2003－DBS－131　(55)
2003／7／17

3 次元地図を用いたビデオコンテンツの自動索引法
－被写体建物オブジェクトの自動抽出－

石黒 玲† 佐藤 有紀子† 増永 良文‡

†お茶の水女子大学大学院人間文化研究科数理・情報科学専攻 〒112-8610 東京都文京区大塚 2-1-1
‡お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚 2-1-1

E-mail: †{ray, yukiko}@dblab.is.ocha.ac.jp

‡masunaga@is.ocha.ac.jp

あらまし 本論文ではビデオカメラによって取得された映像に対して，写し込まれているであろう建物オブジェクトを自動抽出し映像の索引として付与し，それを使って映像の内容検索を実現する索引法を提案する．撮影者はウェアラブルコンピュータを身に付け GPS・ジャイロセンサを装着してビデオ撮影を行う．GPS によって取得された時刻データと位置データ，カメラに装着されたジャイロセンサから得られる姿勢データに加えて，データベースに格納されている 3 次元地図により撮影されているべき建物オブジェクトを計算により自動抽出し，索引とする．写っている建物オブジェクトが変わるごとに映像ストリームをユニットとし分割していく．これにより例えば「銀座の三越デパートの夕暮れ時の映像」をデータベースから検索することが可能となる．

キーワード 3 次元地図，GPS，ジャイロセンサ，建物オブジェクト，自動索引付け

An Automatic Indexing Method of Video Contents using 3-Dimentional Geographic Data
－Automatic Extraction of Buildings for Indexing－

Rei ISHIGURO† Yukiko SATO† and Yoshifumi MASUNAGA‡

†Graduate School of Humanities and Sciences, Ochanomizu University

2-1-1 Otsuka, Bunkyou-ku, Tokyo, 112-8610 Japan
‡Department of Information Science, Faculty of Science Ochanomizu University

2-1-1Otsuka, Bunkyou-ku, Tokyo, 112-8610 Japan
E-mail: †{ray, yukiko,}@dblab.is.ocha.ac.jp

‡masunaga@is.ocha.ac.jp

Abstract This paper proposes an automatic extraction method of buildings shot by a video camera for indexing purpose so that users can retrieve video data by content-based specification. Video are taken by a camera with an wearable computer attached by a GPS and a gyro-sensor. A set of shot buildings are automatically extracted using a 3-dimentional geographic data stored in the wearable computer. Video stream are divided into “unit” where the set of shot building does not change. By this means, the system makes it possible for user to retrieve units whose contents are, for example, “Ginza Mitsukoshi Department Store in the evening darkness” from video database.

Keyword 3-Dimentional Geographic Data, GPS, Gyro-sensor, Buildings, Automatic Indexing

1. はじめに

近年ビデオカメラの小型化やバッテリの長寿命化が進み，人が手持ちで長時間移動しながら多数のビデオを撮影することが可能になっている．それに伴い，莫大な映像データが取得されるようになったが，一方その大量の映像データから所望のシーンを映像内容に基づいて的確に検索する技術は確立されていない．

これまで様々な映像の索引付け手法が研究されてきている．有澤[1]らは画像処理の技術によるカットの自動検出による自動索引付けの手法を，堀内ら[2]は映像情報に付随する開始時刻，終了時刻の情報を場面単位でデータベースに格納している．また井出ら[3]はニュース映像に対する映像中の画像情報及び言語情報を総合的に利用して映像への自動索引付け手法を提案している．Healy ら[4]は発汗という生体情報をセンサで捉え，それを利用して映像イベントのハイライトを作成している．飯島ら[5]はパターン認識によって映像中に現れる人物を検出し，それをもとにエピソード映像を作成する研究を行った．Laerhoven ら[6]は多数の単純なセンサを用いることで小さく，安く，高価な工程なしで異なった状況（walking, running, sitting など）を区別し，1 日の行動日誌を作成している．近年上田ら[7][8]は映像に写っている可能性のある空間オブジェクトを地図データを使って自動的に索引として付与する試みを行ったが，何が写しこまれているかという計算を行わないので，実際に写っているオブジェクトと索引が一致しない場合が生じる．

