I-2.18 映像ショットと文字キーワードの共起に基づく映像索引付け

Video Indexing Based on Co-occurrence of Shots and Keywords

孟 洋 † Hiroshi MO　山岸 史典 ‡ Fuminori YAMAGISHI　井手 一郎 * Ichiro IDE　片山 紀生 † Norio KATAYAMA　佐藤 真一 † Shin'ichi SATOH　坂内 正夫 † Masao SAKAUCHI

† 国立情報学研究所 National Institute o In ormatics　‡ 東京大学 The University o Tokyo　* 名古屋大学 Nagoya University

Abstract: Recently, it has become possible to handle a large amount of video data with a video archiving system. Indexing is very important for useful access to a large video archive. In this paper, we introduce a method to annotate keywords to important shots. The important shots are extracted by detecting repetitious shots in a video archive. The words as candidates for keywords are extracted from closed-caption texts in video segments corresponding to repetitious shots. Correspondence of shots and keywords are realized by evaluating co-occurrence of repetitious shots and words.

1　はじめに

ハードディスクレコーダーの普及などにより、大量の映像情報の蓄積や利用が可能になるにつれ、映像情報の意味内容に基づく検索やブラウジングなど、大量の映像情報の中から必要とする映像情報を効率的に探し出す機能の開発が重要な課題の一つとなっている。意味内容に基づく検索を行うためには、映像内容の理解の自動化が必要であるが、現在の映像解析技術ではその実現は困難である。

本稿では、映像アーカイブに蓄積されたニュース映像を対象に、統計的な処理に基づき、重要な映像ショットを抽出すると同時に、それら映像ショットへ文字キーワードを付与する手法について検討を行ったので報告する。本検討では、重要な映像ショットは、繰り返し利用される傾向がある、という事実に基づいて抽出を試みた。また、文字キーワードの付与は、抽出された映像ショットと文字放送字幕テキストから得られる語の共起に基づいて行った。

2　共起に基づく映像索引付けの枠組

日々放送されるニュース映像は、画像的には様々なショットから構成されているが、同一あるいは関連する話題（トピック）を継続的に取り上げているため、長期間にわたる多量のニュース映像を考えた場合、注目度の高いトピックに関連した重要なショットや貴重なショット、あるいは資料的なショットは繰り返し利用される傾向がある（例えば、同時多発テロ事件やウサマ・ビンラディン氏を撮影した映像など）。このため、映像アーカイブ内で複数回出現する映像ショット（同一映像ショット）は、ニュース映像を代表する重要なショットの候補となり得る。

一方、多くのニュース映像には、映像情報とともに、テロップや文字放送などのテキスト情報が付随している。これらのテキスト情報は、映像内容を説明するものであり、映像索引付けのための重要なキーワードを提供する。繰り返し利用される映像ショットとそのショットの意味内容を示すキーワードの間には、共起関係が存在すると考えられるため、ショットとキーワードの出現パターンを評価することで、同一映像ショットへのキーワードの付与を実現することが可能になる。共起に基づく映像索引付けの枠組を Fig.1 に示す。なお、ここでは、テキスト情報として、音声を書き下ろした文字放送字幕テキストを用いることとし、映像情報とテキスト情報は同期がとれているものとしている。

