論 文

地図とリンクした実写映像のインタラクティブ操作

Interactive System of Real-world Video Based on Maps

谷田部 智之†, 川 崎 洋†, 正会員 坂 内 正 夫†

Tomoyuki Yatabe†, Hiroshi Kawasaki† and Masao Sakauchi†

Abstract The amount of contents we can obtain through TV broadcasting by satellite, CATV, and the Internet is increasing day by day. In this paper, we propose a method and an application for future Interactive TV service, called ADTV. In our method, users not only get information about specific video objects, but they can describe video objects about them using the incomplete description of portion in each frame. We discuss automatic structuring of video objects automatically using incomplete description. We implement the prototype system to provide users with some basic functions such as "Description", "Question and Answer", and "Retrieval" regarding real-world video based on digital maps. The system can link video objects to digital maps so that users can retrieve an image of a building or get information about an unknown building in a video.

キーワード：映像データベース，実世界映像，映像モザイク，インタラクティブ放送

1. ま え が き

放送のディジタル化による多チャネル化やネットワークを使った映像情報の配信などにより，視聴者が手に入れられる映像コンテンツは増加の一途をたどっている．また，コンピュータの性能向上，安価なメモリやディスク，MPEGを中心とした符号化技術などにより，映像を扱うための物理的な制約は少なくなりつつあるため，ディジタル映像を効率良くハンドリングするための技術が求められている．

手に入れられる映像が増える一方で，多数の番組が同時に放送されることにより，放送する映像コンテンツの不足も考えられる．そのため，放送される映像コンテンツとして，編集されていない映像やリアルタイムに映し出される街の映像などが多数用いられることが予想される．従来そのような映像に対しては，放送する前段階において，テロップや地図などの映像に関連した情報を付加するのが一般的であったが，リアルタイムに流れる映像についても同様に関連する情報を得たいという要求が考えられる．また，現在の放送では，すべての視聴者は一方的に放送される同一の情報しか得ることができず，各視聴者それぞれのニーズを満たしているとはいいがたい．そこで，ネットワークを利用した双方向形放送によりこうした要求が満たされることが期待される．

1999年1月11日受付，1999年8月23日最終受付，1999年9月9日採録
†東京大学 生産技術研究所
（〒106-8558 東京都港区六本木7–22–1, TEL 03–3402-6231）
†Institute of Industrial Science, The University of Tokyo
(7–22–1 Roppongi, Minato-ku, Tokyo, 106-8558, Japan)

筆者らは，各視聴者のニーズに答えられるようなシステムとして，映像と関連する情報のデータベースとをリンクする映像情報システムADTV（Advanced Database TV）を提案している[1)2)]．**図1**に示すように，放送される映像に含まれる，映像オブジェクトと呼ぶ対象物それぞれに対して，映像の内容に基づいて記述された情報と動画像から抽出されるオブジェクト情報を合わせて，映像オブジェクトデータベースを構築する．ネットワークを通じて，このようなデータベースを共有するにより，映像中のオブジェクトあるいは映像全体に関して，内容に基づく検索やリアルタイムに行われる質問に対する応答などのインタラクティブな操作が可能である．この時，映像に付与される情報としては，例えば，多数のユーザが放送を見ながら対話的に記述するもの，情報発信者あるいはサービス提供者が正確に記述するものや映像認識技術により自動的に記述されるものなどを想定している．

このように映像に付与された情報はデータベース化され，映像のハンドリングに利用することができる．これまでにも，カットあるいはシーンなどの連続したフレームの集合を単位として，映像集合それぞれに関連した情報を付加することにより映像のハンドリングを行う研究[3)~5)]が行われてきているが，提案しているADTVではフレーム内の映像オブジェクト単位にインデックスを付けることにより映像のハンドリングを行うことを目指している．

しかしながら，各映像オブジェクトに対し情報が付与されたとしても，それらは必ずしも映像の内容に合致した情報であるとはいえず，また，視聴者それぞれのニーズを満

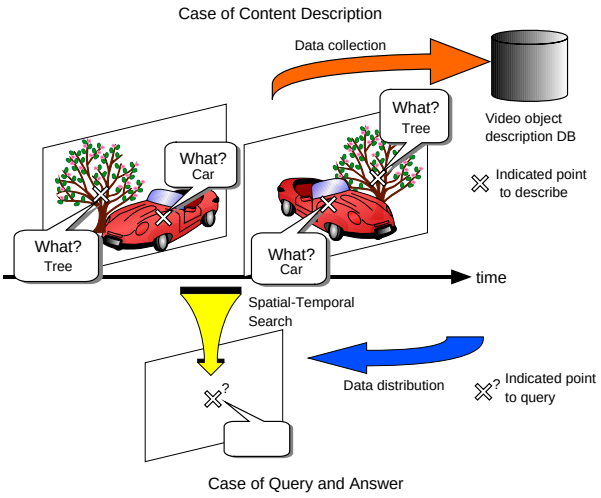


図 1　ADTV の構想
Basic concept of ADTV System.