そこで本研究では GPS（Global Positioning System）とジャイロセンサを用いて得られるビデオカメラの時刻位置・姿勢データに加え，3 次元地図データを用いることにより映像に写しこまれているであろう建物オブジェクトを計算して自動的に索引付けを行い，それらに基づいた内容検索を可能にするビデオデータベースシステムを構築することを目的とする．これによりユーザは例えば「銀座の三越デパートの夕暮れ時の映像が欲しい」という要求に対し，夕暮れとは天文学によれば「日没後太陽の中心が地平線下 7 度 21 分 40 秒の角度にある時刻」と定義されていることと銀座三越デパートという建物オブジェクトをキーワードとして検索を行うことで検索が可能となる．

2. 映像データベースシステムの概略

2.1. システム構成

必要データ取得のために GPS とジャイロセンサを使用する．GPS により撮影者の位置情報・時刻情報を取得することができる．またジャイロセンサによりカメラの姿勢情報を取得することができる．撮影者はウェアラブルコンピュータを身に付け GPS データ，ジャイロセンサデータ，カメラデータ（レンズの画角：今回使用するビデオカメラの画角は横 44°，縦 35°である），3 次元地図データ，人文地理データを総合的に処理し，映像データに内容に基づいた索引付けを行う．

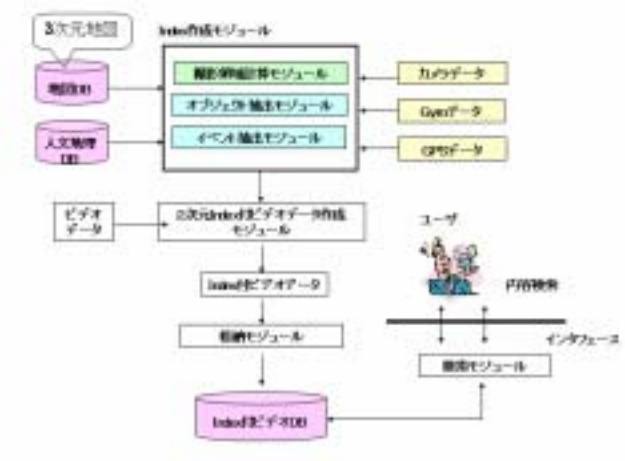

図 1：システム概要
Fig1. A System Overview

図 1 は我々が開発しようとしているシステムの全体像を表している．（ただし，人文地理データの利用によるイベントの自動索引付け機能は現在のところまだ研究を行っていないので詳細は省く．）

本システムを実現するには大別すると 3 つのアプローチが考えられる．①ウェアラブルコンピュータを使用し，撮影と同時に索引付き映像データベースを作成していく．②携帯電話等を使用し，GPS とジャイロセンサのデータのみ PPP（Point-to-Point Protocol）接続によりネットワークの TCP/IP ソケットを介しサーバに送信し，サーバにおいて撮影領域の計算・建物オブジェクト抽出を行い，索引を作成する．その後撮影者が戻ってから，索引付き映像データベースを作成する．③完全にオフライン作業で行う．それぞれの手法のメリット・デメリットを表 1 に示す．いずれの手法においてもパソコンは必要になることや，河内ら[9]の実験での報告によると，都市部では携帯電話による PPP 接続が途切れることもあるので，本システムの実現法としては①ウェアラブルコンピュータによるリアルタイム作業を最終的には考えている．ウェアラブルコンピュータを使用することで，「Hands　Free」が実現され，システムを気にすることなく活動し，撮影を続けることができる．またリアルタイムに索引付き映像データベースを作成していくことで，撮影に出掛けた先で検索システムを利用することも出来る．デメリットとして挙げたメモリ･HDD の容量に関しては，ザイブナー社は 128MB RAM，5GB HDD のウェアラブルコンピュータを提供している．映像データを MPEG4 で圧縮した場合 10 時間の映像データは約 36MB～1.3GB となり，被写体建物オブジェクト抽出アルゴリズムに必要な地理情報システム（ESRI Japan 社 ArcView 8.2）の必要 HD 容量が 700MB なので 10 時間程度の映像を撮影する場合を想定すると問題にはならないと考えられる．

