

映像構造化のためのテロップ属性抽出・分類法の提案*

5M-9

桑野秀豪† 谷口行信† 児島治彦†

NTTサイバースペース研究所‡

1 はじめに

大量の映像情報を効率的に管理し、所望の映像をすぐに取り出して利用できる映像システムへの期待が高まっている。これを実現するためには、映像情報を意味的な区間に分割し、それぞれの映像区間の内容を理解・分類する技術が必要である。

映像中に表示されるテロップ文字は、映像の意味的な区間、いわゆる話題毎に決まったルールに基づいて表示効果が施されており、映像の意味的な構造との相関が高い。例えば、ニュース映像では、ニュース項目の最初に見出しを表すテロップ文字が、画面の下に他のテロップよりも大きく表示される。大きさ以外にも、見出しのテロップだけ、色が異なったり、文字に縁取りが付いていたり、帯状の飾りが付いている等、特別な表示効果が施される場合が多い。ニュース映像であれば、見出し部分を検出し、次の見出しまでを一つの区間とすることで、話題単位の抽出が可能となる。このように、テロップ文字には、文字としての言語情報以外に、その表示効果も映像を構造化する手がかりとして利用できる。

我々は、映像中のテロップ文字の認識結果の言語情報を利用した映像インデクシング法に加え、テロップの表示位置、大きさ、色、飾りの有無等の属性情報を抽出、分類し、映像の意味的な区間を取り出す方法の検討を進めている。本検討では、テロップが表示されている画像から、テロップ属性の中でも特にテロップ文字列の高さ情報を抽出する方法を提案する。

2 手法の要求条件および従来法

数種類のニュース映像中のテロップ文字列の高さを目視により調べたところ、見出しテロップの高さは見出し以外のテロップよりも平均して約10

画素程度大きいことが分かった。したがって、本検討では、テロップ文字列の高さを数画素程度の誤差内で求める必要がある。

我々は、これまでにテロップ表示画像の二値化結果中の各連結成分の膨張処理に基づいた文字列抽出法を提案している [1]。ただし、映像中には、テロップの他に様々な輝度、色の任意背景が存在するため、[1]の方法では、二値化結果中に含まれる背景中のノイズ成分も膨張領域に含まれてしまい、実際の文字列範囲よりも大きな範囲で文字列矩形を抽出する傾向があり、要求される精度を満たさない。また、他にもテロップ文字の輪郭部のコントラストに着目し、エッジ画素の多いラインが集中する部分を文字列範囲とする方法が提案されている [2][3]。ただし、この方法も背景から検出されるエッジの影響を受け、文字列の位置、サイズを正確に求めることは困難な場合がある。すなわち、従来の方法では、文字列検出に用いる特徴量に背景部との差を強調することができる「文字らしさ」が十分に反映されていないという問題があった。

3 提案手法

ここでは、二値化の結果得られる連結成分とエッジの検出結果を両方を用いることで、背景ノイズの影響を極力抑制した文字列検出法を提案する。提案法の前提として、予め従来方法 [1][4] を使って、一つの文字列のみを囲む仮の矩形画像、及び文字列の方向情報が抽出されているものとする。この仮矩形内ではテロップ文字の上限、下限は一直線上に並ぶものと仮定し、提案法でも [2][3] と同様、文字列方向のライン毎に文字らしさを示す特徴を求め、文字のラインと背景のラインの境界を検出する。

矩形内の文字列のライン上では、文字の連結成分、及び文字輪郭部からエッジ画素が共にほぼ確実に得られる。一方、背景部のライン上では、連結成分、エッジの個数は共に文字部に比べ少なく、

*Extraction and Classification of Telop Attributes for Structuring Videos

†Hidetaka Kuwano, Yukinobu Taniguchi, Haruhiko Kojima

‡NTT Cyber Space Laboratories

また、必ずしも両者は同時に得られるとは限らない。すなわち、エッジと連結成分は各々単独で扱うと、文字と背景の区別が付きにくい場合もあるが、両者を同時に評価すれば、文字のラインと背景のラインの境界を際立たせることが可能となる。そこで、提案法では、ライン毎に連結成分の個数とエッジ画素の個数の積を文字列特徴値とし、この値がある閾値より高いラインを文字列ラインと判定する。以下に具体的なアルゴリズムを示す。

