

講義映像のデータベース化と検索手法の動向

柳沼 良知¹⁾

インターネットの高速化やコンピュータ性能の向上により、大学等の講義映像をWeb上で公開する試みが広く行われるようになってきた。一方で、写真、図面、地図といった画像データが広く利用されるようになるとともに、画像検索技術の実用化が進んでいる。このような画像検索は、検索意図にもとづき、「検索したい画像のイメージを明確に表現できる場合」に対応する「こんな検索」と、「検索したい画像のイメージを明確に表現できない場合」に対応する「あんな検索」の2つに大別することができる。このため、本稿では、まず、これまでの画像検索、映像検索技術を「こんな検索」、「あんな検索」に分類し、それぞれの手法の特徴について述べる。次に、講義映像を対象とした検索技術の動向について、スライド検索による講義映像の頭出しシステム等を例にして、「こんな検索」、「あんな検索」の観点から議論する。

キーワード

講義映像、データベース、画像検索、こんな検索、あんな検索

1. はじめに

近年、大学等の講義資料や講義映像をWeb上で公開する試みが広く行われるようになってきた。OpenCourseWareは、MIT（マサチューセッツ工科大学）で行われている講義資料や過去のテスト問題等をインターネット上で無償公開する試みであり、英国の公開大学（The Open University, UK）でも、Web学習システムを使って教材の一部が公開されている。日本国内でも、帝塚山大学などが中心となって設立された大学連携によるeラーニングの教育コミュニティであるTIESや、日本でOpenCourseWareを公開している大学を中心に設立された日本オープンコースウェア・コンソーシアムJOCW等で大学の講義資料や講義映像の配信が行われている。特に、インターネットの高速化やコンピュータ性能の向上により、より多量の、より高画質な講義映像を配信することが容易になってきた。

一方で、出版、デザイン、医療など様々な分野において、写真、図面、地図といった画像データが広く利用されるようになり、これにともない、蓄積された多量の画像データの中から、ユーザが自分の好みや目的に応じて自由に必要な画像データを採り出せるような画像検索技術の実用化が進んでいる。

このような画像検索は、ユーザの検索意図にもとづき、大きく2つに分類することができる。1つは、「検索したい画像のイメージを明確に表現できる場合」であり、

ユーザがこんな画像という検索要求を明示的に与えることができる場合に対応する。もう1つは、「検索したい画像のイメージを明確に表現できない場合」であり、ユーザが検索内容のイメージを抽象的に表現することが難しい場合や、画像を連想的に検索する場合に対応する。前者は、「こんな検索」、後者は、「あんな検索」とも呼ばれ（坂内ほか、1996）、様々な手法が提案されている。

このため、本稿では、まず、これまでの画像検索、映像検索技術を「こんな検索」、「あんな検索」の2つに大別し、それぞれの手法の特徴について述べる。次に、講義映像を対象とした検索技術の動向について、「こんな検索」、「あんな検索」の観点から議論する。

2. 検索手法の動向

画像検索においては、通常、まず、対象画像に対して検索の手がかりとなる検索キーの付与が行われる。検索キーは、一般には、対象画像のもつ特徴量や、画像内容の説明情報等が用いられる。検索キーの作成方法としては、人手により作成する方法、コンピュータによる画像認識技術により自動的に作成する方法等があり、検索の効率や柔軟性を考えてさまざまな手法が提案されている。そして、検索キーを利用して、ユーザは、必要とする画像の絞り込みを行っていく。

このような画像検索は、「検索したい画像のイメージを明確に表現できる場合」と、「検索したい画像のイメージを明確に表現できない場合」の2つに大別することができるが、これら2つの検索では、検索システムにお

¹⁾ 放送大学 ICT活用・遠隔教育センター

いて必要となる機能や役割が異なってくる。

2.1 イメージを明確に表現できる場合：「こんな検索」

「検索したい画像のイメージを明確に表現できる場合」とは、ユーザがこんな画像という検索要求を明示的に与えることができる場合であり、「この商標とデザインが似ている商標を探せ」や「俳優の〇〇が映っている画像を探せ」といった検索要求に対応する。

このような検索の場合、ユーザの検索要求がはっきりしていることから、検索する画像内容を比較的容易に指定することができる。このため、このような検索においては、いかにして有効な検索キーを実現するかが重要になる。ここでは、検索キーを、

- (a) 画像の特徴量に基づく検索キー
 - (b) 画像内のオブジェクトや配置に基づく検索キー
 - (c) 画像と他の関連情報を利用した検索キー
- の3つに分類して説明する。

2.1.1 画像の特徴量に基づく検索キー

「画像の特徴量に基づく検索キー」とは、画像から色、大きさ、形、配置等の特徴量を抽出し、それらの類似度により検索を実現するものである。このような検索としては、例えば、HVC色空間内の色ヒストグラムの類似度によって検索を行った例 (Swain, 1993)、画像のテキスト情報を用いて検索を行った例 (Manjunath et al., 1996)、蝶画像を対象に輪郭情報を用いて検索を行った例 (田辺ほか, 1992)、土器画像を対象とし、形状情報をグラフ構造で表し、それらのマッチングにより検索を行った例 (黒川ほか, 1992)、隣接する画素の色対を特徴量として利用した例 (長坂ほか, 1992) 等がある。