たす情報であるとは限らない．特に，リアルタイムで放送される映像の場合には，関連した情報そのものが与えられない映像オブジェクトが存在することも充分考えられる．

このように付与される情報は不完全であることから，ユーザの質問や検索などのインタラクティブな操作に対応するためには，映像オブジェクトに付与されている不充分なインデックスから推定する必要がある．例えば，**図 2** のように，映像の第 n フレームにおいて，映像中のオブジェクトに対する情報記述がある場合にも，他のフレームに存在する同一のオブジェクトに対する質問や，同じフレームに存在する隣接した他のオブジェクトに対する質問に，システムは答えることができない．この問題を解決するためには，ユーザの記述情報を基に映像オブジェクトの時間的・空間的な連続性を利用し，システムが自動的に映像オブジェクトを構造化する必要がある．時間方向への構造化は，各種のオブジェクト追跡手法により行うことが可能であり，空間方向への構造化に関しては，一般的にはセグメント手法あるいはオブジェクト抽出手法を用いることになる．しかし，それらの時間的・空間的処理により抽出された，3 次元構造を持つオブジェクトが，必ずしもユーザが意図したオブジェクトの構造と一致するとは限らない．そのため，ユーザの意図したオブジェクトと抽出されたオブジェクトが同一のものであることを示すためには，各オブジェクトに特有のモデルを用意する必要がある．

本論文では，これまで述べてきたようなシステムの一実現例として，道路上を移動するカメラから撮影した映像中に存在する建築物を映像オブジェクトとして扱い，地図との対応付けを行うことで，各視聴者がリアルタイムに流れる映像に対し，質問応答や検索などのインタラクティブな操作ができるシステムを提案する．

2.　地図と実世界映像とのリンク

このような ADTV システムの実例として，時々刻々変

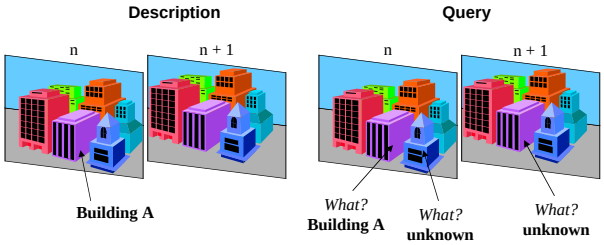


図 2　不充分な記述に基づくインタラクティブ操作
Query based on insufficient description of video objects.

化する実世界を映す映像を対象としたシステムを構築する．本論文では，このような映像を実世界映像と呼ぶことにする．ここでは，進行方向に対して垂直に向けた車載カメラで道路沿いの建築物を撮影した実世界映像を対象とし，映像オブジェクトに対する不完全な記述とモデルとなる地図を利用し，建築物の映像と地図とを自動的にリンクさせることにより，映像をインタラクティブに操作できる環境をユーザに提供することができる．

始点と終点となる建築物の映像と地図とが対応付けされた場合に，以下のような手順により，対応付けされた建築物の間あるいは前後に存在する建築物について自動的に情報を付与することができる．ただし，情報を自動的に付与できる建築物は，撮影された道路に対し同じ側に面している必要があり，映像は交差点で曲がったりすることはなく，同一の道路に沿って移動するカメラから撮影されたものとする．

（1）　映像から，建築物群のパノラマ画像を生成し，それぞれの建築物についての位置を求める．
（2）　パノラマ画像に含まれる建築物の位置を示すモデルを地図から生成する．
（3）　パノラマ画像と地図から生成したモデルの境界が一致するようにマッチングを行う．

2.1　実世界映像のモザイキング処理

ここで対象とする実世界映像は，道路に沿ってカメラが移動して撮影されたものであり，映像オブジェクトである建築物はそれぞれカメラに対して相対的に同一の速度で移動していることになる．そこで，カメラの向きは移動方向に対し，常に垂直であることを利用し，カメラに対する相対的な移動をブロックマッチング法によって算出する．ここで求められたパラメータを用いて映像モザイキングを行う [6]．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と，実世界映像からパノラマ画像を生成することができる．以上の処理により，**図 3** に示すようなパノラマ画像を生成することができ，映像中に含まれる建築物のパノラマ画像における位置関係を求めることができる．

2.2　地図から建築物境界パターンモデルの生成

前述の手法により生成されたパノラマ画像と，地図から生成されるモデルをマッチングすることにより映像と地図

図 3 カメラ移動に基づいたパノラマ画像の生成
Making panorama images based on camera motion.

とをリンクすることができる．ここでは，電子地図における建築物の形状を道路の中心線に射影することにより得られる建築物の境界，つまり，道路から見た場合の建築物の形状をパノラマ状に配置したものを境界パターンモデルと呼ぶことにし，パノラマ画像に対応した範囲の境界パターンモデルをマッチングに用いることにする．

なお，このシステムではどの道路に沿って撮影されたかという情報は直接与えられず，実世界映像中の 2 つの建築物と地図上の建築物を表すポリゴンとの対応付けのみが得られるため，対応付けられたポリゴンの地図上の位置から撮影された経路を決定する必要がある．

そこで，ユーザにより対応付けられた 2 つの建築物を始点，終点とする境界パターンのモデルを生成するために，以下のような作業を行う．

（1） 始点および終点として対応付けされた建築物が同一の道路上の同じ側に面していると仮定し，2 つの建築物が共通に接する道路を探索し，それを撮影された経路とする．