- (1) 仮矩形内で文字列方向の各ライン毎に文字列特徴値を求める。
- (2) 文字列特徴値の上位 M 個の平均値 Ave を求める。
- (3) 閾値 $Th=Ave/N$ 以上の文字列特徴値を持つラインを文字列ラインとして残す。
- (4) 求めた文字列ラインの最高、最低のラインの間の距離を文字列の高さ、あるいは文字列の幅の値として求める。

なお、以下の実験では、文字列特徴値の計算に使用するエッジ画素の検出には [5] の手法を用い、文字らしい部分から選択的に得られるエッジ画素(エッジペアと呼ぶ)を使用した。また、閾値処理では、 $M=10, N=10$ とした。

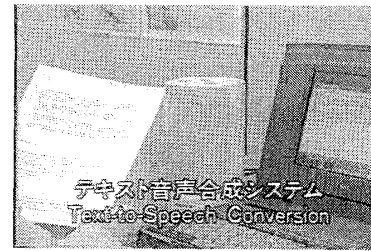
4 実験結果、及び考察

提案アルゴリズムの実データへの適用結果を述べる。図 1 に結果を示す。(a) は入力画像である。(b) は [1][4] の方法を用いて得た文字列矩形の範囲、及び文字列方向のライン毎に連結成分の個数をカウントしたヒストグラムである。(c) は [5] の方法を用いて得たエッジペア検出結果、及びエッジペアのヒストグラムである。(b) では背景ノイズの影響、(c) では文字の縁取り部分からのエッジの影響で文字ラインと背景ラインの区別が付きにくい。(d) は文字列方向のライン毎に得た文字列特徴値のヒストグラム、及び閾値処理により得られた文字列矩形を示す。提案した文字列特徴値により、文字のラインと背景のラインの差が強調されたため、(b) の矩形よりも正確に文字列の上限と下限を抽出できた。

ニュース映像中のテロップ表示画像 140 枚に対し、提案法を用いて全画像中のテロップ文字列の高さ(縦書きの文字列の場合は幅)の値を抽出したところ、双峰性の分布が得られた(図 2 参照)。この分布を [6] の方法で二分割し、高さ値の大きい

方の山に含まれる画像を見出し画像として検出したところ、ニュースの話題の先頭に現れる見出し画像 21 枚を全て正しく検出でき、提案法の有効性が確認できた。

今後は、他のテロップ属性の抽出法、及びニュース以外の映像への適用を検討していく。



(a) 入力画像



(b) 仮矩形と連結成分ヒストグラム



(c) 仮矩形とエッジペアヒストグラム



(d) 提案法の結果と文字列特徴値ヒストグラム

図 1: 実験結果

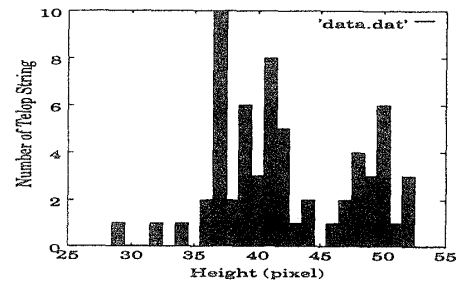


図 2: ニュース映像中のテロップの高さの分布

参考文献

- [1] S.Kurakake, H.Kuwano, K.Odaka, "Recognition and Visual Feature Matching of Text Region in Video for Conceptual Indexing", *Proc.ofSPIE StorageandRetrievalforImageVideoDatabase V*, vol.3022, pp.368-379, 1997.
- [2] Yu Zhong, Kalle Karu, Anli K.Jain, "Locating Text in Complex Color Images", *Proc.of3rdICDAR*, pp.146-149, 1995.
- [3] Y.Ariki, T.Teranishi, "Indexing and Classification of TV News Articles Based on Telop Recognition", *Proc.of4thICDAR*, pp.422-427, 1997.
- [4] 桑野, 新井, 倉掛, 杉村, "ライン単位の二値化による映像中文字領域の抽出方法", 信学秋大, D-12-42, p.264, 1998.
- [5] 桑野, 新井, 倉掛, 小倉, "エッジ密集度によるテロップ文字フレーム検出誤り抑制法", 信学秋大, D-12-22, p.214, 1997.
- [6] 大津, "判別および最小 2 乗基準に基づく自動しきい値選定法", 信学論, D, J63-D-4, pp.349-356, 1980.