

特集：ICTの高度化による先進的学習支援に向けて
—技術と人間の共生を目指して—

部分検索を目指した講義映像の場面の自動分類

柳沼 良知*

Automatic Scene Classification for Retrieval of Lecture Movie Segments

Yoshitomo YAGINUMA*

1. はじめに

大学などにおいて、講義資料や講義映像の配信が広く行われるようになってきている。例えば、OpenCourseWareは、MITで始まった大学などで正規に提供された講義とその関連情報（シラバス、講義資料、テストなど）をインターネット上で無償公開する試みである。また、iTunes Uは、PCやiPadなどの端末に対して、映像、音声、電子書籍などのさまざまなコンテンツを無償で配信する仕組みであり、50万以上のコンテンツを利用することができるようになっていく。

このような教育コンテンツを利用する場合に必要な機能の一つとして、検索機能がある。コンテンツの検索のためには、それぞれのコンテンツに対して、検索のためのメタデータが付与されている必要があるが、特に講義映像の検索を考えた場合、映像全体に対して、映像の概要のような一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、映像の部分ごとにメタデータを付与する手法に関して、従来からさまざまな研究が行われている。

例えば、森本ら⁽¹⁾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を提案している。また、講義映像とその講義で利用されているスライドとの同期を行うことで、スライド中のテキスト情報を用いた講義映像の検索を実現するとともに、利用者が検索のためのキーワードを明示的に表現することが難しい場合や、連想的に検索を行う場合に対応するため、スライドの関連性を視覚的に表示することで、必要な映像の部分検索を実現するシステム⁽²⁾も提案されている。

このような講義映像の検索を、意味的なまとまりごとに行おうとする場合、

- 1) 映像を場面ごとに分割する
- 2) それぞれの映像部分からメタデータを抽出する

といったことが必要になる。このなかで、「それぞれの映像部分からメタデータを抽出する」部分に関しては、例えば、音声認識やOCR（光学文字認識）などを利用することも考えられるが、これらについては今後の課題とし、本論文では、「映像を場面ごとに分割する」部分の自動化手法の提案を行う。具体的には、講義映像は、講師画面や資料画面によって構成されることから、まず、講義映像中から顔検出を行い、講師の検出を行うことで、講師画面かそれ以外の資料画面かの分類を行う。そして、その連続性により、講義映像の場面ごとの分割を実現する。また、評価実験により、その性能評価を併せて行う。

*放送大学 ICT 活用・遠隔教育センター (Center of ICT and Distance Education, The Open University of Japan)
受付日：2012年4月27日；再受付日：2012年7月26日；採録日：2012年9月26日