TRADEMARK (加藤ほか, 1989) は、商標画像の検索システムである。TRADEMARKでは、各利用者の感じる主観的な画像間の類似度を反映させる空間である主観特徴空間を構成し、その空間内での距離が例示画に近い画像を類似画として検索する。画像特徴としては、濃淡分布や周波数分布等を用い、利用者の主観的な類似度を抽出するために、サンプル画像をグループ分けし、サンプル画像集合内の二つの画像がどれくらい近いかの類似度を被験者に評価させるという方法を用いている。画像特徴空間から主観特徴空間への写像にはアフィン写像を用いている。

CueVideo (Poncelon et al., 1998) は、画像検索技術や音声認識技術と人手によるキーワード付与とを組み合わせた映像検索システムである。音声認識によるキーワード付与については、コマンドモードの認識率は90%以上、自由文の認識を行うディクテーションモードの認識率は50%としている。

CueVideoの画像検索に利用されているQBIC (Flickner et al., 1995) は、色やテキスト、形状、物体の位置等、

様々な特徴量によって画像の検索を行うシステムである。具体的には、以下のような検索を行うことができる。

(a) 形状による検索

例えば、魚の画像を提示し、その輪郭が類似している画像の検索を行うことができる。このような検索は、対象画像がイラスト画であったり、背景と物体とのコントラストがはっきりしている場合などに有効である。

(b) 画像の色合いを指定することによる検索

例えば、黄色が15%、青が13%の画像といった検索を実現できる。色合いは、画像の印象を決める重要な要素であるが、一般には、高い適合率で検索を行うことは難しい。

(c) 背景と物体の特徴による検索

複数の属性を組み合わせることで、「物体が赤くて丸く、背景は緑色」といった複雑な検索を行なうことができる。QBICにおいては、視覚的に条件の組合せを行なうインタフェースが用意されている。

以上述べたような特徴量にもとづく検索は、検索対象となる画像の特徴が明らかである場合には有効なもの、問題点としては、画像を記述する観点が素朴であり、複雑な内容検索を行うことが難しい点が挙げられる。

2.1.2 画像内のオブジェクトや配置に基づく検索キー

「画像内のオブジェクトや配置に基づく検索キー」とは、例えば、「青空の下に白い砂浜が広がる画像を探せ」といったように、画像内に映るオブジェクトの指定や、それらの位置関係の指定にもとづき検索を行うものである。

このような検索システムとしては、山を対象とし、画像内のオブジェクトの前後関係、位置関係等を用いて画像検索を行なった例 (美濃ほか, 1991) や、ランドサット画像を対象に、道路や川、空港等の主要な目標物を抽出し、それらの位置関係を指定することにより画像検索を行なった例 (Yang et al., 1984) などがある。

配置による検索は、スポーツ画像やスポーツ映像の検索にも利用されている。サッカー映像から、ボールやゴールポスト、選手の平均位置等の抽出を行い、コーナーキックシーンの抽出を行う試み (丸尾ほか, 1999) や、サッカー映像のブレイク区間から、ブレイク時間やブレイク前後のフィールド位置の変化等から、ゴールキックやコーナーキック、スローイン等のイベントを推定するシステム (跡部ほか, 2007) が提案されている。

自動認識における中間的、あるいは、不完全な認識レベルをも許容する画像検索システム (Yamane et al., 1993) では、状態遷移型と呼ばれる認識モデルに従って、画像中のオブジェクトの認識レベルの上昇が明示的に表現される。認識は、最上位の認識状態まで行えることが望ましいが、その段階に至らなくても、そのオブジェク

トが中間状態である旨のラベル付けを行い、この中間的な認識水準をも検索の際に利用する。例えば、「サッカー」というキーワードで検索を行ったものの、「サッカー」というキーワードが抽出されていなかった場合、システムは「サッカー」を構成する、より低位の概念である「ゴールポスト」や「フィールド」といった状態をも使って検索を実行する。

状態遷移型認識モデルと画像の特徴を表すシーン記述を用いた画像検索システム（小野ほか、1996）では、状態遷移型認識モデルによるキーワードを用いた検索だけではなく、シーン記述を用いた画像中のセグメントの位置や色などの特徴量や、セグメント間の位置関係などを用いた検索が可能となっている。これにより、例えば、自動認識によるキーワード「空」とシーン記述によるキーワード「下部に白い領域」とを組み合わせることで、「青空の下に白い砂浜が広がる画像」の検索を実現することができる。

また、ニュース映像を対象として、顔の大きさや位置により、キャスタショット、歩行ショット、集合ショット等に映像を分類し、画像的に類似した映像は類似した内容を含むというニュース特有の性質を利用して、索引付けを行うシステム（井手ほか、1999）も提案されている。

以上のような画像内のオブジェクトやその配置に基づく検索の問題点は、自動認識により検索キーとなるオブジェクトの抽出を行うことが、現状では必ずしも容易ではないことが挙げられる。

2.1.3 画像と他の関連情報を利用した検索キー

「画像と他の関連情報を利用した検索キー」とは、一般には困難である画像の認識、キーワード抽出を補うものとして、他の関連情報を利用しようとするものである。これらの検索の場合、画像自体がもつ情報のみを利用するのではなく、それに関連する書誌情報等のテキスト情報を利用して検索を行う。

