6S－10

ニュース映像データベースの索引づけ手法の一提案

山本　浩司　　浅野　正一郎
東京大学工学部　　学術情報センター研究開発部

1　はじめに

本発表では現在放送されているニュース映像を、予備的な情報を用いることなく意味のあるまとまりごとに自動的に分割し、ユーザが任意のニュース映像を検索し表示できるようなシステムにおける、映像の分割と索引づけに関する検討を行なう。

映像・画像中の内容を把握し、索引づけする手法としては種々のものが提案されており、将来的には映像中の人物や動作を理解することによる手法が可能になると考えられる。しかし、現在の画像処理技術でそこまでの分析をすることは困難であるため、今回は比較的安定かつ容易に実現できる方法として、映像中に表示された文字を文字認識の技術を用いて認識するという方式を提案し、これについて検討を行なう。

2　ニュース映像データベース

一般に映像のデータベース化は、映像中に多くの意味データが表現されているにもかかわらず原データに内容や構成に関する情報が明示的に含まれないという要因から非常に困難であるといわれている。そこで、対象をニュース放送映像に絞る[1,2]ことによりこの要因を排除し、その特徴を利用して実用的な映像データベースシステムを構築する手法を以下に検討する。

2.1　ニュース映像の特徴

他の多くの番組と異なり、ニュース映像には映像の構成が時間的にも空間的にもある程度規定されているという特徴がある。この制約を利用して一連の映像を単なるカット割りでない、内容的に意味のある部分に分割することを比較的安定に行なうことが可能である。

時間的な規定とは、図1に示すような、それぞれのショット（画面の切り替わりのないひとまとまりの映像）の毎回決まったおおまかな流れのことを指す。このような時間的な構成は、番組ごとに多少の差はあるものの、キャスターのショットと取材映像の繰り返しの部分は多くのニュース番組に共通である。

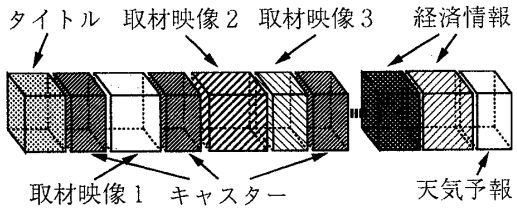


図 1: ニュース番組の流れ

空間的な規定とは、このキャスターのショットにおける画面上の人物などの位置の定常性である。キャスターのショットは一部の、ニュースを軸にした総合番組的なものを除いて、一人のキャスターが画面中央に写っているもの、二人のキャスターが登場しているもの、一人のキャスターの脇に小さな画面が写っているものなどの比較的少数の画面構成を予測しておくことで対処が可能である。

2.2　キャスターのショットの抽出

これらの規定を利用して、ニュース映像を分割する。主な処理は、それぞれのニュースの話題がキャスターのショットから始まっていることに着目したキャスターのショットの抽出である。

まず、処理対象となる映像データは非常に大量のデータになるためにそのままでは扱いづらいため、ショットに分解する。ショットへの分解はすでにさまざまな手法が報告されている[3]。

次に、キャスターのショットを一連の映像中から抽出するために、それぞれのショットについて次のような処理を行なう。

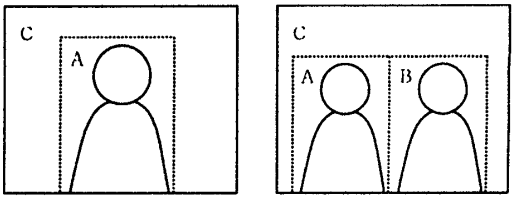


図 2: キャスターのショットの特徴のモデル

キャスターのショットは、図2のようなモデルをあ

An indexing method of news video database
Koji YAMAMOTO [1], Shoichiro ASANO [2]
[1]Faculty of Engineering, The University of Tokyo
[2]Research & Development Department, National Center for Science Information Systems

てはめた場合に、領域(C)では時間的な差分が非常に少なく、(A)や(B)の部分では適当な閾値を上回る差分が検出されると考えられる。このような特徴を用いて一つ一つのショットがキャスターのショットであるか否かを判定することができる。