	メリット	デメリット
①ウェアラブルコンピュータによるリアルタイム作業	・その場で索引付き映像データベースが生成 ・出掛け先で検索が可能 ・自由度が高い	・地図が多数必要（標準各分） ・メモリ･HDD の容量の問題
②携帯電話による PPP 接続使用	・地図はサーバだけにあればよい ・戻った時点で索引は出来ている	・通信の信頼性 ・通信料
③オフライン作業	・地図は処理を行う PC のみ	・戻った時点では好きに出来ていない（作業が行える場所に戻る必要がある）

表 1：実現手法の比較
Table 1. Comparison of Realization Techniques

2.2. 地図データ（2 次元･3 次元）

3 次元地図は三菱商事株式会社製の DiaMap（レベル 2）を使用する．これは道路面などの 2 次元地図情報に加え，ビルの高さなどの 3 次元地図情報を収録する 3 次元地図データである．高さデータとしては建物データ（基本データ），高速・高架鉄道敷，駅プラットフォーム，一般道路橋があるが，建物データ以外のものはオプションデータとなっており，今回使用する高さデータは建物データのみを使用している．

建物は複数面で構成されるブロック（多面体）としており以下のもので構成されている．

- 屋根面
- 側面
- 屋上形状
- 建物全体

ファイルフォーマットは DXFR14 形式である．建物の属性データについては 2 次元地図である国土地理院の数値地図 2500（空間データ基盤）を使用しているが，これは収録属性値が公共建物のみとなっており，今後はより詳しい属性値を収録するゼンリン社製の 2 次元地図 Zmap-TOWN Ⅱを使用する予定である．これは公共建物に限らず属性値は地図上の建物について，その形状 ID，建物名，住所コード，地番等を提供している．

2.3. 3 次元地図を用いた映像データへの索引付与

GPS により取得したビデオカメラの位置データと，ジャイロセンサにより取得したビデオカメラの姿勢データ，それにビデオカメラのレンズの画角を考慮し，3 次元地図 DB を使って映像の各フレームに写っているべき建物オブ

ジェクト（三越デパート，ソニービル等の地理情報）集合の自動抽出を行う．使用する 3 次元地図は Point オブジェクト P_1,・・・P_1, Polyline オブジェクト L_1,・・・L_m，Polygon オブジェクト S_1,・・・S_nによって構成されており，建物の高さ情報は Polygon オブジェクトの属性値として定義されている．Polygon オブジェクトが索引に適したオブジェクトと考えられ，地図 M は次のように表される．

$M=\{S_1,\cdots S_n\}$

2^MでMの巾集合を表すことにすると，映像の各フレームには，そこに写っているべき建物オブジェクト集合，これらは一般的に 2^Mの元，により索引付けられる．

さらに，映像を空間オブジェクト集合が変化するごとに“ユニット”として分割して索引付け格納する．

3. 被写体建物オブジェクトの自動抽出アルゴリズム

3.1. 基本的な考え方

本章ではビデオカメラで撮影されているべき 3 次元建物オブジェクトの抽出アルゴリズムを考える．GPS により ArcView 8.2 の拡張機能である Tracking Analyst を介しログファイルにビデオカメラによる撮影の時刻，緯度，経度のデータを取得する．ログファイルを ArcView 8.2 においてシェイプファイルに変換し，属性テーブルを生成する．また同時にジャイロセンサからは時刻，ヨー角，ロール角のデータを取得する．

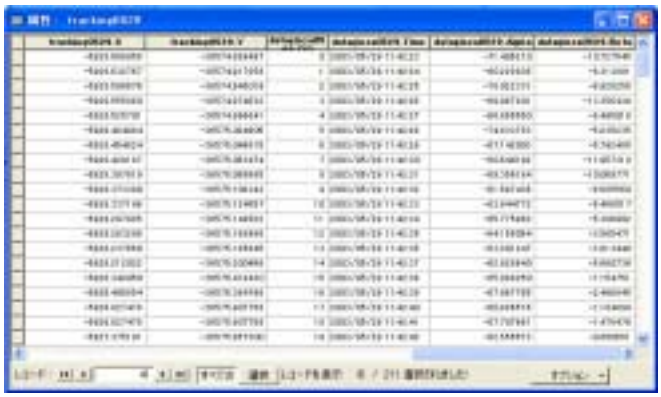

図 2：GPS と Gyro によるデータのテーブル表現
Fig2. Table Representation of GPS and Gyro Data

図 2 はこれら 2 種類のデータを時刻で同期させて得られたテーブルを表す．ただし GPS により取得した緯度・経度データは日本測地系・横メルカトル図法・平面直行座標系（XY 表記）に変換した値である．