Chabot (Ogle et al., 1995) は、画像を管理するリレーショナルデータベースからのキーワードと画像特徴とを用いて画像検索を実現するシステムである。リレーショナルデータベースにおいては、画像のタイトル、作成日時等の情報を管理しており、画像特徴としては、データの投入時に色数を限定し、その色ヒストグラムを属性として利用している。そして検索の際、“SomeColor”と指定することで、その色が画像の一部に含まれること、“MostlyColor”と指定することにより、その色が画像全体に含まれることを表わすことができる。また、このシステムでは、例えば、“Sunset (日の入り)”という概念を、

- (a) “Sunset”というキーワードが付与されているか、
- (b) 赤、黄、紫の色領域が大きな画像

というように、概念をChabotで参照できるリレーショナルデータベースの記述や画像特徴を用いて予め登録し

ておくことにより、概念による検索も実現している。

PICTION (Srihari, 1995) は、新聞記事のデジタル化された画像とその説明文を用いて、画像内の人物の位置を特定し、検索を行うシステムである。

具体的な処理は以下のように行われる。

- (1) 自然言語処理モジュールに画像の説明文を入力し、人数や位置関係の制約を抽出する。
- (2) 画像理解モジュールに画像を入力し、人数等の情報も利用し、顔の位置や特徴の抽出を行う。
- (3) 制御モジュールにおいて、説明文から得られた制約情報と画像から得られた情報とを統合し、対応する顔と名前の絞り込みを行う。

説明文からの制約条件の抽出の際には知識ベースが必要となるが、このシステムでは、語彙や写真に固有な知識等、いくつかの知識ベースを組み合わせて利用することで、文章情報と画像情報の統合を実現している。例えば、画像に付随する文章から画像中に人物の肖像画が存在することが分かった場合、肖像画部分は強いエッジで囲まれているという知識を用いることで、選択的に肖像画に描かれた人物の抽出を実現している。

映像の検索キーとして、画面上に表示される字幕を利用しようとする試みも提案されている（佐藤ほか、1998、茂木ほか、1996）。これらは、ニュース映像等の字幕領域を自動的に検出し、文字認識を行うことで、その結果を検索キーとして利用する。

複数メディアの統合によるドラマ映像のデータベース化システム（柳沼ほか、1996）は、ドラマ映像に付随する音声やシナリオ文書を関連メディアとして利用する。しかし、映像、音声、シナリオ文書といったメディアは、あらかじめ時間的に同期されているわけではない。このため、各々のメディアから複数のメディアから参照できる特徴量パターンを抽出を行い、そのパターンをDPマッチングを用いて対応付けすることにより複数メディアの同期を実現している。

具体的な処理は以下の手順で行われる。

- (1) 各々のメディアの処理、解析を行ない、その結果をメディアに依存しないパターンとして抽出する。例えば、「台詞のあるなしのパターン」は、音声から音量の大小を用いることにより抽出可能であるばかりではなく、シナリオ文書からも台詞の文字数を用いて推定することが可能である。
- (2) 抽出したパターンをDPマッチングにより最適に対応付けることで、メディア間に対応付けを行なう。すなわち、抽出したパターンを時間軸方向に非線形に伸縮し、パターン間の最適な対応付けを求めることにより、あるメディアの時間軸方向の任意の部分が、他のメディアのどの部分に対応するかを求める。

図1に、ドラマ映像を対象とした映像、音声、シナリオ文書の対応付け結果を示す。

1番上のパターンがシナリオ文書から抽出した台詞パターンであり、上から5番目のパターンが、音声から抽出した台詞パターンである。上から2番目のパターンがシナリオ文書から抽出した女性の存在パターンであり、上から6番目のパターンが、音声から抽出した女性の存在パターンである。上から3番目のパターンがシナリオ文書から抽出した場面の変わり目パターンであり、上から7番目のパターンが、映像から抽出した場面の変わり目パターンである。そして上から4番目の枠には、映像、音声、シナリオ文書の対応付け結果を示している。このメディア間の対応付け結果を用いることで、シナリオ文書中に記述されている人物名や場所といった情報の利用が可能となり、人物名による映像の検索等、映像の高度な利用を実現できる。

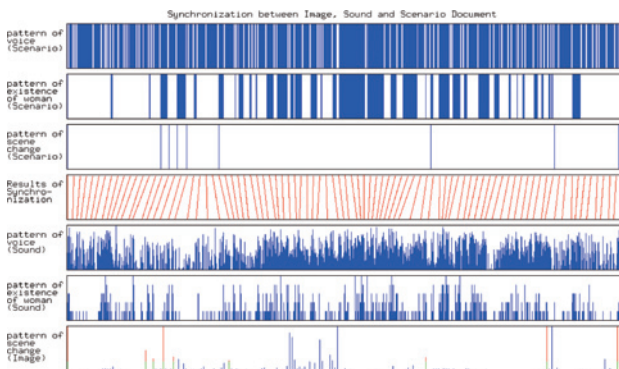


図1 映像、音声、シナリオ文書の対応付け結果

2.2 イメージを明確に表現できない場合：「あんな検索」

「検索したい画像のイメージを明確に表現できない場合」とは、ユーザが検索内容のイメージを抽象的に表現することが難しい場合や、画像を連想的に検索する場合に対応する。このような検索では、ユーザの検索意図をどのように解釈し、くみとるかが重要となる。

このような検索システムとしては、ユーザが検索したい画像の例を指定することにより、ユーザがどの特徴量に注目しているかを推定し検索を行うシステム (Picard et al., 1995)、顔画像を対象とし、指示された画像と目や鼻の形や位置が異なる画像を生成し、ユーザの検索空間内の動きによりユーザの検索意図を抽出しようとする試み (尾田, 1994) 等がある。

