3 次元地図を用いたディジタルビデオコンテンツの自動索引法の提案と検証

A Proposition and Verification of an Automatic Indexing Method for Digital Video Contents using a 3-Dimentional Map

佐藤 有紀子♥ 石黒 玲♦
増永 良文♠

Yukiko SATO Rei ISHIGURO
Yoshifumi MASUNAGA

本論文では，ディジタルビデオカメラによって撮影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計算により自動抽出して，映像の索引として付与するという新しいビデオコンテンツの自動索引法を提案し，その有効性を検証する．撮影者はウェアラブルコンピュータを身に付け GPS とジャイロを装着して街頭でビデオ撮影を行う．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から得られるビデオカメラの姿勢データに加えて，データベースに格納されている 3 次元地図を使うことにより，ビデオのどのフレームからどのフレームまでどのような建物が写っていたかを計算により自動抽出することができ，それを索引として付与する．アルゴリズムを実装して，具体的に街頭で撮影を行い，提案している自動索引法が非常に有効に働くことを確認した．

In this paper, a novel indexing method for video contents is proposed and verified. This method calculates automatically the building objects which should be captured by a digital video camera. A shooter takes a video in a building area with an wearable computer, GPS, and Gyro. Using time data and shooter's position data taken by GPS, posture data of the video camera taken by Gyro attached on the camera, and 3-dimentional geographic data stored in a database, the system can calculate what building objects are captured from which video-frame to which video-frame automatically. These results are used to index the video database. The proposed algorithm is implemented, and it is shown that the proposed method works very well as designed by a concrete experiment done in a central city.

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，ビデオカメラを自在に持ち歩いて長時間移動しながら多数のビデオを撮影することが多くなっている．それに伴い，当然のこととして，大量のビデオが取得されることとなった．また，一方では，MPEG に代表されるビデオの圧縮技術と標準化が進み，ビデオをデータベース化することが，今日の大容量ストレージ技術の進歩と相まって可能となっている．

しかし，大量なビデオは格納されるだけでは価値が無く，的確に検索できうることが肝要である．そのためには，ビデオに的確な索引付けがなされていなければならない．索引付けは大別すると，人手によるものと，自動索引付けの２種類に分けられる．ビデオが少量である場合には，人手による索引付けも可能であるが，大量である場合には，人手が嵩むだけでなく，多くの時間を必要とし，かつエラーの起こりやすいものと指摘されており，自動索引付けが望まれることは言うまでもない．後者の問題はビデオのデータベース化が叫ばれた当初から多大の関心を呼び，これまでに多くの研究がなされてきているが，問題は単純ではない．

ビデオは物理的な単位であるショット，いくつかの連続したショットからなる論理的な単位であるシーン，そして一般には複数のシーンからなるビデオクリップとして成り立っている．ディジタルビデオ処理の研究・開発は次のように分類できる[1, 2]：

(a) ショットを検出する（shot detection）研究．
(b) シーンを検出する（scene or story detection）研究．関連して，シーンの類似検索に関する研究などを含む．
(c) ビデオあるいはビデオの部分列にメタデータを付与する研究．メタデータは索引（index）あるいは注釈（annotation）と読み替えてもよい．従来，ビデオからオーディオビジュアルな特徴量を抽出してその内容記述を行おうとする研究が多数行われてきた．画像理解技術，被写体オブジェクトの抽出と追従技術，音声認識技術，話者認識技術，文字認識技術，あるいはカメラワーク情報などの手法が用いられている．
(d) ビデオ検索技術の研究．問合せを発行して所望のビデオクリップやシーンをビデオデータベースからインタラクティブにあるいは自動で効率よく得る研究．

本研究は上記(c)項に類別されるが，特徴量を抽出して索引付けを行おうとする従来型の方法ではなく，ビデオに何が写っているかを直接計算してビデオの部分列（これをユニットということにする）に 索引を付与しようとする研究である．このような発想を現実のものとするために，我々が想定しているビデオ収録環境は次のとおりである：

- GPS(Global Positioning System)によりビデオ撮影者の位置と時刻がわかる．
- ビデオカメラに取り付けたジャイロセンサによりビデオカメラの姿勢を知ることができる．
- 撮影地点を含む３次元地図があり，建物オブジェクトの３次元データが取得可能である．

次章で示すように，これらのデータを用いることにより，ビデオカメラの視野に入り，実際にビデオに写っている建物オブジェクトを計算して，自動的にユニットにその建物 ID を索引として付与することができる．

