テキスト情報に基づくビデオ映像の構造化ブラウザ

4Z－3

早川 和宏　杉崎 正之　大久保 雅且　田中 一男
NTT ヒューマンインタフェース研究所

1　はじめに

インターネットの普及，映像符号化技術の発達，パーソナルコンピュータのマルチメディア機能の充実により，符号化，配信，復号化を安価に行えるようになってきたことから，個人レベルでも映像の情報発信が可能となりつつある。その結果，今後はビデオ映像が情報洪水をもたらすメディアの中核として登場してくると思われる。

ビデオ映像を視聴する労力の軽減を目標とする技術開発としては，ビデオ映像をショット毎に分割したり，ショットの代表画像を抽出するなどの研究が行なわれている。代表画像を用いれば，利用者は映像コンテンツを概観することができるようになるが，一方すべてのショットから一つずつ代表画像を抽出すると，得られる画像は10分の映像で通常100個前後にものぼり，そこから内容を理解するために多大な労力を要するという別の問題[1]も現れてきた。本稿ではこの問題を解決するための，台本などのテキスト情報によってショットをグルーピングする手法について述べる。

2　マルチメディア速覧

本手法は，テキスト情報の速覧[2]によって，テキスト自体を扱われている話題の変遷に従って階層的に構造化し，次にそのテキストと映像を対応づけることによって映像を意味的に構造化する。この手法を以下マルチメディア速覧と呼ぶことにする。ここで映像とテキストの対応づけには時刻の情報を用いるので，本手法では映像に対応する字幕や台本などの情報が，タイムコード付きで入手できる必要がある。今回は英語ニュース番組の翻訳原稿を入手して実験を行っている。また，現在テレビ放送については2007年をめどに生放送以外の番組についてすべて字幕をつけることが検討されており[3]，いずれほとんどのビデオ映像に字幕がつくと思われる。

マルチメディア速覧の基本的な処理は以下のようになる。まずテキストに対して速覧処理を行う。速覧処理は，テキスト中から「まず」「次に」「ところで」など，話題転換手がかり句（話題の継続・転換を表す語句）を見出して話題の流れを抽出する処理と，一つの話題が継続している中で，「～とは」「～について」などの説明的な語句が付与されているかどうかを手がかりに，話題語（扱われている話題を端的に表す名詞句）を探す処理の二つからなる。話題の階層構造は話題転換手がかり句の属性に基づき決定される。速覧処理の結果として，テキスト全体の目次に相当するような構造を抽出できる。

次に，テキストに付与されているタイムコードを参照し，新たな話題語の出現する文の開始時刻と，その文の直前に隣接する文の終了時刻との時間的な中間点をテキストから見た話題の転換時刻と考える。ただし，これだけでは話題語の出現する文の開始時刻のはるか前に話題の転換時刻が現れることも考えられる。そこで，話題の転換時刻が話題語を含む文の開始時刻に先行できるのは数秒以内（実際には5秒とした）という制約を設けている。

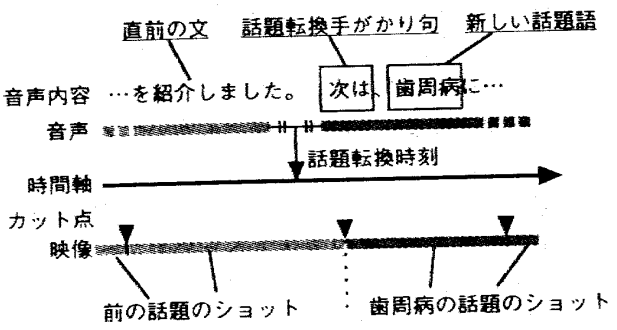

図1: 映像ショットのグルーピング

このようにして得られた話題転換時刻を，ビデオ映像の時間軸にマッピングする。そして，一つの話題転換時刻から次の話題転換時刻までの区間に開始

Video Browser Using Topic-based Structuring of Text Data.
Kazuhiro HAYAKAWA, Masayuki SUGIZAKI, Masaaki OHKUBO and Kazuo TANAKA
NTT Human Interface Laboratories

されるショットを，意味的な一つのまとまりとして扱う（図1）。ショット検出にはPaperVideo[4]を用いた。

また，話題語の出現する文の近辺でショット切換えが起こっている場合，そのショットと話題語とは関係している可能性が高いと考えられる。このことを利用して，話題語と特定のショットとを結び付けることができる。