報道機関の保有している静止画を対象とした、連想的な検索が可能キーワード検索システムとしては FORKS (浦谷ほか, 1987) がある。このシステムは、あらかじめシステムに登録されたキーワードを利用する統制キーワード方式を採用し、キーワードの上位語・下位語の関係を体系化し木構造群で管理している。これらのキーワードは、画像登録の際に人が付与する。

FORKSでは、このような画像に対するキーワードの重みのベクトルである画像ベクトルを作成し、画像ベク

トル間の角度の大小を反映した画像間距離を定義している。連想検索は、検索キーとなる画像とデータベース中の全画像との画像間距離を計算し、距離の最も小さい画像から順に、指定された件数の画像を選択し表示することにより行なわれる。

百科事典の知識をもとに連想的に検索を行うシステム (芥子ほか, 1996) も提案されている。このシステムでは、10万語以上の意味ベクトルの辞書を百科事典のテキストから生成し、3万6千件の写真を対象とした検索を行っている。

意味ベクトルの作成は、頻度解析により用語の説明文中に良く現れる3700語を選択し、まず、それらに266次元の意味ベクトルを人手により付与した。次に、3700語の加重平均により10万語に対して意味ベクトルを付与するという手順で行われた。

例えば、「リゾートの家」というキーワードで検索を行った場合、第6位に検索された画像には、「リゾート」という記述は付与されていなかったものの、この画像に記述されている「マリナデルレイ」と「リゾート」の意味ベクトルが類似していることから検索結果として表示された。このように、このシステムでは、百科事典の知識を用いることで、ユーザの曖昧な画像指定を可能としている。

3. 講義映像のデータベース化と検索

3.1 講義映像のデータベース化

講義映像のデータベース化やWeb配信を行う場合、関連する資料、スライド等を講義映像と同時に配信すれば、画面サイズの小さい映像中に表示するのが困難な、関連する詳細な情報の参照が容易になる。また、資料が映像と別にある場合に必要な、映像に合わせて資料のページをめくったり、映像と資料の交互に視線を移すといった手間が不要になるなど、学習者の利便性は向上する。

このような講義映像の作成に関しては、講義映像と書画カメラの映像を合成して作成する方法 (渡邊ほか, 2001)、PC画面を撮影した映像と講義者の音声を合成して作成する方法 (山本, 2004)、講義映像と同期してスライド等の補助資料を表示する方法 (阿倍ほか, 2006) 等が提案されている。また、無償でダウンロードして使用することができるMicrosoft Producerでも映像とスライドの同期教材の作成を行うことができる。

以下では、このような講義映像を対象としたメタデータ付与手法や、検索技術について、「こんな検索」、「あんな検索」の観点から述べる。

3.2 講義映像の「こんな検索」

教育コンテンツを対象とした「こんな検索」に対応する画像検索システムとしては、例えば、人手で付与したメタデータとともに、画像特徴量として、画像や動画ク

リップのHSI色空間内での距離を用いた類似画像検索システム（福本ほか，2004）が提案されている。

また，映像データベースでは，同様の映像が複数投稿されている場合，映像の取り込み方法の違いなどによって，横長になっていたり，映像が黒い枠に囲まれている場合があるなど，単純な類似検索手法では検索を正しく行えない場合がある。このため，画像の縦横比の変更や黒枠が付加されることを考慮した類似検索手法（柳沼，2008）も提案されている。この手法では，黒枠がある可能性のある範囲を設定し，もしそこに含まれる画素が黒である場合に，その画素を用いないで特徴量となる色相のヒストグラムの作成を行う。これにより，黒枠がある場合でも，ない場合でも，色相のヒストグラムがほぼ同じになり，黒枠の有無にかかわらず類似した映像を検索できるようになる。

画像間の類似度の尺度としては，2つの画像のヒストグラムを $h1(n)$ ， $h2(n)$ ，ヒストグラムの段数を N とした場合，

$$\sum_{n=1}^N \frac{(h2(n) - h1(n))^2}{h1(n)}$$

を計算し，この値が小さいほど画像は類似しているとしている。この式は映像からの場面の変わり目を検出するシーンチェンジ検出等で用いられる式である。

図2は，色特徴による検索画面である。左上が検索キーとなる画像であり，その画像との類似度順に検索結果の画像が並べて表示されている。そして，これらの画像をクリックすることで，クリックした画像を検索キーとした類似検索結果を表示することができる。また，それぞれの画像の下の「Open HTML」のアイコンをクリックすることで，対応する映像を表示することができる。

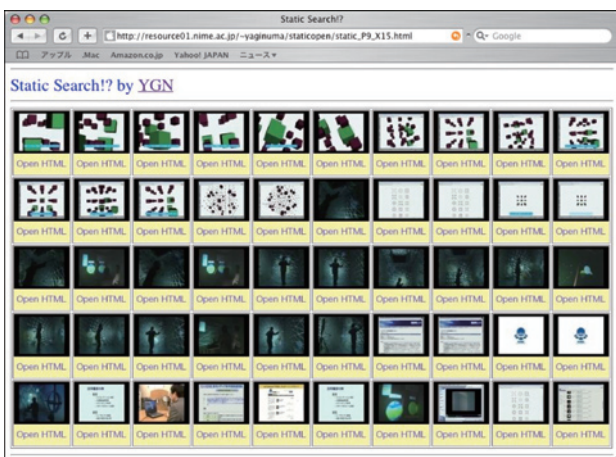


図2 色特徴による検索

講義映像と他の関連情報を利用して検索を行う場合，映像全体に対して，映像の概要のような1つのメタデータだけが付与されているのであれば，映像中の必要な部分だけを検索し，頭出しして部分的に再生することが