ここで，人間が映像として認識できる最低の時間は 3 秒ということ[10]を考慮に入れ，すべてのデータは 1 秒ごとに取得していく．（人間は 3 秒間で向いている方向をかなり動かすことが出来ることを考え，3 秒ではなく 1 秒ごとにデータを取得することにした．）

これらのデータと数値地図 2500（国土地理院）の属性データ（建物の名称：お茶の水女子大学等）用い， ArcView 8.2 により時刻 t における 2 次元平面地図で写っているべき空間オブジェクトの抽出を行う．図 3 の 4 角錐の部分がビデオカメラで撮影される空間であり，斜線の部分は 3 次元で写されている空間を 2 次元平面（X-Y 平面）に投影したことにより求まる．今回は使用しないが将来ビデオカメラのズーム機能を使用することを考慮して，ジャイロセンサのヨー角（α）は横の画角の中心，ロール角（β）は縦の画角の中心からの角度と設定する．

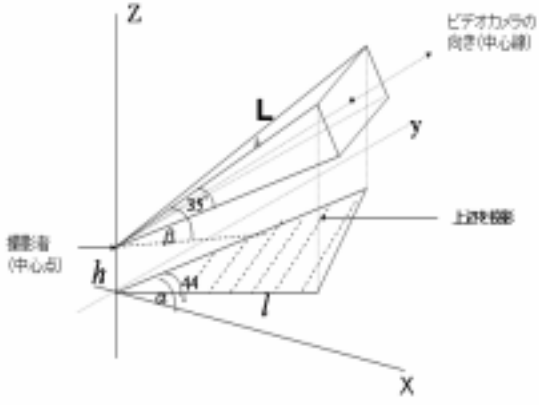

図 3：3 次元から 2 次元への投影
Fig3. Projection from 3-Dimention to 2-Dimention

L とは視野の距離であるが，ビル街等建物の密集している地域では L が小さく，野原など見晴らしの良い地域では L を大とする必要がある．この値は実験結果によって最適値を求める．これは後の建物オブジェクト抽出アルゴリズムにおいて被写体オブジェクトが取り残すことのない値である．

使用するビデオカメラの横の画角 44°，2 次元上への投影視野 *l* は， *l* =cos(β+17．5)＊L で求まることにより，オブジェクトはｙ＝sin（α+22）＊ *l*，y＝sin（α-22)＊ *l* で囲まれた 3 角形中にあることがわかる．ここで時刻，緯度，経度，オブジェクト名の属性からなるテーブルがでる．

次に，図 4 に示すように 3 次元的な視点により 3 次元地図 DiaMap からオブジェクトの高さ情報を加え，実際に写っているオブジェクトを抽出する．ビデオカメラの縦の画角（35°），撮影者の身長と高さ情報（ｈ）により，Z=sin（β+17．5）＊L＋ｈ，の範囲内にあるオブジェクトが，実際に写っている空間オブジェクトとなる．

