

# 3D 仮想空間を用いた動画蓄積システムの試作

住野 優<sup>†</sup> 明石 康久<sup>†</sup> 上島 紳一<sup>†</sup>

<sup>†</sup> 関西大学大学院総合情報学研究科 〒569-1095 大阪府高槻市霊仙寺町 2-1-1

E-mail: †yutaka.s@mb.newweb.ne.jp, ††skya@ff.iij4u.or.jp, †††ueshima@res.kutc.kansai-u.ac.jp

あらまし 蓄積された数多くの動画から求めているシーンを検索するには、予めシーン毎にコンテンツ情報を記述しておくことが望ましい。しかし、長期間に及ぶ動画に対してこの情報を手入力することは困難である。本稿では、動画と同期させた GPS と電子方位計のデータをシーンに対する基本的な索引とし、重要なシーンにのみコンテンツ情報を付与した動画蓄積システムについて述べる。本システムでは、3D 仮想空間を用いて利用者が、ビデオカメラの位置（視点）、或いは被写体の位置（注視点）をキーとした検索が行うことのできる特徴を持つ。また、このシステムを用いた時空間非同期のコミュニケーションを目指した動画コミュニケーションシステムについても述べる。

キーワード ユーザインターフェース, Web とインターネット, 位置情報, 情報検索, マルチメディア DB

## A prototype for video storage system utilizing 3D virtual space

Yutaka SUMINO<sup>†</sup>, Yasuhisa AKASHI<sup>†</sup>, and Shinichi UESHIMA<sup>†</sup>

<sup>†</sup> Graduate School of Informatics, Kansai University Ryozenji 2-1-1, Takatsuki-City, Osaka, 569-1095 Japan

E-mail: †yutaka.s@mb.newweb.ne.jp, ††skya@ff.iij4u.or.jp, †††ueshima@res.kutc.kansai-u.ac.jp

**Abstract** In case accumulating a large volume of videos, it is preferable to define the contents information scene by scene for later retrievals. However, the work to input such information manually costs a long period of time. In this paper, we propose a video storage system equipped with indexed data by use of GPS and Digital Magnetic Compass, scene description, and meta data of contents. In our system, we employ virtual collaborative space to search for a specific scenes from given viewpoints, as well as from given gazing points as a key. We also show an application of our system aiming at time-asynchronous communication by use of accumulated videos, and show a prototype system.

**Key words** user interface, Web and Internet, positional retrieval, information retrieval, Multimedia DB

### 1. はじめに

近年、インターネットの高速化と動画圧縮技術の発展やハードディスク容量の増加により、動画を用いた情報交換が容易になりつつある。この結果、日常生活を動画として蓄積し、不特定多数の人々と動画を共有できるようになると考えられている<sup>[1]</sup>。しかし、大量に蓄積された動画を検索する方法やインターフェースは、まだ十分に確立されているとは言えない。全てのシーンに対して人がコンテンツ情報を書き加えるのは煩わしく、動画が長時間に及ぶ程困難な作業となるため現実的ではない。そこで動画メタ情報や位置情報<sup>[4]</sup>等を用いる方法が提案されている。また位置情報についてはカメラ位置だけでなく方向も取得することで、被写体の位置をキーとした検索が可能となる<sup>[3]</sup>。

本稿ではコンテンツ情報は記述者が必要と感じた重要なシーンに対してのみ MPEG-7 形式で記述し、それ以外のシーンに

対しては位置情報と撮影時刻を用いて検索を行えるようにするだけでなく、検索キーとする位置情報を緯経度などの数値で入力するのではなく、実空間を再現した 3D 仮想空間内の検索したい場所へ自分のアバターを移動させることで、自動的にその位置が検索キーとして入力されるシステムを提案する。これにより実空間を探索するような感覚でシーンを検索することが可能となる。更に、これらの動画の蓄積検索方法を用いて仮想空間内で動画による会話を行えるようにし、時空間を越えた非同期型の動画コミュニケーションの実現を試みる。

2 章では本研究の研究動機を説明し、3 章で動画索引情報に用いているコンテンツ情報や動画メタ情報、さらに位置情報について述べ、4 章で 3D 仮想空間を用いて、位置情報から動画を検索する方法を紹介する。5 章ではその検索方法を用いた時空間非同期のコミュニケーションを提案する。6 章は動画のインデキシング及びシーンの検索、さらに時空間非同期のコミュニケーションを行うプロトタイプシステムを紹介する。7 章で