できない。このため，学習者が自分の興味のある部分のみを検索して視聴したり，教師が映像の一部分のみを検索して授業の中で使うといった利用を実現するためには，メタデータを映像中の時間情報とともに付与する必要がある。

このようなメタデータの付与方式としては，講義映像の音声認識結果と書き起こしテキストとを時間的に同期することによりメタデータ付与を実現する方法（森本ほか，2005）が提案されている。また，スライドを使った講義映像を対象として，講義映像中から読み取った文字と，スライド中の文字との位置関係の類似度により対応付けを行うことで，講義映像とスライドの同期をとる手法（小澤ほか，2002）が提案されている。

講義映像とスライドとの同期を，文字が箇条書きされていたり，図が等間隔で配置されているといった空間的な周波数特徴の類似度を用いて行い，講義映像の検索を実現する手法（柳沼ほか，2010）も提案されている。この手法では，まず，講義映像から一定間隔で抽出したフレーム画像中のスライド部分と，もとのスライド画像に対してフーリエ変換を行い，空間的な周波数特徴の抽出を行う。次に，各フレームの空間周波数特徴に最も近いスライド画像を選び出すことで，講義映像とスライド画像との対応付けを行う。これにより，それぞれのスライド画像が，講義映像の何フレーム目から何フレーム目に対応するかが分かり，その結果，それぞれのスライドが講義映像のどの時間に対応しているかを知ることができる。

この結果により，スライドファイル中のテキスト情報を検索用のメタデータとして利用できるようになり，テキスト検索により講義映像を部分的に検索し，必要な部分のみを再生することが可能となる。講義映像の部分検索は，抽出したメタデータを，RSS (Really Simple Syndication) と呼ばれるメタデータ形式へと変換することで実現している。

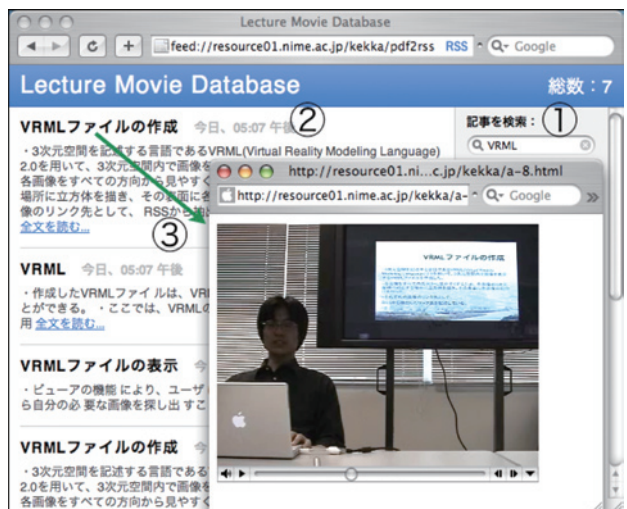


図3 講義映像の部分検索

講義映像の部分検索を行った例を図3に示す。①の部分がWebブラウザによって用意された検索窓であり、ここでは、キーワードとして「VRML」を入力している。②の部分には、検索されたスライド画像のタイトルや内容、更新日時等の情報が表示されている。そして、検索された項目のタイトルをクリックすることで、③のように、対応する映像部分を再生することができる。

3.3 講義映像の「あんな検索」

教育コンテンツを対象とした「あんな検索」に対応する検索システムとして、先修科目、後修科目といったシラバスの前後関係や、習得項目等の共通キーワードによりシラバスの関係を線で結ぶことでシラバスを可視化するシステム（木下ほか，2007）が提案されている。

また、東京大学で開講されている授業科目のシラバスや教材などを公開するためのWebサイトであるUT OpenCourseWareでは、シラバスを横断的に検索し、俯瞰的に可視化できるMIMA Search (Mima, 2006) を実装している。従来の検索システムでは、指定されたキーワードに関連するシラバスをリスト表示するだけなのに対し、MIMA Searchでは、検索されたシラバスの構造を「点」と「線」によるグラフ構造により表現し、シラバス間の意味的な関連性が強ければ近くに配置することで、どの講義が内容的に近いかわ、履修の順序、他大学のシラバスとの関連等を直感的に把握できるようにしている。

また、講義映像を対象としたものとしては、講義映像の要約のため、映像から離散コサイン変換により抽出した45次元の画像特徴と、提示資料のタイトルに含まれる用語の頻度を要素とする57次元ベクトルを用いて自己組織化マップを作成し、類似する映像を近くに配置することで映像の要約を行うシステム（中山ほか，2005）も提案されている。

このような講義映像の「あんな検索」システムとして、現在、スライド検索による講義映像の頭出しシステムの開発を行っている。このシステムは、スライドと講義映像との同期結果をもとに、スライド中のテキスト情報を用いた講義映像検索を実現するとともに、検索結果のスライドの関連性を視覚的に表示することで、必要な検索を行いやすいようにしている。