関連研究として上田らの研究[3]がある．そこでは，ビデオ撮影者が自分の過去の行動を振り返る際に必要な機能としてビデオ検索を位置付け，撮影場所に注目した索引機構を提案している．これは，ビデオ撮影時に GPS を用いて撮影場所の位置情報を緯度・経度の形で取得し，それを地名やランドマーク名といった地理情報に変換し，それらをキーワードとしてビデオ検索を実現するというものである．しかしなが

♥ 学生会員 お茶の水女子大学大学院人間文化研究科博士前期課程 yukiko@dblab.is.ocha.ac.jp
♦ 日本アイ・ビー・エム株式会社
♠ 正会員 お茶の水女子大学理学部情報科学科 masunaga@is.ocha.ac.jp

ら，この研究では，用いた地図は２次元であり，また被写体建物オブジェクトの抽出計算も行わなかったので，実際にビデオに写っているオブジェクトと索引が一般には一致せず，たとえば，我々の方法では答えられる「銀座三越デパートが 10 秒以上写っているビデオが欲しい」などという検索要求に答えられない．

2．被写体建物オブジェクトの自動抽出と索引付け

2.1 自動抽出・索引付けシステムの概念

図１は我々が開発しているシステムの概念図である．ビデオ撮影者の位置と時刻を取得するために GPS, ビデオカメラの姿勢を知るためにジャイロを使用する．撮影者が身につけているウェアラブルコンピュータで，GPS データ，ジャイロデータ，カメラの画角データ（レンズの画角：今回使用するビデオカメラ（Sony 社製 DCR-PC1）の画角は横 44°，縦 35°），３次元地図データ（三菱商事製 DiaMap）を，本論文で提案するアルゴリズムで総合的に処理して，被写体建物オブジェクト ID を取得し，その ID とその建物が写っているユニット（次節）を対応付けるインデックステーブル XBuildings が作成される． ビデオはショット単位で処理する．

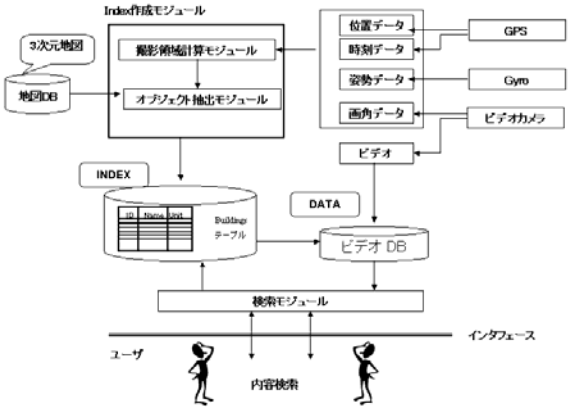

図1. ３次元地図を用いた被写体建物オブジェクトの自動抽出・索引付けシステムの概念

Fig.1 Concept of an Automatic Extraction and Indexing System of Building Objects using a Three-Dimensional Map

2.2 ユニット：索引付けの単位

ビデオ撮影者は市街地を歩行しながら，かつビデオカメラを左右上下に振りながらビデオ撮影をする．$u_{v,o,i}$は建物オブジェクト（oとする）が，ビデオ（vとする）のあるビデオフレーム（bとする）から始まり，あるビデオフレーム（eとする）まで連続して写しこまれている，第i番目の部分とする（i≧1）．このビデオフレームの連続を$u_{v,o,i}$ =(v, o, i, b, e)で表わし，ユニット（unit）と呼ぶ．図２に被写体建物オブジェクトとユニットの関係を示す．例では，建物O_1とそれが連続して写っているユニットの対が図 1 のINDEXデータベースのXBuildingsテーブルに記録されるので，少なくとも(O_1, $u_{v,o,1}$）と(O_1, $u_{v,o,2}$)の２つのタップルが存在する．その結果，建物IDで問い合わせると，その建物が写っているユニット全てを知ることができる．ビデオは 1 秒間に 30 フレームづつ撮影されるので，ショットの撮影開始時刻をGPSデータより取得すれば，各フレームの撮影時刻も割り出せる．