3 字幕の認識

実際にデータベースを利用するためには、分割された映像が何を表すものかを何らかの形で索引づけする必要がある。映像の作成者が索引情報を同時に放送している場合はそのまま用いることができるが、そうでない場合は映像から索引を抽出する必要がある。そこで、本発表ではキャスターのショットで表示される字幕を背景と分離し、市販の文字認識ソフトウェアによって認識する手法を提案する。

3.1 画像中の字幕の特徴

画像中の字幕はそのほとんどが次のような特徴を持つ。

- 画像中の位置が固定されている
- 色が単一である
- ふちどり、エッジを持つ

これらの特徴を用いて、文字とその背景を分離する。

3.2 画像中からの字幕の抽出

まず、画像中の文字以外の部分を削除するために、画像中から文字の存在する矩形領域を切り出すことを考える。画像と文字の混在文書の複写技術でも提案されているように、エッジの強さによって文字部分と画像部分を分離することを考える。

このときキャスターの洋服の柄によっては多くのエッジが生じ、誤認識の原因となるため、同様のエッジ強調画像を同じショットの中から複数用意し、各画像(A,Bとする)の各画素について式(1)のような演算を行ない、画像Cを生成する。ただし$A(x,y)$は画像Aの(x,y)座標の輝度値という意味である。

$$C(x,y) = \frac{1}{2}\{A(x,y)+B(x,y)\} - |A(x,y)-B(x,y)| \quad (1)$$

このようにして得られたエッジ画像に対し、横方向に走査を行なう。文字が横一列に並んでいることを仮定すると、他の部分との差は大きく、安定に文字部分を抽出できる。この結果をさらに縦方向にも走査することによって矩形領域を切り出すことができる。ニュースの字幕にアンダーラインがついている場合が多いが、この部分については横方向の走査の際に上限を設けることで対処している(図3-a)。

3.3 文字と背景の分離

文字を正しく認識させるためには、文字と文字の間から見えている背景と文字を分離する必要がある。こちらも前処理として、式(1)を用いて、動きのある部分の輝度値を減少させる。

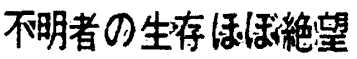
背景の多くはキャスターの洋服であるが、分離が困難になるのはこの色が文字の色に近い場合である。このような場合の処理を以下に検討する。

まず、文字の色が単一であることを利用し、分析対象画像を適当な閾値を用いて二値化する(図3-b)。文字の周囲の背景を除去するためには、さきほど求めた文字領域の境界を大きく越えている部分を除去すれば良い(図3-c)。さらに、文字同士の間から見える背景を除去するために、この画像のヒストグラムから文字の境界を認識し、境界上に広がっている領域を除去する(図3-d)。

このような処理を行なうことにより、入力画像からノイズの少ない出力画像を得ることができる。これを市販の文字認識ソフトにより認識させることにより、自動的に索引をつけることが可能になる。この比較的背景と文字の区別のつきにくい画像の場合でも、処理結果は「布明者の生存ほぼ絶望」であった。

a 

b 

c 

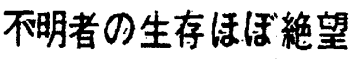
d 

図3: 処理結果

4 おわりに

ニュース映像データベース構築のための索引づけ方法として、キャスターのショットに注目した映像の分割と画像中の字幕を抽出し認識する手法を提案した。今後の課題としてはより一層の文字抽出の安定化、より細かい内容の認識方法の開発と、本手法の適応範囲の統計的な調査による客観的な評価が挙げられる。

参考文献

[1] D.Swanberg, C.-F.Shu and R.Jain: "Knowledge Guided Parsing in Video Databases", SPIE, Vol.1908, pp.13-24(1993).

[2] H.J.Zhang, Y.Gong, S.W.Smoliar, S.Y.tan: "Automatic Parsing of News Video", Proc.IEEE Int'l Conf. Multimedia Computing and Systems, pp.45-54(1994).

[3] 長坂,田中:"カラービデオ映像における自動索引付け法と物体探索法", 情処論, Vol.33, No.4, pp.543-550(1992).