時間

処理時間については、CPU 2.5GHz Dual, メモリ 1.5GB の環境で測定を行った。表3に示すように、フレーム画像とスライド画像の対応付けは、いずれも40秒以下で処理を行うことができた。また、3.1で述べたスライド部分の抽出については、最も多い場合で19頂点の処理でスライド部分の抽出を行うことができた。処理時間は、いずれも6秒以下であった。

表3 処理時間

	対応付け処理 (秒)	スライド部分の抽出 (秒)
スライドA	40	6
スライドB	22	5
スライドC	11	6
スライドD	11	3
スライドE	21	5

5. まとめ

本稿では、画像検索技術を用いて講義映像とスライドの同期を行うことで、講義映像へのメタデータ付与の自動化を実現した。また、評価実験を行い、講義映像とスライドの同期を99.7%の精度で行えることを明らかにした。

今後は、講義映像データベースの画像特徴による検索機能の開発等について、更に検討を行っていく予定である。

引用文献

- インターネット教育支援サービス TIES (2008).
<http://www.tiesnet.jp/>
- Lienhart, R., Effelsberg, W., and Jain, R. (1998). Towards a visual grep: A systematic analysis of various methods to compare video sequences, Proc. of SPIE, **3312**, pp.271-282
- 森本容介・室田真男・清水康敬 (2005). 教育用動画像検索システムと時間情報同期方法の開発, 電子情報通信学会論文誌, **J88-D-I(10)**, pp.1515-1524
- 長坂晃朗・宮武孝文・上田博唯 (1996). カットの時系列コーディングに基づく映像シーンの実時間識別法, 電子情報通信学会論文誌, **J79-D-II(4)**, pp.531-537
- 日本オープンコースウェア・コンソーシアム JOCW (2008).
<http://www.jocw.jp/>
- 杉山善明・有木康雄 (1998). 多重部分空間法に基づくテレビニュース映像の自動分類, 電子情報通信学会論文誌, **J81-D-II(9)**, pp.2112-2119
- 谷口行信・南 憲一・佐藤 隆・桑野秀豪・児島治彦・外村佳伸 (2001). SceneCabinet: 映像解析技術を統合した映

像インデクシングシステム, 電子情報通信学会論文誌,
J84-D-II(6), pp.1112-1121

Wactlar, H.D., Kanade, T., Smith, M.A., and Stevens, S.M.
(1996). Intelligent access to digital video: The informedia
project, *IEEE Computer*, **29-5**, pp.46-52

柳沼良知・鈴木一史・清水康敬 (2005). 教育用映像コンテ
ンツのデータベース化のためのテロップ認識手法の検討,
電子情報通信学会技術研究報告, **105-374**, pp.13-18

柳沼良知・坂内正夫 (1996). DPマッチングを用いたドラマ
映像・音声・シナリオ文書の対応付け手法の一提案, 電
子情報通信学会論文誌, **J79-D-II(5)**, pp.747-755



やぎぬま よしとも
柳沼 良知

昭63東大・工・計数卒。平2同大大学院修士課程(理・物理)了。同年東大生産技術研究所技官。平5同助手。平12メディア教育開発センター助教授。平13より総合研究大学院大学助教授併任。工博。マルチメディア処理, マルチメディアデータベースなどの研究に従事。

Developing of Lecture Movie Database by Synchronizing Lecture Movies and Slides

Yoshitomo Yaginuma

By the improvement of internet and computer environments, many universities have started Web delivery of lecture movies. To retrieve such movies, metadata are indispensable. But, manual extraction of metadata is difficult, if the number of movies is large. Therefore, automatic metadata extraction method of lecture movies is proposed in this paper. In the proposed method, at first, synchronization between lecture movies and slides is carried out. Then, retrieval of lecture movies is realized using text information described in slides. Experimental results revealed that the accuracy of the synchronization between lecture movies and slides using the proposed method is 99.7%.

Keywords

Slide, Lecture Movie, Database, Image Retrieval, Synchronization