（2） 例えば，**図 4** において，始点および終点として対応付けされた建築物をそれぞれ A，G とすると，手順（1）で求めた撮影経路に沿って，A–G 間に存在する建築物の地図上における形状を道路の中心線に射影することにより，図 4 下部のような建築物境界パターンモデルを生成する．

2.3 境界パターンモデルと実世界映像とのマッチング

モザイキング処理により得られたパノラマ画像と建築物の地図上における形状を道路の中心線に射影して得られる境界パターンモデルとは，縮尺と始点と終点の位置を合わせることにより全体を一致させることができると考えられる．

例えば，**図 5** の上部のような映像において，リアルタイムに流れる映像中の建築物 A，G と地図上の図形とを対応付ける場合には，ユーザは建物の中央付近を指示すると仮定する．ここで指示された点を図 3 に示したパノラマ画像における座標に変換し，ユーザの指示した始点および終点の 2 点を，図 5 の下部のように，境界パターンモデルにおけるそれぞれに対応した建築物の中央の点に対応させる．ここで対応付けられた始点終点を基にパノラマ画像と**図 4** に示した境界パターンモデルとのマッチングを行う．そこ

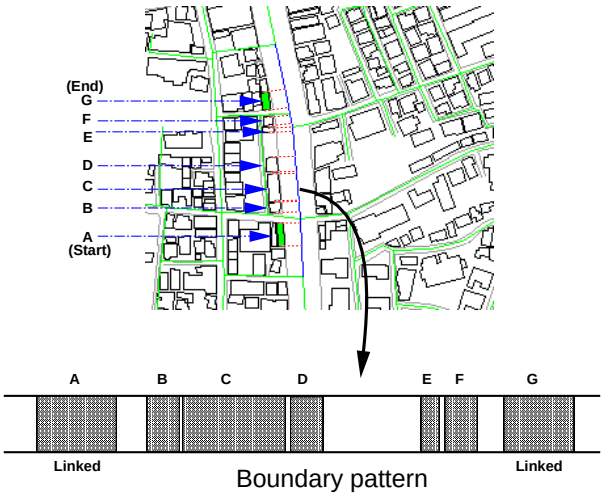


図 4 建築物の境界パターン
Boundary pattern of buildings.

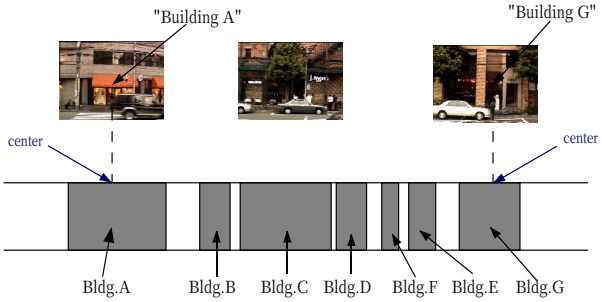


図 5 モデルと実世界映像とのマッチング
Matching between models and real-world images.

で，パノラマ画像における始点から終点までの長さに対し，境界パターンモデルにおける対応する長さを線形に変化させることにより，パノラマ画像と境界パターン全体を対応させる．この対応に基づき，実世界映像に対して建築物の境界パターンモデルを重ね合わせることができ，実世界映像中の建築物 B から F についても，自動的に地図と対応付けることができる．

このように，実世界映像とモデルとを対応させることにより，対応付けのない建築物に関しても，ユーザの質問や検索といった要求に応じることができるようになる．

3. 実験システムの実装

これまで述べたように，リアルタイムに流れる実世界映像中に存在する一部の建築物と地図との「対応付け」をユーザが行うことにより，その間に存在する他の建築物についても自動的に対応付けを行うことができる．このように対応付けがなされた場合に，リアルタイムに流れる実世界映像中に存在する建築物それぞれに関して「質問応答」や地図上で示された建築物に関する「検索」を行えるような実験システムを実装した．Silicon Graphics 社のグラフィックスワークステーション上の X-Window System のアプリケーションとして開発を行った．

境界パターンモデルを生成するための地図データとして，ゼンリン住宅地図 Zmap-TOWN と数値地図 2500 を使用した．建築物を表すポリゴンと建築物の属性情報，街区情報に関してはゼンリン住宅地図 Zmap-TOWN のデータを用い，道路ネットワーク情報に関しては数値地図 2500 のデータを用いた．なお，両者は，国土調査法施行令で定められた 19 座標系の IX 系原点をもとに，重ねあわせることができる．

3.1 建築物の対応付け

図 6 に，建築物と地図との対応付けを行っている場合の画面を示す．左下のウィンドウは，リアルタイムで再生されている映像である．左上のウィンドウは撮影された周辺の地図を示す．

対応付けを行うユーザは，知っている建築物が映像に映っていた場合に，その建築物が映っている領域の中央付近をマウスなどにより指示する．次に，その建築物に対応する地図上での図形を選択することにより，データベースに地図上の位置と映像の対応関係を入力する．

なお，中央下のウィンドウは，ユーザが再生映像を指示した際のフレームと指示した座標を示すためのウィンドウである．これは，確認のために補助的に表示している．

図 6 対応付けを行う例
An example snapshot of linking images of buildings to map.

