ニュースビデオにおける実世界建物に関する情報検出*

5K－3

金 浩民, 徐 旭, 柳沼 良知, 坂内正夫 †
東京大学生産技術研究所

1 はじめに

現在, 我々は放送されているニュースビデオを処理対象とする実世界媒介情報システムを開発している. システムの基礎部分である市街地建物データベースの中には, 実世界建物に関する情報 (名前, 所在地, 所有者, 地図, 画像等) を含める. 本システムはこの市街地建物データベースに基づいて, 放送されているニュースビデオの中に映す実世界建物に関するコンテンツを認識して, それに関連する地図情報をユーザへ自動的に提供することを目標としている. 具体的には, 本システムでは, 二種類の情報メディアを利用している. 一つは画像情報 (テクスチャ, カラー等) であり, もう一つは映像の中に映す字幕情報である. 本文では, この二種類の情報メディアに対して, 実世界建物に関する情報の検出手法を提案する. 二章では情報検出のために用いた識別モデルについて. 三章では実験について. 最後, まとめと将来の研究について述べる.

2 情報検出のために用いた識別モデル

2.1 シーンチェンジの検出

ショットの検出技術は, 映像データベースの構築領域で, よく研究されている. 本システムに対して, まずニュースビデオをショットと言う部分に分割する. そして, ショットごとに第一番目のフレームの上で, 内容解析を行う. 本システムでは, 文献 [1] に提案された"step-variable"と言うアルゴリズムを用いて, シーンチェンジの検出を行う. この手法でシーンチェンジを検出して, ショットごとに第一番目のフレームを選んで, 後の処理で使う.

2.2 "特徴フレーム"と顔を映すフレームの検出

ニュースビデオの中に特別な意味を代表しているショット, 例えば, 図 (1) に示しているような特定的な放送内容を表すショットを多く映している. これらのフレーム (ここで, "特徴フレーム"と言い) は実世界建物情報と関係がないので, 検出されて取り除く必要がある. このため, 本システムではまず, このようなコンテンツに関するいくつのフレームを選んで, "特徴フレーム"データベースを作る. そして, "wavelet transform"モデル [2] を用いて構築されたデータベースのビジョンインデックスに基づいて, "特徴フレーム"を検出する.

図 1: Some examples of "feature frame" representing special meaning.

また, ニュースビデオの中にアナウンサーや人物などの顔を映す映像部分が多いので, これに関するフレームも取り除く必要がある. このため, YES カラー空間 [3] の E、S カラー要素に基づく, いくつのアナウンサーと人物の顔を映すフレームの上で顔の部分をマニュアルでマックして、顔の色の学習を行う. 顔の色を検出する際に, "bayesian decision rule for minimum cost"[4] の手法を用いて, フレームの中から顔の色表す画像点を検出する. このように検出されて来た画像点の分布状況によって, 統計的手法でフレームの上で顔を映すかどうかを求める.

2.3 建物オブジェクトを映すフレームの検出

ニュースビデオの中に実世界に関する重要な情報ソースは建物を映すフレームである. 本システムでは, エッジの情報を利用して, あるフレームの中に建物を映すかどうかの識別処理を行う. このため, まず, フレームから, 全てのエッジを検出して (図 2), エッジの角度を計算する. 次に, 特定な閾値で, 長いエッジと短いエッジを判定し, 映している建物は少し傾きがある状況を考慮した上, 三つの角度範囲の内に, 或は, $-10°-+10°$, $80°-100°$, 他の角度範囲の内にある長いエッジの数, 短いエッジの数を各々に計算する. 最後

*Detection of information relating to building object in news video

†Haomin Jin, Xu Xu, Yaginuma Yoshitomo, Masao Sakauchi
Institute of Industrial Science,The University of tokyo
7-22-1 Roppongi, Minato-ku, tokyo, 106 Japan

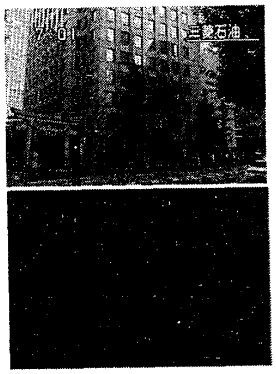

図 2: An example of frame containing building objects and the edges extracted from the frame

に, これらのエッジ角度の分布状況によって, 建物オブジェクトをフレームの中に映すかどうかを判断する.