は関連研究について述べ、8章でまとめと今後の課題について述べる。

## 2. 研究動機

本研究では、長期間の動画を以下のように活用することを考えられる。

(1) 動画を用いて不特定多数の人々と時間非同期の意見交換を行う

(2) 1日に放送される全てのテレビ番組を蓄積し、後で好きな番組だけ視聴する

(3) 日常生活の全てを動画として蓄積し、記憶として利用する

(1)については、5章で述べる。(2)は放送局がインデキシングを行うことで、視聴者は全ての番組を動画にして索引情報と共にハードディスクへ蓄積しておき、後で見たい番組を選択して視聴する方法である。(3)は、人が見る映像を動画として撮影しておき、記憶の代わりに利用する方法である。さらにこの記憶動画をインターネットに公開することにより他人との記憶の共有も可能になると考えられる。

本稿では、動画をこのように利用することを前提とする。しかし、そのためには動画に対して効率の良いインデキシングを容易に行い大量の動画から1つのシーンを簡単に抽出できる動画蓄積・検索システムが必要である。

現在、動画の検索方法として、コンテンツ情報を手入力で記述しておくコンテンツ検索<sup>[5][6]</sup>や、位置検索<sup>[4]</sup>、イメージマッチング、脳波検索<sup>[1]</sup>などが提案されている。これらの検索方法に用いる索引を記述する際の容易性と、シーン検索に対する有効性の比較を表1に示す。表1のようにコンテンツ情報を手入力する方法は様々な情報を柔軟にインデキシングできるが、索引付けに時間や労力がかかる。位置検索は、例えばグラウンドを検索キーにするとスポーツをしている動画が抽出されるといったように、ある程度動画の内容を絞り込むことができる。しかし誰が写っているのか、どんなスポーツが行われているのかまでは特定できない。イメージマッチングは、映っている人物やモノを検索する場合は有効であるが、イベントなどを検索したい場合は有効とは言えず、検索にも時間がかかる。脳波情報を用いた検索は、何に対して興味を持った映像なのかまでは不明なため特定の人物やイベント等でシーンを抽出するのは難しい。

このように現在のところ、1つのインデキシング方法のみで容易性と有効性を両立させるのは難しい。そこで我々はコンテンツ検索の有効性と位置検索の記述時の容易性を組み合わせることで、容易性と有効性を両立させたシステムの構築を試みる。これにより、これまでただ蓄積していた動画ファイルを、少ない時間で効率良くシーン単位で検索することが可能になると考えられる。

また、我々は3Dの仮想空間をインターフェースに用いることで、より自然に近い感覚で動画を検索できるシステムの構築を目指している。

表1 検索方法による容易性と有効性の比較

	索引付けの容易性	シーン検索時の有効性
コンテンツ検索	×	
位置検索		
イメージマッチング		
脳波検索		

## 3. 動画の索引情報

本研究では動画作成者(放送ならば放送局)が、動画だけでなく動画の索引情報も公開することで、探しているシーンを誰でも好きな時に検索できるようにすることを試みる。そこでより多くの人々と索引情報を共有できるように動画メタ情報とコンテンツ情報を国際標準規格であるMPEG-7形式でXMLファイルに記述し、他のシステムとも共有できるようにしている。本節では、この動画メタ情報とコンテンツ情報、さらに位置情報を加えて、これらの情報がどのような役割を果たしているのか述べる。

### 3.1 動画メタ情報及びコンテンツ情報を用いた索引

不特定多数の人々と動画を共有する場合、どんな環境であっても再生できる動画ファイルを蓄積していくことが望ましい。しかし、実際にインターネットで蓄積・公開されている動画ファイルを調べると、MPEG(Moving Picture Experts Group)形式をはじめ、RM(RealMedia)形式、AVI(Audio Video Interleaving)形式、WMV(Windows Media Video)形式等あらゆる形式の動画ファイルが混在しており、それぞれ再生できる環境が異なる。そこで本研究では、利用者の環境に合った形式の動画ファイルが抽出できるように、動画メタ情報としてコーデックに関する項目を設けている。その他、動画メタ情報として以下のような項目がある。

- (1) データレート、解像度
- (2) アスペクト比
- (3) フレームレート
- (4) オーディオ情報(形式、チャンネル数、サンプル率)