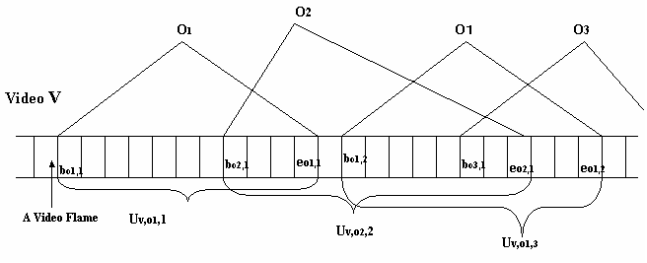

図2. 被写体建物オブジェクトとユニットの関係

Fig.2 Relation between Building Objects and Units

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 基本的考え方

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とその実装を示す．基本的な考え方は，建物は宙に浮いているわけではないので，図 1 の Index 作成モジュールに示したように，まず，３次元のビデオ撮影領域を２次元の地図上に写像して撮影領域を特定し，続いて３次元地図から建物の高さ情報を取得して，その領域に存在して実際に写っているべき建物オブジェクトを計算する．

(1) ２次元地図上での撮影領域の特定

ビデオ撮影時刻τで，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 の四角錐（P）の部分がビデオカメラで撮影される３次元領域であり，斜線の３角形の部分（T）はその３次元領域を２次元平面（X-Y平面）に投影することにより求まる．今回は使用してい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．Lは視野の距離を現すが，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．本研究ではL=80mでビル街（銀座）での実験を行っている．使用するビデオカメラの横の画角は 44°で，２次元平面上への投影視野 lは，$l = L \times \cos(\beta)$で求まることにより，Tは３つの頂点：T'=(0, 0)，T_B=($l \times \cos(\alpha\text{-}22)$，$l \times \sin(\alpha\text{-}22)$)，$T_A$=($l \times \cos(\alpha+22)$，$l \times \sin(\alpha+22)$)で囲まれる三角形と定められる．被写体建物オブジェクトはその底面がTの中もしくはTと交わる建物オブジェクトでなければならないことがわかる．

(2) 被写体建物オブジェクトの抽出

次に，３次元地図DiaMapから建物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（変数hで表す）により，５つの頂点：P_1=(0,0,h)，P_2=($l \times \cos(\alpha\text{-}22)$，$l \times \sin(\alpha\text{-}22)$，$L \times \sin(\beta+17.5)+h$)，$P_3$=($l \times \cos(\alpha+22)$，$l \times \sin(\alpha+22)$，$L \times \sin(\beta+17.5)+h$)，$P_4$=($l \times \cos(\alpha+22)$，$l \times \sin(\alpha+22)$，$L \times \sin(\beta\text{-}17.5)+h$)，$P_5$=($l \times \cos(\alpha\text{-}22)$，$l \times \sin(\alpha\text{-}22)$，$L \times \sin(\beta\text{-}17.5)+h$) からなる四角錐$P$の範囲内，もしくは$P$と交わる建物オブジェクトが，(1)で得られた候補のうちで，実際に写っている建物オブジェクトの候補となるが，自分より前，つまり自分より撮影者に近い位置に自分より高い建物がある場合には実際にはビデオに写っていない．

次節ではこれらを考慮した被写体建物オブジェクト抽出のアルゴリズムを説明する．

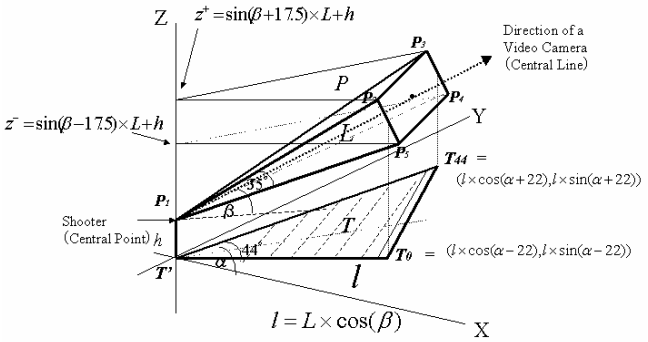

図3. ビデオ撮影領域の3次元空間から2次元空間への投影
Fig.3 Mapping of a Video Captured Area from Three-Dimensional to Two-Dimensional Space

3.2 被写体建物オブジェクトの抽出アルゴリズム

図4に, 前節の考え方に基づき実装した被写体建物オブジェクトの自動抽出アルゴリズムの概略を示す.

3.2.1 建物オブジェクトに関するパラメタ

抽出アルゴリズムは, 建物オブジェクト（o）に関して次のパラメタを有する.