3.2 映像中の建築物に対する質問応答

建築物に関する質問応答を行っている画面の例を**図 7** に示す．これは，リアルタイムに再生されている映像中のある建築物に対して,「その建築物が何であるか」という質問を行う例を示している．このように映像中に含まれる映像オブジェクトに対して「質問」を出し，それに対する答を得る操作を質問応答と呼ぶことにする．

リアルタイムに再生されている左下の映像ウィンドウに質問したい建築物が映っている場合に，その建築物が映っている映像中の建築物領域内の 1 点を指示すると，前述した手法により，システムがその部分に映っている建築物と地図との対応関係を推定し，ユーザに対し答を提示する．実験用システムでは，ユーザに対する答として，地図データベースに含まれるその建築物の名称および対応した地図上の領域を提示する．

図 7 映像中の建築物に対する質問応答する場合の例
An example snapshot of question and answer about buildings.

3.3 建築物画像の検索

地図上で指示された特定の建築物が映っている画像を検索する場合の例を**図 8** に示す．

左上ウィンドウに表示されている地図上で，検索したい建築物を表す領域を指示，選択すると，その建築物が映っている画像が右下のウィンドウに表示される．対象としている実世界映像では，特定の建築物が映っているフレームは連続しているため，検索結果として提示するのはこの映像集合中の 1 フレームのみであり，時間的に集合の中央に位置するフレームとしている．なお,「ある建築物が映っているフレーム」とは,「建築物の幅の半分以上がフレーム内に映っている」あるいは「フレーム全体に建築物の一部が映っている」ものとした．

4. 実　験

4.1 実験方法

実験に用いた実世界映像は，道路に沿って車を移動させ，撮影車の進行方向に対してカメラを垂直に固定して，建築物が正面から映るように撮影したものである．毎秒 6 フレーム，全フレーム数は 280 枚で，道路に面している 18 軒の建築物が映っている．

このような実世界映像に対して，ユーザが 2 つの建築物を地図との対応付けを行った場合に，実装したシステムを用いて，それら対応付けられた建築物の間に存在する建築物に関して「質問応答」および「検索」の操作を行った．

ユーザにより地図と対応付けされた建築物の間に存在す

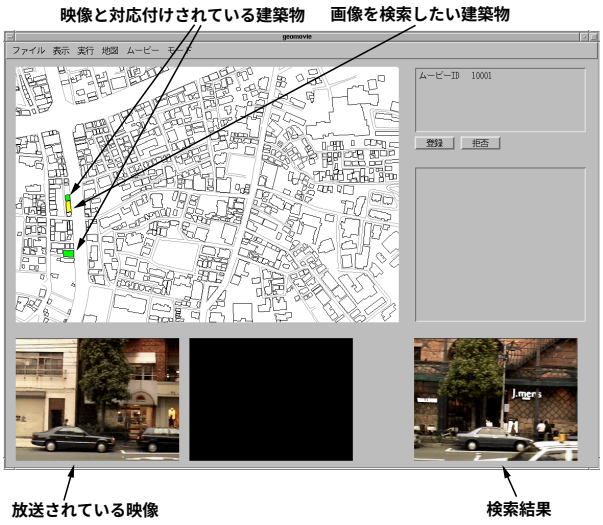


図 8 建築物の画像を検索する場合の例
An example snapshot of image retrieval.

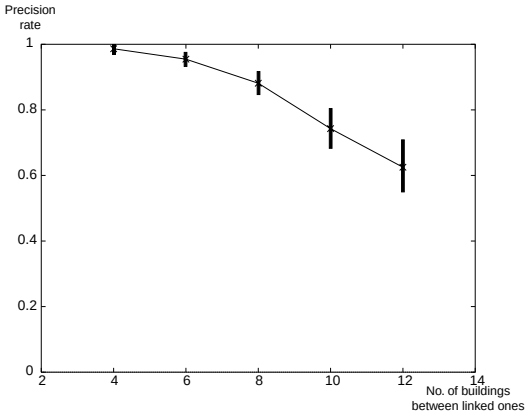


図 9 質問に対する正答率
Precision rate of question and answer.

る建築物の軒数を 4 軒，6 軒，8 軒，10 軒，12 軒と変化させ，それぞれ場合について，始点と終点となる建築物と地図との対応付けを行い，質問応答の操作を試行した．ただし，対応付けされた建築物の間に存在する建築物の軒数には対応付けされた建築物も含むものとする．

なお，実験結果の精度は，280 枚のすべてフレームから，建築物の境界を人手により検出した結果をもとに算出した．

4.2 結　果

以上のような条件のもとで,「映像に映っているこの建築物は何か」という質問に対して，自動的に対応付けされた建築物についての情報が正しく得られるかどうかを表す正答率を**図 9** に示す．ここでは，始点および終点に対応付けされた建築物の間に存在する建築物の数を変化させた場合の正答率の平均値およびその標準偏差 σ を重ねて示す．

この場合の質問応答の正答率とは，リアルタイムに流れる映像中で建築物が占める領域の中央付近を指して質問を行った場合に正しい答が得られる確率とする．また，ここで指し示す領域としては，建築物の中心から水平方向にそれぞれの建築物の幅 1/2 以内の領域であり，対応付けされている建築物以外の建築物を対象としている．

また，**表 1** 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．この場合の適合率とは，地図上で指示した建築物に対応した正しい画像が得られる確率を示す．表 1 の値は，これまでの実験と同様の試行をした場合の平均値と標準偏差を示している．この時，全ての試行に関して検索の正答率は 1.0 であった．

表 1 検索における適合率
Recall rate of retrieval.