図4がWebブラウザ上で動作するスライド検索による講義映像の頭出しシステムである。画面上側には、検索対象となる講義映像を選択するメニューや、検索キーワードを入力するためのテキストエリア等があり、「検索」ボタンをクリックすることで、画面下側に検索キーワードを含むスライドが表示される。検索されたスライドをクリックすることで、対応する映像部分の再生が行われる。

一方で、「あんな検索」を実現する場合、検索結果の個々

の関係性の表示方法が重要となる。このような、類似するものを近くに配置するような表示手法としては、自己組織化マップによる方法、項目間が仮想的なバネで繋がれているとして項目間の距離を決める方法、対応分析による方法等がある。ここでは、検索結果のスライドの関連性を視覚化する手法として、対応分析を用いている。

図4の「Visual検索」ボタンをクリックすることで、図5のような、「あんな検索」用のインタフェースが表示される。この図は、対応分析により、1枚1枚のスライドのテキスト情報が類似するスライドを近くに配置している。具体的には、まず、検索対象となる184枚のスライドのテキスト情報を、形態素解析プログラムである「茶筌」を用いて品詞ごとに分け、上位20位のキーワードを抽出した。抽出したキーワードは、「権利、情報、技術、倫理、論文、科学、学術、学会、目的、著作、出版、研究、利用、放送、保護、管理、通信、複製、条、項」である。

ここで、スライドが I 枚、抽出したキーワード数を J として、 i 番目のスライドに対する j 番目のキーワードの個数を f_{ij} とすると、 f_{ij} は、 I 行 J 列のクロス表となる。この時、行のプロファイル、すなわち、行の比率のパターンを考えると、各行の総和が1となる制約から、行のプロファイルは、 $(J-1)$ 次元空間に分布する I 個の点の集合となる。また、列のプロファイル、すなわち、列の比率のパターンを考えると、各列の総和が1となる制約から、列のプロファイルは、 $(I-1)$ 次元空間に分布する J 個の点の集合となる。これらの点を、より次元数の低い同じ空間上に配置して、これらの点の関連性を見るのが対応分析である。これにより、類似するスライドを近くなるように配置できるだけでなく、関連するキーワードがどこに存在するかを重畳表示することができる。

例えば、図5では、大きく分けて、画面左側には、「倫理」について述べたスライドが配置されており、その上側には、「学会」のキーワードがあり、この部分には、主に学会の倫理規定に関するスライドが配置されている。画面左下には、「学術」、「論文」等のキーワードがあり、この部分には、主に学術論文の倫理に関するスライドが配置されている。画面右側には、「著作」、「条」、「項」のキーワードが配置されており、この部分には、著作権の条項に関して述べているスライドが主に配置されている。このように、対応分析を用いることで、内容の類似度に従ってキーワードと関連付けながらスライドを配置できていることが分かる。なお、対応分析結果は、キーワード検索と連動しており、対象となる講義映像をメニューで選択したり、キーワードを指定することで、対象となるスライドを絞り込むことができる。

対応分析結果を表示するウインドウの左上の、「+」、「-」ボタンは、画面の拡大縮小を行うボタンである。また、矢印ボタンは、表示されている画面を上下左右に

移動させるボタンである。これらのボタンにより、興味がある部分を拡大し、スライドの内容を確認しながら、必要なスライドを探すことができる。図6は、キーワードとして「学会」を含むスライドのみを検索し、その一部を拡大表示したものである。これらのスライドをクリックすることで、画面右側のムービー再生ウインドウに示すように、スライドに対応する講義映像を表示することができる。



図4 スライド検索による講義映像の頭出しシステム

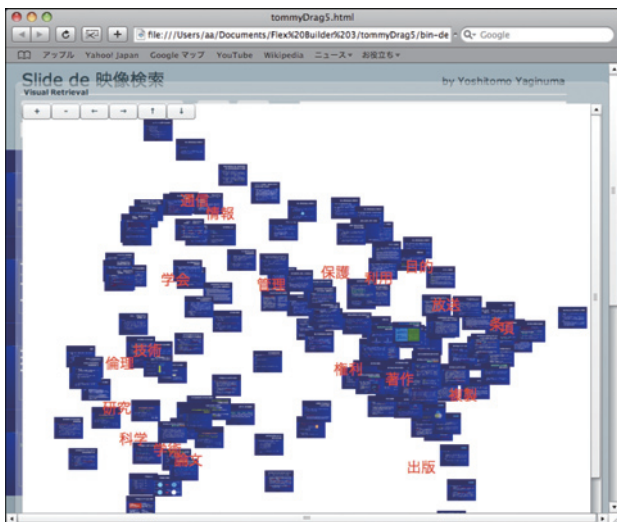


図5 スライドの「あんな検索」インタフェース

4. まとめと考察

本稿では、これまでの画像検索、映像検索技術を、ユーザの検索意図にもとづき、「こんな検索」、「あんな検索」に分類し、それぞれの手法の特徴について述べるとともに、講義映像を対象とした検索技術の動向について、「こんな検索」、「あんな検索」の観点から述べた。