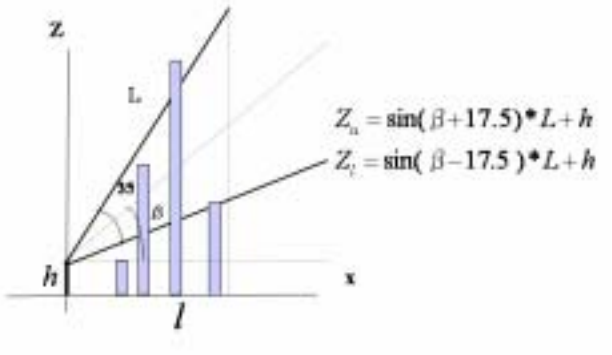

図 4：建物オブジェクトの抽出
Fig4. Extraction of Building Objects

3.2. 抽出アルゴリズム

建物オブジェクトは以下の属性を持つ．

- 建物オブジェクトの重心座標データ(x1, y1)
  （撮影地点から建物オブジェクトへの距離の比較時のみ重心座標データを用い，建物オブジェクトを選択する際には建物オブジェクトの幅も考慮に入れて計算する．）
- 高さデータ　height
- 撮影地点から建物オブジェクトへの距離 R
- 建物オブジェクトが撮影地点から見えている･見えていないのフラグ visible(初期値は on)
- カウンター　int　I

3.2.1. 2 次元地図上での候補建物オブジェクトの抽出

はじめに，2 次元地図上（DiaMap の 2 次元地図のみを使用）で建物オブジェクトを抽出する．使用するビデオカメラの横の画角（44°），2 次元地図上への投影視野長 *l* は *l* ＝cos（β＋17.5°）*L で求まることにより，オブジェクトはｙ＝sin（α±22°）＊ *l* で囲まれた三角形内にある．この三角形内でｙ＝sin（α－22°＋I）＊ *l*（0≦I≦44）として直線を 1 度ずつ動かしていき，直線と交差する建物オブジェクトを各 I に対して配列ａ[ｉ][ ]にＲの昇順に 2 次元地図における建物オブジェクトを格納する．

3.2.2. 高さ処理に基づく被写体オブジェクトの抽出

次に，3 次元的な視点により，DiaMap からオブジェクトの高さデータを加え，実際に写っているべき建物オブジェクトを抽出する．ビデオカメラの縦の画角（35°），撮影者の身長（ｈ）によりＺ＝sin（β±17.5°）*L＋ｈの範囲内にあるオブジェクトを抽出していく．配列ａ[ｉ][ ]で配列 2（[ｉ][1]）のオブジェクトＢの高さ（height）を配列 1（[ｉ][0]）のオブジェクト A と比較し，高ければそのまま，低ければフラグ（visible）を降ろす．次に配列 3（[ｉ][2]）のオブジェクト C と，フラグが立っているオブジェクトと高さをくらべフラグを決める．このように配列の先頭から高さを比べていき，最終的に visible のフラグが立っているすべてのオブジェクトを抽出する．図 5 の大きい楕円で囲まれているある I=i の y 値：ｙ＝sin（α－22°＋i）＊ *l* での場合を考えてみる．

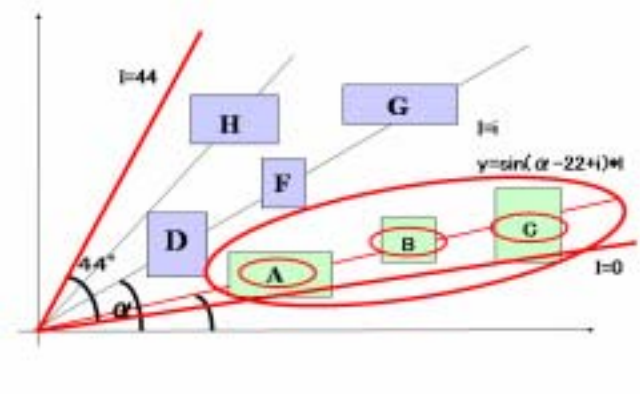

図 5：抽出アルゴリズム
Fig5. Extraction Algorithm

1. 撮影地点からの距離 R が小さいものから順番に配列に格納する．この場合 A・B・C の順番で格納する．
2. R の 1 番小さい A と 2 番目に R の小さい B と高さを比較し A より B のほうが高ければ B のフラグはそのまま，低ければ B のフラグを下げる．
3. その次に R の小さい C とフラグのたっているオブジェクトと高さを比較しフラグを決定する．