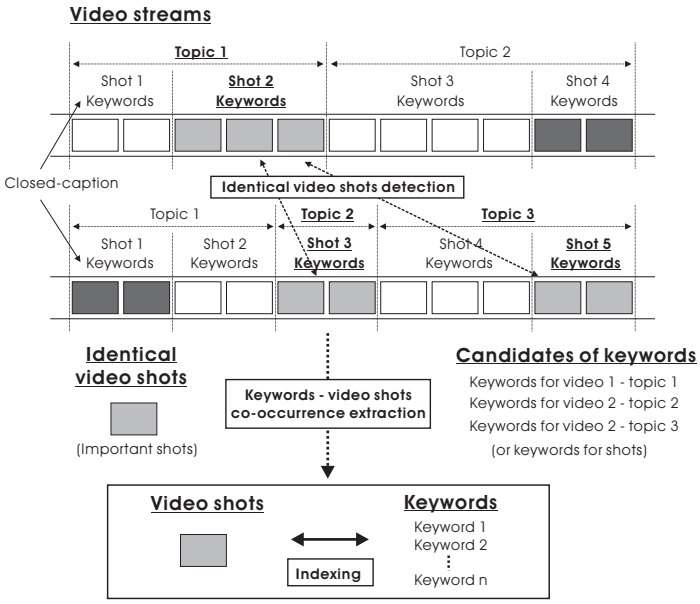

Figure 1: Framework of indexing based on co-occurrence

3　重要な映像ショットの抽出

繰り返し利用される映像ショットは重要なショットの一つとなり得るとの考えに基づき、映像アーカイブ内の異なる映像ソースにおいて、二箇所以上で繰り返して利用されている同一映像ショットを抽出することで、重要なショットの候補を求める。同一映像ショットの抽出は、画像的にほとんど同一のフレームを検出 [1] し、輝度情報を用いたショット分割の結果とあわせ、それらをショット単位へ統合することにより実行する。“ 同一 ”の判定には、全画面での精密な照合を行う必要があるため、輝度情報を用いた

フレーム間の正規化相互相関を利用し、相関値がきわめて1に近い場合を“ 同一 ”であると判定することとした。

4　キーワードの抽出と対応付け

映像内容を説明するキーワードの抽出は、形態素解析により、文字放送字幕テキストから名詞列を抜き出すことで実行する。同一映像ショットへのキーワードの付与は、それらの共起を評価することにより行うが、そのためには映像ソース全体にわたって得られるキーワード列から、抽出された同一映像ショットの内容を指し示すキーワード列の範囲を求めておく必要がある。

ショット単位で内容が変化する映像情報の場合、その同一映像ショットと同時にあらわれるキーワード列をその範囲とするのがよいと考えられるが（例えば、Fig.1において、video1-shot2のショットに対し、video1-shot2-keywordsのキーワード列が対応）、ニュース映像の場合、トピックを単位として意味的に切り分けられるため、ここでは、その同一映像ショットが含まれるトピックと同時にあらわれるキーワード列をその範囲とする（例えば、Fig.1において、video1-shot2のショットに対し、video1-topic1(shot1,shot2)-keywordsのキーワード列が対応）。つまり、同一映像ショットの内容を指し示すキーワードの候補は、その同一映像ショットを含むトピックに対応する文字放送字幕テキストから抜き出された名詞列ということになる。

トピックの分割は、文字放送字幕テキストから得られる名詞列に基づいて行う。具体的には、形態素解析により得られた名詞列から、出現頻度付きのキーワードベクトルを求め、前後の文を結合させるなどして隣接する文間の距離（cos距離）を評価することで、分割点を求める [2]。

同一映像ショットに対する映像索引付けは、そのショットが利用されている各トピックに対応する名詞列から、共起する可能性の高いキーワードを求め、ショットとの対応付けを行うことで実現される。

5　実験結果

5.1　同一映像ショットの抽出

2001年9月1日から2003年2月5日までにNHKで放送された「ニュース７」のニュース映像を用いて、同一映像ショットの取得実験を行った。約500日分のデータから、オープニングやエンディングのショットを除いて、約1547種類、それぞれ2 〜46ヶ所で出現している同一映像ショットを抽出することができた。抽出された同一映像ショットの例をFig.2に示す。

5.2　映像ショットへのキーワードの付与

同一映像ショットとして抽出されたオサマ・ビンラディン氏のショット（2001年9月15日のNHK「ニュース７」など）に対して、キーワードの付与の実験を行った。同一映像ショットの画像と付与されたキーワード上位10語をFig.3に示す。名詞として取得された語をすべて利用しているため、「こと」、「の」などの語も出現しているが、一方で「ビンラディン」、「タリバン」などショットの内容を示すのに有効な語を取得することができている。

(a) Important shot　(b) Rare shot　(c) Reference shot　(d) Others