- 底面の重心座標: (o.x, o.y)
- 高さ: o.height
- 撮影地点から重心への距離: o.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: o.visible

3.2.2 2次元地図上での建物オブジェクト抽出

2次元地図上（DiaMapの2次元地図のみを使用）に投影された撮影領域中の建物オブジェクトを抽出する. 3.1で述べたように, Tの中もしくはTと交わる建物オブジェクトがビデオカメラに写るべき候補である. 図4に示したように, T内で直線$T'T_i$ (T_i=(l×cos(α-22+i), l×sin(α-22+i)) を1°づつ(0≦i≦44)動かしていき, 直線$T'T_0$ から直線$T'T_{44}$まで交差する建物オブジェクトを地理情報システムArcView3.2（ESRI社製）に備わっているフィーチャ選択機能を用いて抽出し, 各iに対して配列a[ ][ ]={a[i][ ]| i=0, …,44}にrの昇順にその建物IDを格納していく.

3.2.3 高さ情報を用いた被写体建物オブジェクトの抽出

次に, 3次元空間的視点により, DiaMapから建物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.

あるiにおいて直線$T'T_i$ 上にある建物オブジェクト（o）を全て配列{a[i][0], a[i][1],…,a[i][n_i]}にo.rの昇順に格納する. ここでn_iとは, a[i][n_i+1]=nullであるが, あるn(n ≦ n_i)においては, どのようなa[i][n]もnullではない数である. すなわち直線$T'T_i$ と交わった建物オブジェクトはn_i+1個ある. 次に配列の1番目に格納された建物オブジェクト（a[i][0]）の高さとその位置でのPの底辺の高さ (a[i][0].r×tan(β-17.5) +h)を比べ, 高ければa[i][0].visibleの値を“on”とし, そうでなければ“off”として, 変数sの初期値を0とする.

次に図4の内側のループにおいて, m=s+1とおく. 配列のはじめの建物オブジェクトの高さ（a[i][s].height）と配列の2番目以降のオブジェクトの高さ（a[i][m].height）を比べ, 後者が高く且つPの底辺の高さ (a[i][m].r×tan(β-17.5) +h)以上ならば, 後者のフラグの値（a[i][m].visible）を“on”にし s=m とする. そうでなければ“off“とし, m=m+1 とする. これを a[i][m]が null でない限りループをまわして配列（a[i][]）に格納されている全ての建物オブジェクトの visible のフラグを設定する. すべての i の値で調べ終わり, 最終的に visible のフラグが立っている全てのオブジェクトが, 時刻τにおいてビデオに写っているべき建物オブジェクト群である.

3.2.4 過少誤認の事実上の無発生

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してしまう可能性が理論上ある. しかし, 例えば視野長L=80mとすると1°間隔で配列を作成しているので80m先の隣接する配列間の距離は2πL/360　1.4mとなる. 実世界では, 1.4m 幅の建物は考えにくいので, 本来抽出されねばならない建物が抽出されないで残ってしまうという過少誤認の発生は事実上ないと考えられる. 換言すれば, 市街地を画角 44°でビデオ撮影した場合には, 本アルゴリズムで実装したように, 1°毎に被写体となるべき建物オブジェクトを抽出していって問題はないことを言っている.

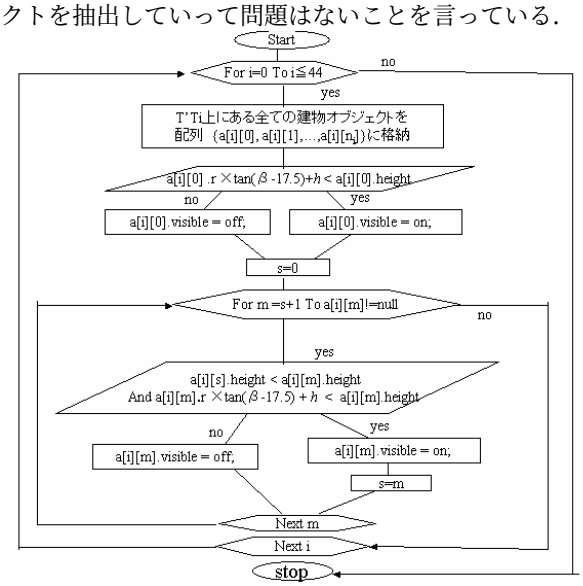

図 4. 自動抽出アルゴリズムの概略
Fig.4 Outline of Automatic Extraction Algorithm of Building Objects

4. 被写体建物オブジェクトの自動抽出と検証実験

4.1 実験環境と方法

提案したアルゴリズムは Windows XP マシーン上で ArcView3.2 のシステム開発用プログラミング言語 Avenue で実装した. アルゴリズムの有効性を検証するために, 以下の実験を行った：

⑴ 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.
⑵ 撮影位置・ビデオカメラの姿勢情報を取得する.
⑶ 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.
⑷ 抽出された建物オブジェクトと実際の映像との比較検証を行う.