建築物数	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

4.3 考　察

質問応答の正答率に関しては，対応付けされた 2 つの建築物の間に存在する建築物が少ないほど良い結果が得られる．したがって，建築物について対応付けが多数ある場合には，なるべく近くの建築物から推定する方が良いということがいえる．このことは，始点および終点では，必ず正しい対応付けがなされており，いずれかに近い程ずれが小さくなっているためである．10 軒程度ごとに建築物が対応付けされている場合には，本手法は充分有効であるといえる．

しかしながら，対応付けされた建築物の間に存在する建築物が増加した場合には，誤った答を得る確率が高くなる．この原因としては，主にモザイキング処理によるパノラマ画像の生成における精度の低さが挙げられる．対応付けを行う際の中心のずれを除けば，始点および終点では，対応付けが正確に行われているため，モザイキング処理の誤差は，特にパノラマ画像の中央部に生じることになる．

ここで，質問応答の正答率とモザイキングの精度との関係について検討する．**図 10** に示すように，生成されたパノラマ画像における i 番目 $(i = 1, \cdots, N)$ の建築物の中心に対応する水平方向の座標を $C_p(i)$，その建築物の幅を $W_p(i)$，同様に，地図から得られる境界パターンモデルをパノラマ画像に重なるように変換した場合の i 番目の建築物に対応する座標を $C_m(i)$ ，その建築物の幅を $W_m(i)$ とすると，各建築物に対して質問応答の操作を行った場合，正しい答が得られるのは，パノラマ画像と境界パターンモデルとの共通する部分を指示した場合である．なお，N は対応付けされた建築物間にある建築物の数を表す．

実験では，建築物の中心にそれぞれの建築物の幅の 1/2 以内の領域を指示するようにしたため，指示する領域と境界パターンモデルとの共通する部分 $L(i)$ は，(1) 式のように表せる．ただし，各建築物の中心における誤差 $|C_p(i) - C_m(i)|$ を $\Delta C(i)$ と表す．

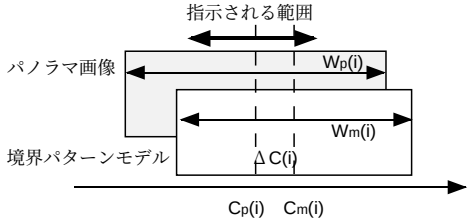


図 10 各建築物のパノラマ画像における座標の定義
Parameter definition of each buildings in panoramic image.

$$L(i)=\begin{cases}\frac{1}{2}W_p(i) & \left(\Delta C(i)\leqq\frac{1}{2}W_m(i)-\frac{1}{4}W_p(i)\right)\\ \frac{1}{2}W_m(i)+\frac{1}{4}W_p(i)-\Delta C(i) & \\ \quad\left(\frac{1}{2}W_m(i)-\frac{1}{4}W_p(i)<\Delta C(i)\right. & \\ \quad\left.<\frac{1}{2}W_m(i)+\frac{1}{4}W_p(i)\right) & \\ 0 & \left(\Delta C(i)\geqq\frac{1}{2}W_m(i)+\frac{1}{4}W_p(i)\right)\end{cases}\tag{1}$$

各建築物における正答率は,(1) 式で示される共通部分 $L(i)$ と指示される領域 $\frac{1}{2}W_p(i)$ との比で表される．つまり，パノラマ画像に対して，境界パターンモデルを重ね合わせた場合には，$W_p(i)\approx W_m(i)$ であるといえるので，各建築物における正答率は（2）式で表される．また，全体の正答率は各建築物に対する正答率の平均で表すことができる．

$$\begin{cases}1 & \left(\Delta C(i)\leqq\frac{1}{4}W_p(i)\right)\\ \frac{3}{2}-\frac{2\Delta C(i)}{W_p(i)} & \left(\frac{1}{4}W_p(i)<\Delta C(i)<\frac{3}{4}W_p(i)\right)\\ 0 & \left(\Delta C(i)\geqq\frac{3}{4}W_p(i)\right)\end{cases}\tag{2}$$

それぞれの建築物におけるモザイキング処理の誤差を表す関数 $f(x)$ は,(3) 式のように定義できる．ここで，x はパノラマ画像における各点の水平方向の座標とする．これにより，各建築物の中心における誤差だけではなく，それ以外の点に関しても隣接する建築物における誤差を利用して表すことができる．

$$f(x)=\begin{cases}\Delta C(i) & \\ \quad(x=C_p(i),\ i=1,\cdots,N) & \\ \left|\frac{x-C_p(i)}{C_p(i)-C_p(i-1)}\right|\Delta C(i-i) & \\ \quad+\left|\frac{x-C_p(i-1)}{C_p(i)-C_p(i-1)}\right|\Delta C(i) & \\ (C_p(i-1)<x<C_p(i),\ i=2,\cdots,N) & \end{cases}\tag{3}$$