以上のようにして，

- 話題語の有効範囲に存在する複数のショット
- 話題語と直接関係するただ1つのショット

を抽出することができる。次に，このようにして話題語と関係づけられたショットの代表画像を通じて，ビデオ映像をブラウジングすることを考える。

3　ビデオ映像の構造化ブラウザ

ビデオ映像を階層的にブラウジングする際に，マルチメディア速覧では画像とテキストという二つの異なる情報間の関連付けが得られる。この関連付けを用いて，ショット代表画像，ビデオ映像，台本テキスト，話題語という4つの異なる情報を統合しなければならない。ブラウザとしては，これらをどのように組み合わせて提示するのが望ましいだろうか。

今回は話題語とショット代表画像を主な表示対象として選び，ニュース映像約4時間分について話題を分析した結果，地名・人名等の固有名詞が階層構造の葉に近い方の話題語に多く見られる傾向があった。このことから，話題構造の階層の上位（根）がより粒度の高い，意味的な抽象度の高いものを表し，下位（葉）がより粒度の低い，具象的な情報を表すと考え，前者をテキスト（話題語）で，後者を画像（ショット代表画像）で表すことにした。これは，テキスト情報はより抽象的な内容を表現することができ，画像は具象的な情報の提示に向いていると考えたためである。

図2は試作したビデオ映像ブラウザの全体像である。下半分はビデオ映像を表示するためのスペースであり，上半分は話題構造にしたがってショット代表画像を閲覧するブラウザとなっている。この部分は，Windows95のエクスプローラ風になっており，エクスプローラにおけるディレクトリの階層構造に話題の階層構造が対応している。ファイルに対応するものはショット代表画像である。

左側の話題語をクリックすると，その話題の区間に存在するショットの代表画像が表示される。ショット代表画像をクリックすると，対応するショットからビデオ映像が再生される。

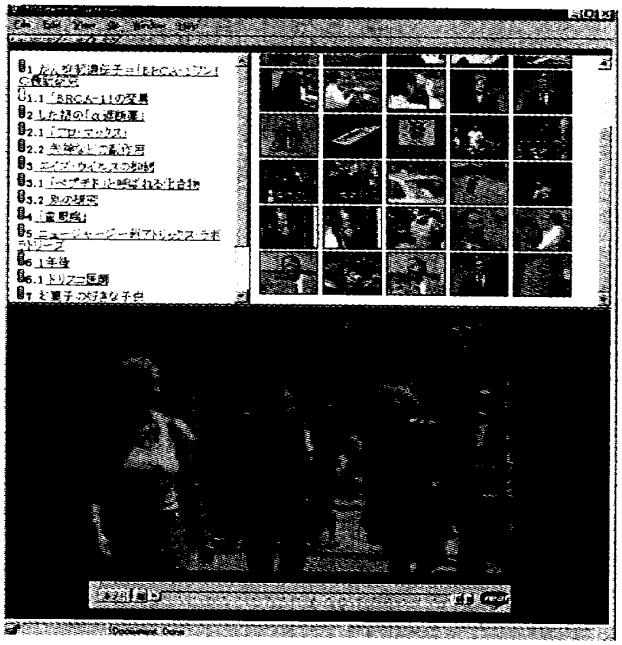

図2: 構造化ビデオブラウザ

4　まとめ

ビデオ映像の内容を効率よく把握するためのマルチメディア速覧とそれを用いた構造化ビデオブラウザについて述べた。現在，テレビ放送における字幕放送と組み合わせたテレビ番組速覧を実験中である。今後，ニュース以外のより広範囲の番組について，話題構造と映像の統合的な提示方法を検討していく。

参考文献

[1] A. Takeshita, T. Inoue, K. Tanaka: "Topic-based Multimedia Structuring", Chap.13 of "Intelligent Multimedia Information Retrieval", AAAI Press, 1997.

[2] 竹下, 井上, 田中: "テキストの概要把握支援のための話題構造抽出", 情報処理学会論文誌, Vol. 37, No.11, 1996.

[3] "2007年めどに全番組に字幕", 朝日新聞, 1997年11月18日朝刊.

[4] 谷口, 外村, 浜田: "映像ショット切換え検出法とその映像アクセスインタフェース", 電子情報通信学会論文誌, Vol. J79-DII, No.4, 1996.