2.4 ニュースビデオに映す字幕情報の利用について

ニュースビデオに映す字幕情報は一つの重要な情報ソースである. 字幕の中に, 実世界に関する情報, 例えば, 場所の名前, 建物の名前等を映すことがよくある. この実世界に関する情報を自動的に検出して, 認識できれば, それに関する地図情報を提供できると考えられる. このため, 我々は本研究室のテロップシステムを用いて, ニュースビデオのフレームから字幕情報を検出する. そして, この検出されて来た字幕情報を用いて, 我々に構築されている市街地建物データベースを検索する. 本システムでは, このテロップシステムで字幕を検出する際に, 仮名と漢字を間違って検出することが多いので, "Finger prints"[5] と言う手法で, 市街地建物データベースの建物名前に関する情報インデックスを構築する. 検索する際に, 検出して来た字幕テキストの中から, 同じ"Finger prints"手法で, 相応な"Finger prints"情報を抽出して, データベースを検索する.

3 実験

実験では, 我々は二時間の NHK ニュース番組のビデオから実験用のニュースビデオを製作して実験をしていた. 実験では, 処理の計算量と複雑さの角度から考えると, まず, 本文に述べたシーンチェンジの検出手法を用いて, シーンチェンジを検出する. カメラが快速的に動くことがあるのため, シーンチェンジではない部分をシーンチェンジとして検出したことがあったが, シーンチェンジの部分は全て検出されていた. 次に, これらの検出されて来たショットごとに, 第一番目のフレームを選んで実世界建物に関する情報検出を行う. このため, まず, 二つの手法で建物情報と関係がないフレーム画像を検出して取り除く. 一つの手法では, "特徴フレーム" を検出するため, ニュースビデオから 10 種類の"特徴フレーム"を選んで, "wavelet transform" モデルに基づく, "特徴フレーム"データベースの視覚インデックスを構築する. そして, このデータベースに基づく, "特徴フレーム"の検出を行う. 実験では, 特定な閾値を設定して, 全ての"特徴フレーム"が検出できた. もう一つの手法では, 顔の色の学習によって, 統計的な手法でアナウンサーや人物などの顔を映すフレームを検出する. 検出率は 83 %である. 最後に, 以上の処理で残って来たフレームから, エッジの情報に基づいて, 建物オブジェクトを映すフレームを検出する. 実験では, 建物情報に関するフレームの検出率は 77 %である. また, テロップ情報の利用に関する実験では, 我々は本研究室のテロップシステムを用いていた. まず, 2 時間の NHK ニュースビデオから, 字幕が 80 %以上に認識できた 14 枚フレームの字幕テキストを用いて, 市街地建物データベースを検索していた. 相応な建物に関する情報が検索できた.

4 まとめと将来の研究

本文では, ニュースビデオにおける実世界建物に関する情報の検出手法について述べていた. 特に, いくつのモデルを用いて, 映像フレームのコンテンツを分析し, 相応なオブジェクトを検出することができた. 将来の研究について, 実世界に関する情報を検出するため, 他の識別情報 (例えば, オブジェクトの動き情報など) を利用する手法を研究する. また, 建物認識の研究も一つの重要な研究テーマである.

参考文献

[1] W. Xiong and J. C. Lee, "Efficient Scene Change Detection and Camera Motion Annotation for Video Classification", Computer Vision and Image Understanding, Vol.71, No.2, pp. 166-181, 1998.

[2] Jun-Wei Hsieh, Hong-Yuan Mark Liao, Kuo-Chin Fan,m Ming-Tat ko, and Yi-Ping Hung, "Image Registration Using a New Edge-based Approch", Computer vision and image understanding, Vol. 67. No. 2, pp. 112-130, 1997.

[3] E. Saber, A. M. Tekalp, R. Eschbach, and K. Knox, "Automatic Image Annotation Using Adaptive Color Classification", Graphical Models and Image Processing, V0l.58, No.2, pp.115-126, 1996.

[4] H. Wagn, S. F. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video", IEEE Trans. on circuits for video technology, Vol. 7, No. 4, pp. 615-628, 1997.

[5] J. T. Wang, C. Chang, "Fast Retrieval of Electronic Messages That Contain Mistyped Words or Spelling Errors", IEEE Trans. on Systems, man, and cybernetics, Vol. 27. No. 3, 1997.