このように講義映像の検索を、これまで行なわれてき

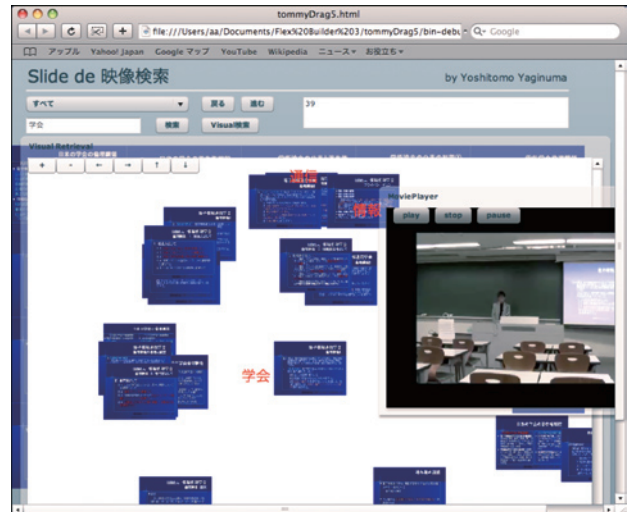


図6 スライドに対応する映像の表示

た検索の流れの中で見た場合、「こんな検索」に含まれる「画像内のオブジェクトや配置に基づく検索」は、これまで講義映像を対象としては、あまり積極的に行なわれてこなかった。これは、例えば教室の講義風景を撮影した場合、大きな画面の変化がなく、配置による検索のニーズがあまり無かったことが原因の1つと考えられる。しかしながら、放送大学の講義映像のように、放送番組と同様に撮影され、編集された講義映像については、「A先生が1人だけ映ってBについて説明をしている場面を探したい」、「Cについて説明しているスライドが映った場面を探したい」など、キーワードによる検索とともに、配置等による検索も併用できれば、利用者の利便性がより向上すると考えられる。

また、講義映像を頭出しして検索する場合、1つの映像から頭出し用の複数の画像を抽出して検索することになるため、シラバス検索等に比べて、検索対象となる画像の枚数が大きくなるという問題がある。このため、「あんな検索」で表示する際の粒度（1つの映像から何枚の画像を表示するか）は、最初は少ない枚数の特徴的な画像のみを示し、拡大した場合により多くの画像を表示するような適応的な表示方法や、注目点である画面の中心に近い部分では画像を大きく表示し、周辺部にいくほど画像を小さく表示するといった表示方法についても検討が必要になる。また、最近では、タッチパネル式の携帯端末等も普及しはじめていることから、例えば、ダブルクリックで画面が拡大、ドラッグで移動するというような、より直感的に検索操作を行いやすいユーザインタフェースについても検討する必要がある。

また、大きな流れとして、検索対象を明示的に指定する「こんな検索」から、「こんな検索」と「あんな検索」を併用したシステムへと検索システムが移っていく傾向もある。これは、「こんな検索」では、記述の曖昧性により検索できない場合があったり、指定したキーワード

を含んだ検索結果の順位付けが難しい場合があることなどが要因として挙げられる。これに対し、検索結果を「あんな検索」型の関連性によって表示することにより、異なる分野で同じキーワードが使われていたり、言葉のゆらぎがある場合でも、ある程度対応が可能になる。また、ユーザの好みや検索履歴などにより検索結果の配置を最適に変えることもこれからの研究課題となり、特に教育用の検索システムでは、教育内容によって必要とされるコンテンツが大きく違ったり、初等中等教育、高等教育、生涯教育等の幅広い利用者が想定されるため、技術的な開発とともに、その利用者による評価をあわせて行なっていくことの必要性が高いと考えられる。