(1)は画質に関するメタ情報である。(2)は映像の縦横比、(3)によって動画の動きの滑らかさを表し、(4)によって音質を表現している。これらの動画メタ情報を併せることにより、利用者は求めている画質や動き精度、及び音質をある程度指定でき、その条件内で自分のPCで再生できる動画ファイルに検索結果を絞ることが可能となる。

一方コンテンツ情報は、以下の2つに分けることができる。

- (1) 動画ファイル全体に対する説明
- (2) シーンに対する説明

(1)は動画のタイトルやあらすじなど、1つの動画ファイル全体に共通する情報である。しかし動画の内容は時々刻々と変化し、1つのファイルに複数の内容が撮影されているのが一般的である。そこで動画をシーン単位に分割し(2)を記述することでシーン単位での検索を可能にしている。しかし、全てのシーンにコンテンツ情報を手入力するには、非常に多くの時間を要するため、記述者が重要と感じたシーンに対してのみ記述する

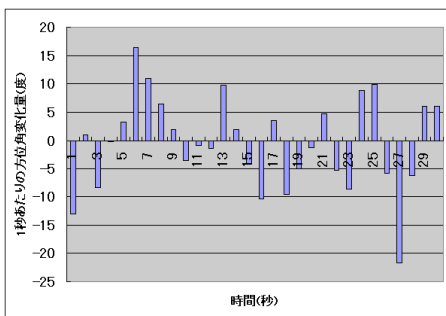


図 1 カメラ方位角の変化量

こととする。

このメタ情報とコンテンツ情報を記述している MPEG-7 を図示すると図 3 の「MPEG-7.xml」のようになる。

### 3.2 位置と方向を用いた索引

3.1 節で述べたコンテンツ情報で検索できるのは、コンテンツ情報の記述者が重要と感じたシーンに限られる。しかし、記述者が重要と判断したシーンと、利用者が見たいと思うシーンが必ずしも一致するとは限らない。そこで全てのシーンに位置情報を付加することで、コンテンツ情報がないシーンに対しても検索を行えるようにする。本研究で用いる位置情報は GPS で取得するカメラの位置（緯度・経度・高度）と、電子方位計で取得するカメラの向き（方位角・仰角・回転角）である。

ここで、実際にキャンパス内の風景を撮影したときのカメラ方向の変化量を図 1 に示す。

図 1 のように動画は、静止画と異なりカメラの向きが一点に定まらず常に変化しており、方位角は 1 秒間で 20°以上変化する場合があることが解る。そこで位置情報の取得間隔を 1 秒間隔とした。次に、GPS と電子方位計のデータを蓄積する際に問題となるのが、動画と位置情報の時間同期である。動画、GPS、電子方位計の記録開始時刻を同時にすることは事実上不可能であり、必ず記録を開始するタイミングにズレが生じる。さらに全てのシーンに位置情報を付加する必要があるため GPS と電子方位計は撮影開始時刻  $T_1$  より早く記録を開始し、撮影終了時刻  $T_2$  より遅く記録終了しなくてはならない。結果、位置情報のログファイルには図 2 の灰色の部分で示した不必要な部分が必ず生じ、XML に記述する際には灰色の部分削除しておく必要がある。そこで、まず  $T_1$  と  $T_2$  を以下のようにして求める。

$$T_1 = \text{動画ファイルの作成日時} \quad (1)$$

$$T_2 = T_1 + (\text{動画ファイルの長さ}) \quad (2)$$

このようにして求めた  $T_1$  の時刻の GPS と電子方位計のログデータを 0 秒とし 1 秒間隔で  $T_2$  の時刻まで XML に記述することで、動画と位置情報の同期化を行っている。

上記の方法で作成された位置情報 XML は図 3 の「GPS 1.xml」と「Commpass 1.xml」のようになる。この XML を用いれば、利用者はカメラの緯度、経度、高度、方位角、仰角、回転角を検索キーとして、カメラの角度からシーンを検索することが可能となる。

