

地図情報を利用した 実世界映像のインデックス付けの一手法*

3Q-5

大場 敏文, 谷田部 智之, 坂内 正夫†

東京大学生産技術研究所‡

1 はじめに

筆者らは、放送された映像フレーム内に存在する個々のオブジェクトに対して複数のユーザが記述した情報をネットワークを通じて収集し、それらをインデックスとして利用して、内容に基づく検索や映像に対する質問応答を可能とする映像データベースシステム ADTV(Advanced Database TV) を提案している [1][2]。

ADTV では、多くのユーザが互いに情報を共有できるため、各ユーザは豊富な情報を入手できるが、入手できる情報はユーザが記述した情報に限定される。ユーザが記述した情報をもとにシステムが映像を理解して、それから得られる情報を提供すれば、利用価値が高まる。

本稿では、対象とする映像を実世界映像に限定し、ADTV において、ユーザが記述した情報と地図から得られる情報を用いて映像を理解する手法について述べる。

2 提案する手法

2.1 実世界映像

実世界映像とは、屋外の情景を撮影した映像を指す。本稿では、進行方向に対して垂直にカメラを向け、道路沿いの建物を車載カメラで撮影した映像を利用した。

2.2 地図における建物

地図において、建物は閉折線として表現されている。閉折線は 2 次元座標の集合であるが、建物を表す閉折線にはその建物に関する情報が関連付けられている。

2.3 実世界映像に対する建物に関する情報の記述

実世界映像を対象とした ADTV では、建物が映っている矩形領域を指定することにより、建物に関する

記述を行う。実世界映像に対して建物の情報を記述するという事は、映像中のあるフレームにおける建物の映っている矩形領域に地図の中のひとつの閉折線に対応させるということの意味する。

2.4 地図から得られる建物のエッジパターン

図 1 は使用した地図の一部分を示す。図 1 の中央部の大きな矢印は撮影の進行方向を表し、左部の文字は各矢印が指し示す建物の名前を表す。ここでは説明のためにアルファベットを用いているが、地図では建物の実際の名称が関連付けられている。



図 1: 地図

ある建物が記述されれば、その映像がどの道路に沿って撮影されたものかがわかる。その道路の方向がカメラの進行方向であり、さらに、映像の大域的なオプティカルフローからカメラの進行の向きが判断できる。また、カメラの向きは道路と垂直な方向であるため、道路沿いの建物を表す閉折線群に関して、カメラの方向の直線との接線のパターンをとれば、道路に沿った建物のエッジのパターンを得ることができる。図 2 は、このようにして得られた建物のエッジパターンである。

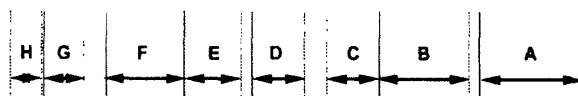


図 2: 地図から得られる建物のエッジパターン

* A Method of Indexing Real-World-Video Using Map Information

†Toshifumi OHBA, Tomoyuki YATABE, Masao SAKAUCHI

‡Institute of Industrial Science, University of Tokyo

2.5 空間統合映像座標系

対象とした実世界映像は進行方向に対して垂直にカメラを向けて撮影した映像なので、静止した被写体は映像中では水平方向に動く。

映像中の静止した被写体は、各フレームの座標系における被写体の位置座標からフレーム間の大域的なオプティカルフローを積算したものを減じることによって得られるような座標系に変換することにより、実世界における位置関係をほぼ再現することができる。

このように、オプティカルフローを利用して実世界の位置関係を再現した座標系を、空間統合映像座標系と呼ぶことにする。

2.6 地図から得られる建物のエッジパターンと空間統合映像座標系の対応

建物を記述する時に指定した矩形領域の枠の2本の垂直線分は、地図から得られる建物のエッジパターンの中の2本の線分と対応する。したがって、それぞれの線分間の距離をもとに線形変換をほどこすことにより、地図から得られる建物のエッジパターンと空間統合映像座標系を一致させることができる。図3は、図1の建物Dに関して記述が行われた場合の例を示している。建物Dは図3の下部に示すように、あるフレームにおいて矩形領域を指定して記述が行われている。

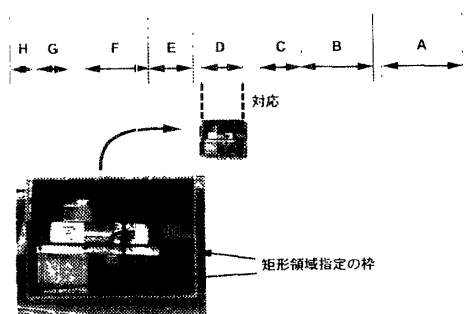


図3: 地図から得られる建物のエッジパターンと空間統合映像座標系の対応

地図から得られた建物のエッジパターンの各領域は、地図における閉折線と対応し、その閉折線には建物に関する情報が関連付けられているため、上の変換によって実世界映像中の被写体を地図の建物の情報と関連付けることができる。

3 適用例

図1の建物Dに関して記述を行った場合の適用例を示す。

3.1 映像中に映っている建物への質問

あるフレームにおいて建物の映っている部分を指して「映っている建物が何か」と尋ねる質問に対して、建物Aから建物Hに関して正しい結果が得られた。ただし、建物が映っている領域の端の方を指すと、正しくない結果が返ってくる場合がある。

3.2 特定の建物が映っているフレームの検索

建物Aから建物Hまでの建物に関して、それが映っているフレームを検索する要求に対して、建物がフレームインしてからフレームアウトするまでの連続したフレームすべてが、ほぼ正確に得られた。検索対象の建物が記述された建物から離れれば離れるほど、検索結果のフレーム群が時間方向に多少ずれる傾向がある。

3.3 考察

良好な結果が得られたが、映像中の建物の映っている領域が完全に地図の情報と対応したわけではない。これは、地図から得られるエッジパターンと空間統合映像座標系が完全には一致しないからである。

4 まとめ

実世界映像を対象としたADTVにおいて、システムが地図情報を利用して映像を理解して、それから得られる情報をユーザに提供することにより、ユーザが記述した建物以外の建物の情報を得ることができるようになった。

本手法を適用することにより、数軒おきに代表的ないくつかの建物が記述されれば、ユーザはほとんどすべての建物に関する情報を得ることができるようになるであろう。

参考文献

- [1] 谷田部智之, 佐藤隆, 坂内正夫, “放送間のリアルタイムリンクを可能とするアドバンスドTVの構想”, 信学ソサイエティ大会, D-279, Sep. 1996
- [2] 谷田部智之, 大場敏文, 坂内正夫, “ネットワーク上での構造化を用いた対話型映像情報システムの提案”, 信学技報, PRMU-97(IE-97, MVE-97), Jul. 1997