一方で，パノラマ画像を生成する際に，始点と終点では正しい対応付けが行われているため,(4) 式,(5) 式が成り立つことがいえる．

$$\Delta C(1)=0\tag{4}$$

$$\Delta C(N)=0\tag{5}$$

図 11 に，対象とした映像において求めた誤差 $f(x)$ と，それぞれの建築物の幅 $W_p(i)$ を示す．横軸には，パノラマ画像における座標を取り，縦軸には誤差および建築物の幅を示す．なお，各建築物の幅 $W_p(i)$ を幅および高さに持つ棒グラフとして示す．ただし，グラフ上の棒グラフが存在しない部分は，建築物が存在しないことを示している．

（2）式で示したように，モザイキング処理における誤差 $f(x)$ が建築物の幅の 1/4 よりも小さい場合には，正答率が 1 であり，常に正しい答が得られる．一方で，建築物の幅の 3/4 よりも誤差 $f(x)$ が大きくなる場合には，指示する領域と境界パターンモデルとの共通する部分がないため，その建築物については正答率は 0 となり，常に誤った答となる．それ以外の場合には，各建築物における正答率は,(2) 式に示す通り $f(x)/W_p(i)$ の関数として表わされ，$W_p(i)$ に対して $f(x)$ が占める割合が大きいほど正答率が低くなることがいえる．

N が大きくなるにつれ，モザイキング処理における誤差が増加していくため，生成されるパノラマ画像の精度は低くなる．そのため，N が大きい場合には，全体の正答率も低くなり，特に中央部の建築物に対する正答率は低くなることになる．

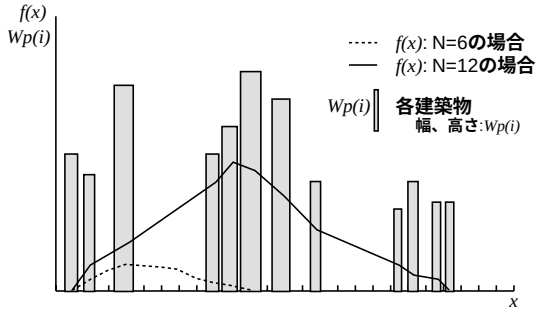


図 11 パノラマ画像における誤差
Error on panaramic image.

一方，地図上で指示された建築物画像を検索する場合の適合率が非常に高いのは，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中の中心となるな 1 枚のみを検索結果として提示しているためである．つまり，**図 12** に示すように，検索対象の建築物が実際に映っている n フレームの映像集合と指示された建築物が映っていると推定される m フレームの映像集合が完全に一致しなくても検索結果としては正しいことを示している．

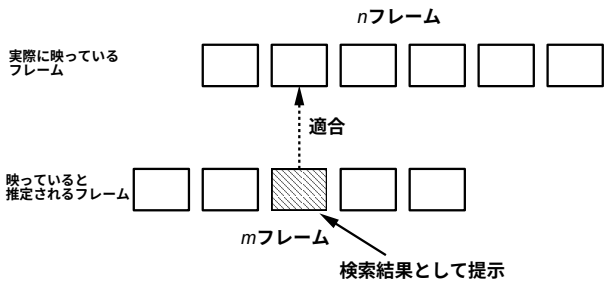


図 12 提示する検索結果
Present a result of retrieval.

5. 境界抽出による補正手法

これまで提案している手法は，一般的なブロックマッチング法による映像モザイキング処理の結果に基づくものであるため，すべての場合において実世界映像と地図を正しくマッチングできるわけではない．そこで，対象が建築物であることを利用して，映像と地図とのマッチングをより精密に行う手法を検討する．以下のような手順により，映像中から建築物の境界と推定されるエッジを抽出し，これまで述べたようにパノラマ画像と境界パターンモデルとを重ね合わせて得られる建築物の境界を補正する手法について述べる．

（1） 各フレームにおいて垂直エッジを検出する．
（2） 余計なエッジの除去およびグループ化により建築物の境界と考えられる垂直エッジを抽出する．
（3） パノラマ画像へ重ね合わせることにより建築物の境界を推定する．

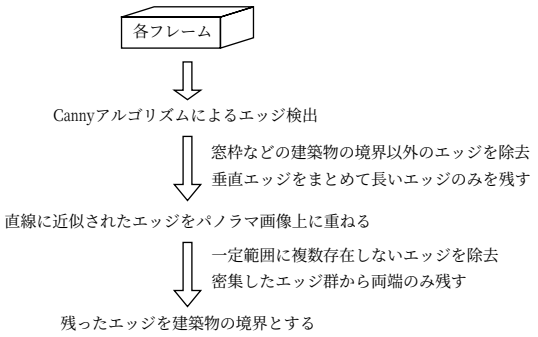


図 13 建築物の境界抽出の流れ
Flowchart of boundary extraction.