図5に示されているように, 本実験では銀座のビル街をソニービル近辺からプランタン銀座の方向に向かって歩きながら撮影した（撮影者の軌跡はGPSで取得したもの）. 収録されたビデオの長さは約10分で, これが一つのショットを構成している.

4.2 比較検証結果

検証実験は，ある撮影時刻τを定めて，その時刻に撮影されたビデオフレームに実際に写しこまれている建物と前章のアルゴリズムによって計算されて写し込まれているべき建物オブジェクトを実際に比較して行った．この時刻τは，被写体となった建物群の密集度や高低差，ビデオカメラ（の視野）と建物群の距離などを勘案して，検証結果が普遍性をもつと考えられる地点での時刻とした．

図5で薄いグレー表示されている建物オブジェクト群が本実験において時刻τにおいて計算により抽出された被写体建物オブジェクト群を表している．ビデオカメラに向かって前方の高い建物オブジェクトによって抽出されない建物オブジェクトがあることがわかる．この時刻では合計 15 個の建物オブジェクトが抽出されたことがわかる．左下に表されているテーブルは抽出された建物オブジェクト群を表すテーブルである．そこには，時刻τで抽出された被写体建物オブジェクト ID，その重心座標，高さなどが記録されている．

図 5. ビデオカメラの軌跡と自動抽出された建物オブジェクト群
Fig.5 Trajectory of a Video Camera and Building Objects Automatically Extracted

図 6 はその時刻τにおける実際のビデオ映像である．比較を行ってみると，実際の映像では一番手前の建物が工事でなくなっていたが，それ以外の建物オブジェクトについては写っている建物オブジェクトが抽出されていることがわかる．3 次元地図を使用し高さを比較したことで，周囲の建物オブジェクトに隠れて写っていないはずの建物オブジェクトが抽出されることなく，実際に写しこまれている建物オブジェクトが抽出されていることが確認でき，本研究において提案した被写体建物アルゴリズムの有効性が確認できた．

5. まとめと今後の課題

本論文では，GPS とジャイロセンサデータに加えて，3 次元地図を使うことで，ビデオに写し込まれている建物オブジェクトを自動計算する考え方と，その自動計算アルゴリズムを提案し，実際に検証実験を行い，その有効性を確かめた．

今後の課題として，このアルゴリズムに基づいたビデオの索引法の提案と実装，それに基づいたビデオの効率の良い格納・検索方法の研究・開発が挙げられる．また，今回の実験を通して判明したのだが，街路樹などの障害物に対する対処法，建物屋上に設置された看板類の扱い，視野長 L の設定法，建物オブジェクトの重要度の考慮や富士山・東京タワーといったランドマークの扱い方，さらに 3 次元地図に加えて人文地理データ（花火大会や盆踊りなど）を利用して，さらに高度な索引付けの研究・開発が挙げられる．

図6. 時刻τにおけるビデオの映像
Fig.6 Video Image at time τ

[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． 3 次元地図 DiaMap は三菱商事のご好意による．ここに記して謝意を表する．

[文献]

[1] Gaughan, G., Smeaton, A., Gurrin, C., Lee, H., McDonald, K.: “Design, Implementation and Testing of an Interactive Video Retrieval System,” Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp.23-30, November 2003.

[2] Wang, Y., Ostermann, J., and Zhang, Y-Q.: “Video Processing and Communications,” (book) Prentice Hall, 2002.

[3] 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 年 3 月．

佐藤 有紀子 Yukiko SATO

お茶の水女子大学大学院人間文化研究科博士前期課程在学中．3 次元地図を用いた建物オブジェクトの自動抽出の研究・開発に従事．2003 お茶の水女子大学理学部情報科学科卒業．日本データベース学会学生会員．

石黒 玲 Rei ISHIGURO

日本アイ・ビー・エム（株）勤務．2004 お茶の水女子大学大学院人間文化研究科博士前期課程修了．在学中 3 次元地図を用いた建物オブジェクトの自動抽出の研究・開発に従事した．

増永 良文 Yoshifumi MASUNAGA

お茶の水女子大学理学部情報科学科教授．1970 東北大学大学院工学研究科博士課程修了，工学博士．データベースシステムの研究・開発に従事．情報処理学会データベースシステム研究会主査，情報処理学会監事，ACM SIGMOD 日本支部長などを歴任．情報処理学会フェロー．電子情報通信学会フェロー．日本データベース学会現会長．著書に「リレーショナルデータベース入門[新訂版]」(サイエンス社)など．