

6Y-5

実世界映像中の建築物に基づいた インデックス付けの手法とその評価*

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、ネットワーク上での情報共有を可能とする対話型映像情報システムである ADTV を提案している [1][2]。

文献 [3] では、実世界映像を対象とした ADTV において、ユーザが記述した情報をもとに、地図をモデルとして用いて実世界映像に対してより詳細な記述を行う手法について述べた。しかし、その手法では、ユーザが映像中の建築物の映っている領域を矩形で指定する必要がある、ユーザにとって煩雑に感じるものであった。

本稿では、記述の際のユーザのインタラクションの負担を軽減した場合に実世界映像中の建築物の領域を推定する手法を提案し、その手法の実験・評価を行う。

2 映像中の対象物の指定

文献 [3] では、ユーザが実世界映像中の建築物の情報記述を行う際に、映像画面中でその建築物が映っている領域を矩形で指定する必要がある。しかし、この作業を流れている映像の上で行うことはユーザにとってあまり容易なことではなく、煩雑に感じるものであった。記述したい対象物の中心付近1点を指定するという方がより自然な指定の仕方であり、この方法ならば流れている映像の上でも簡単に行うことができる。本稿では、このように指定の仕方を改良し、それに応じた推定の手法を提案する。

*An Evaluation of a Method of Indexing Real-World-Video by Building Information

†Toshifumi OHBA, Tomoyuki YATABE, Masao SAKAUCHI

‡Institute of Industrial Science, University of Tokyo

3 実世界映像中の建築物領域の推定

ユーザによって記述された建築物の情報をもとに、実世界映像中に映っている他の建築物の領域を推定する手法について述べる。

地図に対して幾何的な演算を実行することにより、道路から見た場合の建築物の境界のパターンを得ることができる。このようにして得たパターンを実世界映像中の建築物領域推定のためのモデルとして利用する。

実世界映像の静止被写体のカメラに対する相対的な移動を算出し、その情報をもとに静止被写体の共通領域を重ね合わせるように各フレームを配置するような座標系を想定すれば、実世界映像をパノラマ画像のように扱うことができる。このような処理により、被写建築物の実世界における位置関係を再現することができる。

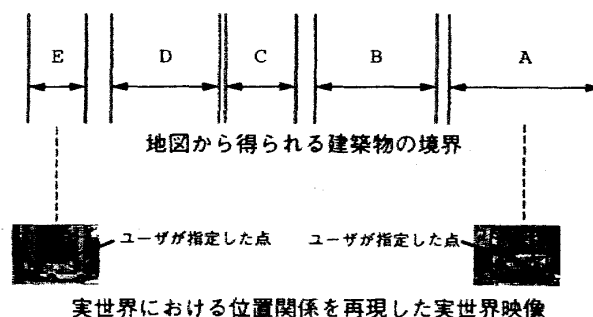


図 1: 建築物境界の実世界映像への重ね合わせ

地図から得られる建築物の境界と実世界における位置関係を再現した実世界映像は、スケールをそろえることによって一致させることができる。スケールをそろえるには対応する点が2組与えられれば良い。地図から得られる建築物の境界に対する点は、ユーザが記述した建築物をあらわす区間の中心点とする。実世界における位置関係を再

現した実世界映像に対する点は、記述を行う時にユーザが指定した点とする。それぞれ対応する点の位置が一致するようにすることによって、実世界映像に対して建築物の境界を重ね合わせることができ、実世界映像中の建築物の領域を推定することができる(図1)。

地図から得られる建物の境界の各区間は、地図中の建築物に関する属性情報と関連付けられている。したがって、この対応付けによって地図の属性情報が実世界映像中の建築物に対しても関連付けられる。

4 実験・評価

実世界映像に対して2軒の建築物を記述し、それらの間に存在する建築物の領域を上記の手法により推定する実験を行った。記述する建築物の間に存在する建築物の軒数を変化させ、それぞれに対して20回の記述を試行した。ただし、記述する建築物の間に存在する建築物の軒数には記述された建築物も含む。そして、ADTVの機能として想定している質問応答や検索の精度を算出し、この手法の評価を行った。

表1は「映像に映っているこの建築物は何か」という質問応答に対する正答率である。表の数値は20回の試行の平均値と標準偏差を示している。

表 1: 質問応答正答率

建物数	4	6	8	10	12
平均	0.986	0.955	0.881	0.743	0.625
標準偏差	0.020	0.065	0.038	0.078	0.056

ここで、質問応答の正答率とは、ユーザが映像中で建築物が占める領域の中央付近(建築物領域の水平方向中央2分の1以内)を指して質問を行った場合に正答が得られる確率とした。記述を行うユーザは記述をする対象の建築物の中央を正確に指定するわけではないので、記述の際の指定誤差により実世界映像中の建築物領域推定の精度にばらつきが生じる。

ユーザに記述された2軒の間に存在する建築物が少ないほど正答率が高くなるということがわかる。したがって、ネットワーク上で多数の建築物

に記述が行われたら、推定の際にはなるべく近い建築物から推定する方が良いということが言える。記述された建築物の間に存在する建物が10軒程度以下だったら、この手法は有効だということと言える。

表2は「ある建築物が映っているフレームが見たい」という検索に対する適合率である。表の数値は20回の試行の平均値と標準偏差を示している。

表 2: 検索適合率

建物数	4	6	8	10	12
平均	1.000	1.000	1.000	1.000	1.000
標準偏差	0.000	0.000	0.000	0.000	0.000

実世界映像中で特定の建築物が映っているフレームはある連続したフレーム群となる。ここでは、検索結果としてその建築物が映っていると推定される連続したフレーム群の真中のフレームをユーザに提示するものとした。実験では全ての試行に関して検索の正答率は1であり、このシステムの検索機能に関しては十分な精度が得られるということが言える。

5 まとめ

実世界映像を対象としたADTVにおいて、ユーザの記述の仕方を単純化した場合に、記述された建築物の間に存在する建築物の領域を推定する手法について述べ、その実験・評価を行い、手法の有効性を確認した。

参考文献

- [1] 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, IE97-33, PRMU97-64, MVE97-49, Jul. 1997
- [2] 谷田部智之, 大場敏文, 坂内正夫, “ネットワーク上での情報共有を可能とする対話型映像情報システムの提案”, 情報処理学会第55回全国大会, 3R-10, Sep. 1997
- [3] 大場敏文, 谷田部智之, 坂内正夫, “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”, 情報処理学会第55回全国大会, 3Q-05, Sep. 1997