5.1 エッジ検出による境界抽出

実世界映像中では，建築物の境界は強い垂直エッジとなる．したがって，実世界映像から垂直なエッジを抽出することにより建築物の境界を検出することができる．ここでは，代表的なエッジ検出法である Canny アルゴリズムによって，エッジの向きが垂直から両側に $\frac{\pi}{16}$ の幅の間にあるものを垂直エッジとして検出する．

実世界映像から検出された垂直エッジの中には，窓枠などの建築物の境界ではないエッジが多く含まれる．まず，このようなエッジを削除する必要がある．**図 14** に示すように，強い水平エッジが存在する水平方向の区間に存在する垂直エッジは建築物の境界ではないことが多いという建築物の画像の特徴を利用し，建築物の境界以外のエッジを削除する．

図 14 のような特徴を持つことを仮定すると，ある長さ（ここでは，便宜的に画像幅の約 1/10 とした）以上の水平エッジが存在する領域に含まれる垂直エッジは，ドアや窓の枠ということになるため，このような建築物の境界以外と考えられる垂直エッジを削除する．この処理を適応した例として，**図 15** に示す．これにより，建築物の境界以外

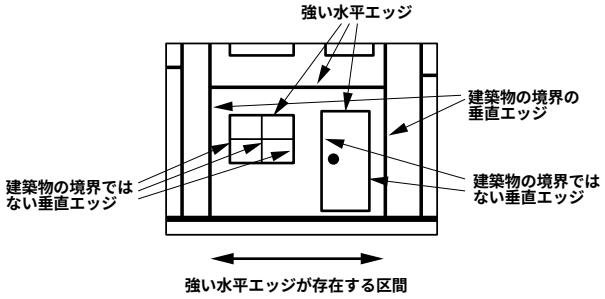


図 14 建築物の境界ではない垂直エッジ
Vertical edges of boundary except buildings.

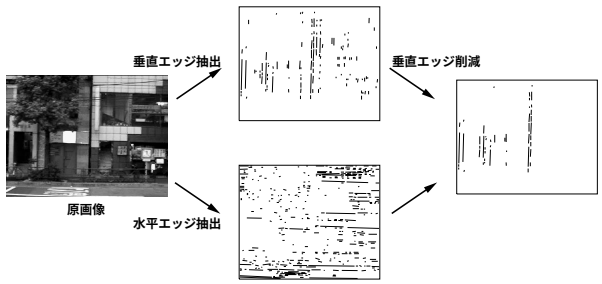


図 15 窓枠などのエッジの削除
Eliminate edges of windows frame.

図 16 直線近似によるグループ化
Grouping of edges by liner.

のエッジをある程度削除できることがわかる．

ここまでの処理により，建築物の境界以外の窓枠などの垂直エッジの多くは削除されていると考えられるが，残っている垂直エッジの中から境界であると推定されるエッジのみを抽出する必要がある．**図 16** に示すように，複数の垂直エッジから境界となるエッジにまとめる．建築物の境界となるエッジはある程度の長さを持つこと，多数の境界が隣接することはないという特徴を仮定し，長いエッジとなるように隣接する垂直エッジをグループ化を行うことにより，垂直な直線に近似する．ただし，ここでグループ化されたエッジの長さが短い場合には，境界ではないとして削除する．

しかしながら，ここで抽出された直線には建築物の境界ではない垂直エッジが含まれる．そこで，モザイキング処

理で求めた移動量を用いて，各フレームごとの処理で抽出された垂直エッジをパノラマ画像上に重ね合わせると，同一と考えられる垂直エッジは，パノラマ画像上ではほぼ同一の場所になるため，それらを 1 つの垂直エッジとしてまとめることにより余分な垂直エッジを削減することができる．

このようにパノラマ画像上に重ね合わせたエッジにも，依然として建築物の境界でないものが含まれている．手順（2）により窓枠などの不要なエッジとして削除された垂直エッジが，他のフレームでは映っている範囲が異なるため，削除されない場合がある．そのため，パノラマ画像にそのまま重ね合わせるのではなく，ある範囲に複数の垂直エッジがない場合には，ノイズとして削除する．

また，**図 17** のように建築物と建築物の隙間は，垂直エッジが密集している場合が多く，そのエッジ群の両端のみが，建築物の境界となる場合が多い．これは，建築物と建築物の隙間には長い水平エッジが現れないため，処理（2）では垂直エッジを削除できず，ノイズによるエッジとして残ってしまう．したがって，垂直エッジが密集している場所に関しては，密集した部分におけるエッジのうち両端のみ残して，内側の垂直エッジを削除することにする．

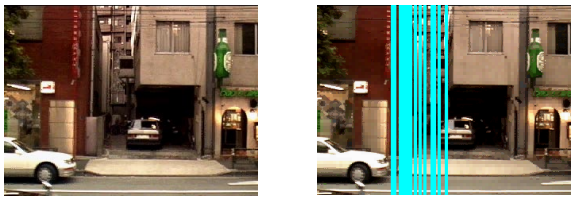


図 17　建築物の隙間における垂直エッジ
Vertical edges between buildings.

5.2　推定される境界の補正

以上の手順により境界として抽出された垂直エッジをパノラマ画像に重ね合わせると，検出されたエッジは建築物の境界全体の約 58%と正しく対応付けられる．そこで，境界パターンモデルとの対応付けから推定された境界を，それからある一定の範囲内の距離にある最も近いエッジと対応付けることにより補正する．ただし，この手法では境界として抽出できないエッジが存在するが，その場合には境界パターンモデルから得られる境界から一定の距離の範囲内にはエッジが存在しないため，補正処理は行われない．一方，誤って境界として抽出されているエッジもあり，それらの約 30%が境界として対応付けされるため，誤った補正処理となっている．

前述と同様の方式により質問応答の実験を行い，このような補正処理を行った場合と補正処理しない場合における正答率を比較したものを**図 18** に示す．

建築物の数が 4 軒の場合には，正答率の向上が見られなかったが，これについてはすでにほぼ 100%正答率が得られているため，本来補正する必要がないといえる．また，