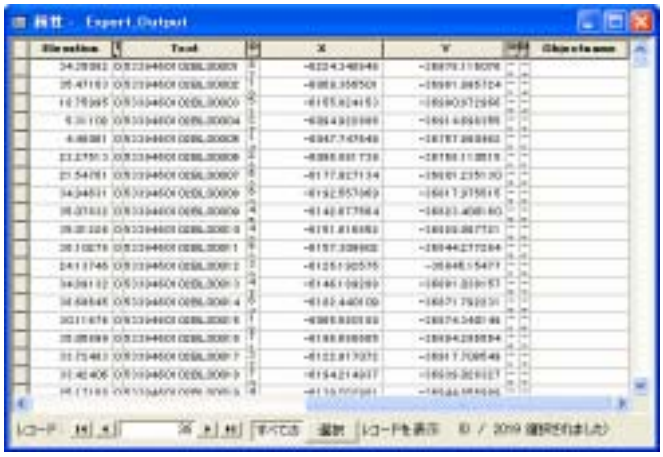

図 6：時刻 t における被写体建物オブジェクト
Fig6. Shot Buildings at Time t

これをすべてのカウント i の値で調べ，時刻 t における写っているべき建物オブジェクト群のテーブルとする．（図 6）

```
For(i=0;i≦44;i++) {
//put all objects on the line y=sin(α-22+i)
Array[i][]
//put order from small R
   a[i][0];
//put the smallest R, Object A
A=a[i][0];
//put the second smallest R, Object B
B=a[i][1];
if(A.height > B.height)
B.height = 0
//put third smallest R, Object C
For(int m=2; m<M; m++)
//find out object whose visible flag is 1
//largest  so far, Object P
if(P.height > C.height)
C.height = 0;}}
```

図 7：被写体建物オブジェクト抽出アルゴリズムの概略
Fig7. An Outline of Extraction Algorithm of Shot　Buildings

図 7 は上記の被写体建物オブジェクトの抽出アルゴリズムである．

このアルゴリズムでは隣接する配列に登録されない建物オブジェクトが存在する可能性が理論上ある．しかし，例えば視野長 L=100m とすると 1 度間隔で配列を作成しているので 100m 先の隣接する配列間の距離は 2πL／360 1.7m となる．したがって実際には後述する過少誤認の原因にはならない．

4. 被写体建物オブジェクト抽出実験

提案手法のうち，建物オブジェクトの抽出に関する実験を行っている．

4.1. 実験概略

本実験は概略以下のように行っている．

1. ウェアラブルコンピュータと GPS・ジャイロセンサを装着し銀座を移動し，ビデオカメラによって家並みを撮影する．
2. 撮影位置・ビデオカメラの姿勢情報を取得する．
3. 撮影終了後，すべてのデータから建物オブジェクトを抽出する．
4. 抽出された建物オブジェクトと実際の映像との検証を行う．

ウェアラブルコンピュータには Sony 社製のノートパソコンを使用した． 図 8 は被写体建物オブジェクトの抽出ウィンドウである．濃く強調表示されている部分が被写体建物オブジェクト抽出アルゴリズムによって抽出された建物オブジェクトとなっている．

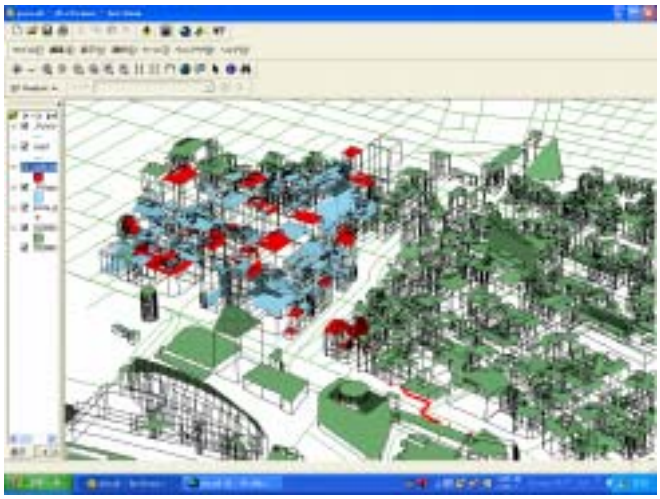

図 8：被写体建物オブジェクトの抽出ウィンドウ
Fig.8 Extraction Window of Shot Buildings

建物オブジェクト抽出に関して以下のような問題点が挙げられる.