引用文献

- 阿倍博信, 濱谷英次 (2006). 映像アノテーションモデルに基づく講義映像を用いたマルチメディア教材オーサリングシステム, 教育システム情報学会誌, 23 (3), 141-147
- 跡部裕貴, 泉正夫, 福永邦雄 (2007). サッカーの放送型映像における試合中断区間に注目したイベント推定, 電子情報通信学会技術研究報告, 106 (606), 25-30
- Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by Image and Video Content: The QBIC System, IEEE Computer, 23-32
- 福本徹, 赤堀侃司 (2004). 画像データベースにおけるメタデータ付与システムの構築と評価, 教育システム情報学会誌, 20(4), 364-370
- 井手一郎, 山本晃司, 浜田玲子, 田中英彦 (1999). ショット分類に基づく映像への自動的索引付け手法, 電子情報通信学会論文誌, J82-D-II(10), 1543-1551
- 加藤俊一, 下垣弘行, 藤村是明 (1989). 画像対話型商標・意匠データベースTRADEMARK, 電子情報通信学会論文誌, J72-D-II(4), 535-544
- 芥子育雄, 池内洋, 黒武者健一 (1996). 百科事典の知識に基づく画像の連想検索, 電子情報通信学会論文誌, J79-D-2(4), 484-491
- 木下聡, 刈谷悠, 妻鳥貴彦 (2007). シラバス可視化システムの構築, 電子情報通信学会技術研究報告, 106 (583), 99-104
- 黒川雅人, 洪政国 (1992). 形状情報を用いた画像の類似検索システム, 情報処理学会論文誌, 32(6), 721-730
- Manjunath, B. S. and Ma, W. Y. (1996). Texture features for browsing and retrieval of image data, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), 837-842
- 丸尾二郎, 岩井儀雄, 谷内田正彦, 越後富夫, 飯作俊一 (1999). サッカー映像からの特定映像イベントの抽出, 電子情報通信学会技術研究報告, 99 (183), 31-38
- Mima, H. (2006). MIMA search: a structuring knowledge system towards innovation for engineering education, Proceedings of the COLING/ACL on Interactive presentation sessions, 21-24
- 美濃導彦, 岡崎洋, 坂井利之 (1991). 対象物の属性特徴による画像検索法—風景画像中の山を例として—, 情報処理学会論文誌, 32(4), 513-522
- 茂木祐治, 有木康雄 (1996). ニュース映像中の文字認識に基づく記事の索引付け, 電子情報通信学会技術研究報告, 95 (582), 33-40
- 森本容介, 室田真男, 清水康敬 (2005). 教育用動画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), 1515-1524
- 長坂晃朗, 上田博唯, 宮武孝文 (1992). 部分領域の色組み合わせに着目した映像中の特定対象物検索手法の提案, テレビジョン学会技術報告, 16 (64), 1-8
- 中山実, 米川孝宏, 清水康敬 (2005). 自己組織化マップを用いた講義映像要約の検討, 電子情報通信学会技術研究報告, 104 (643), 7-12
- 尾田政臣 (1994). 人間のイメージ形成過程の特性を利用した画像検索システム, 情報処理学会論文誌, 35 (7), 1449-1456
- Ogle, V. E. and Stonebraker, M. (1995). Chabot: Retrieval from a Relational Database of Images, IEEE Computer, 28(9), 40-48
- 小野敦史, 天野督士, 斗谷充宏, 佐藤 隆, 坂内正夫 (1996). 状態遷移モデルとシーン記述言語による自動キーワード付与機能を持つ画像データベースとその評価, 電子情報通信学会論文誌, J79-D-II(4), 476-483
- 小澤憲秋, 武部浩明, 勝山裕, 直井聡, 横田治夫 (2002). 文字認識を利用した講義動画中のスライド同定, FIT (情報科学技術フォーラム) 2002, 133-134
- Picard, R. W. and Minka, T. P. (1995). Vision Texture for Annotation, Multimedia Systems, 3(1), 3-14
- Ponceleon, D., Srinivasan, S., Amir, A., Petkovic, D., Diklic, D. (1998). Key to Effective Video Retrieval: Effective Cataloging and Browsing, ACM Multimedia '98, 99-107
- 坂内正夫, 柳沼良知 (1996). 画像の内容検索技術, ビット別冊マルチメディアコンピューティング, 共立出版, 111-119
- 佐藤隆, 新倉康臣, 谷口行信, 阿久津明人, 外村佳伸, 浜田洋 (1998). MPEG符号化映像からの高速テロップ領域検出法, 電子情報通信学会論文誌, J81-D-2(8), 1847-1855
- Srihari, R. K. (1995). Automatic indexing and content-based retrieval of captioned images, IEEE Computer, 28(9), 49-56
- Swain, M. J. (1993). Interactive indexing into image databases, Proceedings of SPIE, 1908, 95
- 田辺勝義, 大谷淳, 石井健一郎 (1992). 多次元心理空間を用いる類似画像検索法, 電子情報通信学会論文誌, J75-D-II(11), 1856-1865

- 浦谷則好, 柴田正啓, 野口英男, 相沢輝昭 (1987). 静止画検索システムFORKSの試作, 情報処理学会論文誌, 28(7), 758-767
- 渡邊晶, 矢吹道郎, 藤村直美 (2001). 簡便な方法で作成可能なオンラインビデオによる復習・自習支援システム, 教育システム情報学会誌, 18(2), 179-188
- 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), 747-755
- 柳沼良知 (2008). 映像データベースとその類似検索手法の開発, メディア教育研究, 5(2), 131-136
- 柳沼良知, 鈴木一史, 児玉晴男 (2010). 講義映像とスライドの空間的な周波数特徴を用いた同期による検索手法, 教育システム情報学会誌, 27(1), 118-127
- 山本芳人 (2004). ビデオカメラを使わないVOD講義コンテンツの作成とその教育利用, 教育システム情報学会誌, 21(2), 117-121

- Yamane, J., and Sakauchi, M. (1993). A Construction of a New Image Database System which Realizes Fully Automated Image Keyword Extraction, IEICE Transactions on Information and Systems, E76-D(10), 1216-1223
- Yang, Y. and Fu, K. S. (1984). A low cost geometry-preserving image database system, International Conference on Data Engineering, 604-610



やぎぬま よしとむ
柳沼 良知

1988年東大・工・計数卒。1990年同大大学院修士課程(理・物理)修了。同年東大生産技術研究所技官。1993年同助手。2000年大学共同利用機関メディア教育開発センター助教授。2001年より総合研究大学院大学助教授併任。2004年独立行政法人メディア教育開発センター助教授。2009年放送大学ICT活用・遠隔教育センター准教授。工博。マルチメディア情報処理, マルチメディアデータベースなどの研究に従事。

Survey on Lecture Movie Databases and Retrieval Techniques

Yoshitomo Yaginuma¹⁾

Nowadays, many universities provide Web delivery services of lecture movies. On the other hand, retrieval techniques of image data, such as photos, drawings, maps etc., are applied to many applications. Such image retrieval techniques are classified into two categories. One is the retrieval whose retrieval keys can be described clearly, such as "Retrieve this". The other is the retrieval whose retrieval keys can't be described clearly, such as "Retrieve that". In this paper, at first, features of above two categories of image retrieval are shown. Then, retrieval techniques for lecture movie databases are discussed.

Keywords

Lecture Movie, Database, Image Retrieval, "Retrieve this", "Retrieve that"

¹⁾ Center of ICT and Distance Education, The Open University of Japan