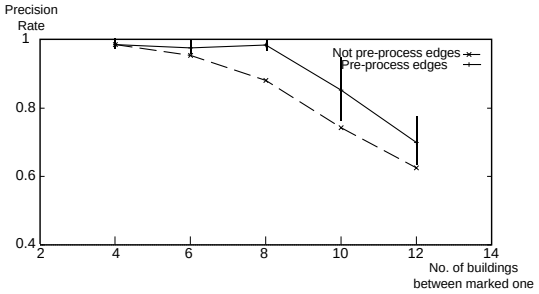


図 18　質問応答正答率（補正前と補正後の比較）
Precision rate of question and answer (original method vs under correction).

そのほかの場合においては正答率の向上が見られ，この手法の有効性が示されている．しかしながら，この手法においても，対応付けされた建築物の間に存在する建築物の数が 10 軒以上の場合における正答率が急激に低下する傾向が現れている．これは，モザイキング処理における精度が低いため，検出されたエッジがモデルから推定されている境界から離れた位置になり，一定の範囲内に入らない場合にはエッジが検出されていないとしたことにより，補正されないためである．特に，境界パターンモデルとの誤差が大きくなる中央付近の建築物では，正しくエッジが抽出されている場合でも補正されないことがあり，そのため正答率がそれほど向上していない．

6.　む　す　び

放送される映像とそれに関連する情報のデータベースとをリンクすることにより，映像に対する質問応答や情報検索などの操作を可能とする対話的な映像情報システム ADTV（Advanced Database TV）について述べた．ADTV システムにおいて，インタラクティブな操作を行えるようにするためには，映像に付加されている情報を共有するだけでは充分ではなく，映像を自動的に構造化する必要があることを述べた．

また，ADTV システムの一例として，実世界映像を対象に，建築物の映像と地図とを自動的に対応付ける手法を提案し，その手法の実験・評価を行った．ADTV の機能として想定されている，映像に対する質問応答や対応付けされた建築物の画像を検索するという操作が充分な精度で行うことができ，手法の有効性が確認された．さらに，対象とした実世界映像の特徴を利用して，対応付けの補正手法について述べ，評価を行った．その結果，質問応答の操作に関する精度の向上が見られた．

今後，モザイキング処理の高精度化，個々の建築物に対する境界パターンと映像との対応付け手法の高度化および ADTV システムの拡張に取り組む予定である．

なお，本研究の一部は，文部省科学研究費補助金（創成的基礎研究費・課題番号 09NP1401）による．

〔文　献〕

1）谷田部, 大場, 坂内：“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, PRMU97–64, IE97–33, MVE97–49, pp. 63–68 (1997)

2）T. Yatabe, H. Kawasaki, and M. Sakauchi：“Interactive Video Description on the Network.”, Proc. of IEEE Multimedia Computing and Systems ’99, vol. 2, pp. 194–198 (1999)

3）上田, 宮武, 吉澤：“認識技術を応用した対話型映像編集方式の提案”, 信学論 (D-II), **J76-D-II**, 2, pp. 216–225 (1992)

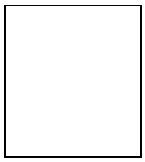
4）柴田：“映像の内容記述モデルとその映像構造化への応用”, 信学論 (D-II), **J78-D-II**, 5, pp. 754–764 (1995)

5）E. Oomoto and K. Tanaka: “OVID: Design and Implementation of a Video-Object Databese System”, IEEE Transactions on Knowledge and Data Engineering, **5**, 4, pp. 629–643 (1993)

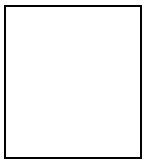
6）R. Szeliski: “Video Mosaics for Virtual Envrionment”, IEEE Computer Graphics and Applications, pp. 22–30 (1996)

7）富井, 有澤：“マルチメディアデータベースにおける映像モデリングと操作言語”, 信学論 (D-II), **J79-D-II**, 4, pp. 520–530 (1996)

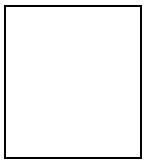
8）S. Pradhan, K. Tajima and K. Tanaka: “Managing Multimedia Objects in Invremental Instance-based Object Database Systems”, Proc. of IPSJ Multimedia Japan’96, Yokohama, pp. 202–209 (1996)



谷田部智之（やたべともゆき） 1995 年，東京大学工学部電子情報工学科卒業．1997 年，同大学院修士課程修了．現在，同大学院博士課程在学中．主として映像メディア処理およびネットワークアプリケーションに関する研究に従事．



川崎　洋（かわさきひろし） 1994 年，京都大学工学部電気電子工学科卒業．現在，東京大学大学院修士課程在学中．主として映像メディア処理に関する研究に従事．



坂内　正夫（さかうちまさお） 1975 年，東京大学大学院工学系研究科博士課程修了．同年同大学工学部電気工学科専任講師，その後，横浜国立大学工学部情報工学科助教授，東京大学生産技術研究所助教授を経て，現在，同大学生産技術研究所教授．1998 年より同大学生産技術研究所所長．マルチメディアデータベースなどの研究に従事．工学博士．正会員．