- 過少誤認（false dismissal）：実際には写しこまれているのに被写体建物オブジェクトとして抽出されない場合
- 過多誤認（false alarm）：実際には写しこまれていないのに被写体建物オブジェクトとして抽出されてしまう場合

現在このような事象が発生しているかどうか，もし発生していたとすればどのような原因であったのかを検討中である.

5. まとめと今後の課題

これまでにシステムの全体像の構想提案とGPSデータ・ジャイロセンサデータ・ビデオカメラの画角データを用いた撮影領域計算を行い，2次元・3次元地図データを用い写っているべき建物オブジェクトの抽出アルゴリズムを提案した．今後の課題として以下のようなものが挙げられる.

- 視野長Lの最適値
- 効率の良い建物オブジェクトの抽出方法の考察
- Index付ビデオデータ作成モジュールの構築
- 重要度の考慮
- 街路樹などの障害物に対する対処
- 富士山・東京タワーなどの特別な建物オブジェクトの扱い（これは視野長Lを大きくすれば建物オブジェクト抽出の計算量が増えるため単純にLを大きくすれば良い問題ではない．このような特別な建物オブジェクトにかんしては3次元地図とは別にデータベースを独自に作成する必要があると考えている.）
- ユニットのタップル数によるデータベースサイズの効率の良い検索方法
- リアルタイム化の拡張
- 人文地理データ（花火大会や盆踊りなど）の索引としての利用

6. 謝辞

本研究は，文部科学省科学研究費補助金萌芽研究「ウェアラブルデータベースとその可能性」（平成14・15年度）および科学技術振興事業団（JST)の戦略的基礎研究推進事業(CREST)「高度メディア社会の生活情報技術」プログラムの援助を受けている．ここに記して謝意を表す.

文　献

[1] 有澤博，由井仁，富井尚志：“映像データベースシステムの構成の一方式”, Proceedings of Advanced Database Symposium ‘93，pp. 181-190，Dec. 1993

[2] 堀内優希，友田政明，石川佳治，植村俊亮：“映像データベースシステムのための論理データモデルとその実装，電子情報処理学会第6回データベース工学ワークショップ（DEWS)，pp. 79-86，Mar. 1995

[3] 井出一郎，山本晃司，浜田玲子，田中英彦：“ショット分類に基づく映像への自動索引付け手法”，電子情報通信学会論文誌，Vol. J82-D-Ⅱ，No.10，pp. 1543-1551，Oct. 1999

[4] Jennifer Healey and Rosalind Picard：“Startlecam：A cybernetic wearable camera”, International Symposium on Wearable Computers（ISWC)，pp42-49，1998

[5] 飯島俊匡，石上陽一，川嶋稔夫，青木由直：“日常生活映像から検出された人物像によるエピソード想起”，パターン認識・メディア理解研究会（PRMU)，1998

[6] Kristof Van Laerhoven，Kofi A.Aidoo and Steven Lowette：“Real-time Analysis of Data from Many Sensors with Neural Networks”, International Symposium on Wearable Computers（ISWC), IEEE Press, pp115-122, 2001

[7] 上田隆正，天笠俊之，吉川正俊，植村俊亮：“位置情報と時刻情報を用いた映像データの索引付け手法”，電子情報通信学会第12回データ工学ワークショップ（DEWS2001)，2001.3

[8] 上田隆正，天笠俊之，吉川正俊，植村俊亮："位置情報と地理情報を用いたウェアラブルカメラ映像のダイジェスト作成"，情報処理学会第125回データベースシステム研究会報告，No.70，2001.7

[9] 河内聡恵，服部麻衣子：“GPSを用いた移動体データベースシステムの構築”，お茶の水女子大学卒業論文，2001

[10] HonJian Zhang，Chien Yong Low，Stephen W.Smoilar and JianHua Wu，“Video Parsing，Retrival and Browsing：An Content-Based Solution” Intelligent Multimedia Information Retrieval, ed. Mark T. Maybury, pp.139-158, MIT Press，Massachusetts，1997