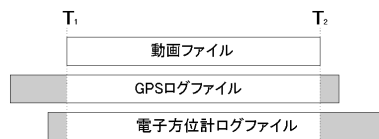


図 2 動画ファイルと位置データの時間の同期

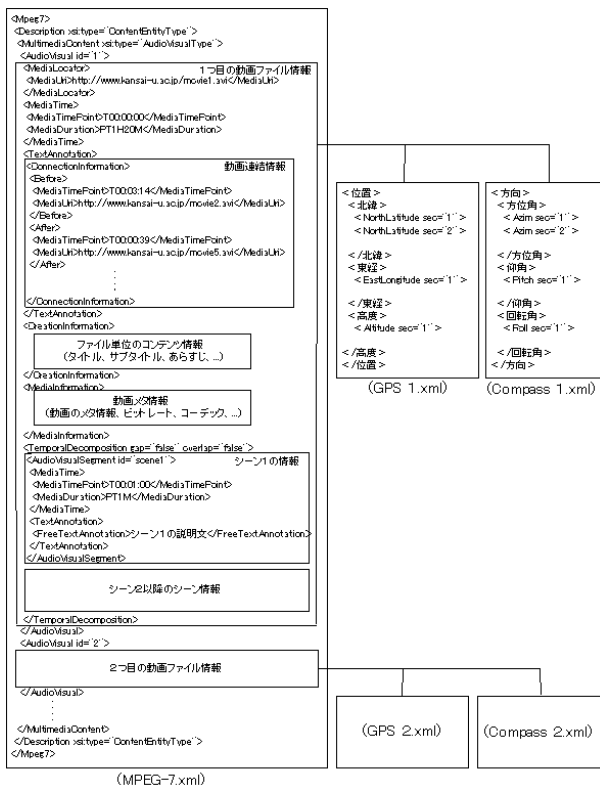


図 3 XML の構成

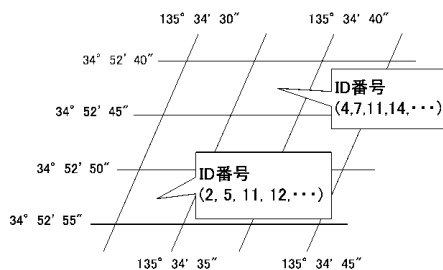


図 4 矩形単位の ID 管理

しかし、位置情報を用いて検索すると検索処理に時間がかかるという問題がある。これは緯経度や方位角などの位置情報が 1 秒間隔で蓄積されているため XML の要素数が増加し、パーサへの負荷が大きくなるためである。この問題を解決するため図 4 のように、緯経度を 5 秒間隔で矩形に区分けし、矩形内を通過した動画の ID 番号を XML とは別にデータベースにも格納しておくことで、全動画の XML を検索せずに指定した範囲内の動画にある程度絞り込み、検索時間を短縮することができる。

### 3.3 位置通過型と位置参照型

3.2 節で述べてきた位置情報を、本研究では位置通過型と呼

んでいる。位置通過型とは、既述したように動画撮影時におけるカメラの位置情報である。つまり「どこで」撮影したかという情報を表している。

しかし2章で述べた、人々とのコミュニケーションに用いる動画やテレビ番組などに対してインデキシングを行う際、「どこで」撮影したかではなく「どこの」話題をとりあげているかが重要となる場合がある。そこで本研究では位置通過型以外に、特定の位置に関連した内容が述べられている動画に対して、その位置座標も記述している。この動画の内容に関連した位置情報を、本研究では位置参照型と呼んでいる。

この位置参照型の位置情報を記述しておくことで、ある事件のニュース番組を蓄積した場合、その事件現場の位置座標からそのニュース番組を検索することができる。また個人的に地点Aにいる人が地点Bの話題を蓄積した場合、地点Aで位置通過型の検索を行う方法と、地点Bで位置参照型の検索を行う方法の2通りの方法で位置検索することも可能となる。ただし位置通過型はGPS等を用いて自動的に蓄積されるが、位置参照型は位置情報自体が意味的情報であるため4章で述べる3D仮想空間を利用して、XMLの記述者が位置座標を指定しなければならない。

また検索時には位置通過型と位置参照型、更に両方型の3つの選択肢を用意する事で、より柔軟な検索結果を得る事ができるようにしている。

#### 4. 仮想空間を用いた動画検索

3.2節で述べた位置情報を用いて検索を行う場合、検索キーとなる緯経度などを手入力するのは困難である。そこで本研究では検索者が実際にいる位置をGPSで取得し、検索キーにする方法を用いている。しかし、この方法では検索者がいる地点を中心とした検索しか出来ない。本節では、この問題を3D仮想空間を用いて解決する方法を述べる。また3D仮想空間を用いた注視点からの検索についても述べる。

##### 4.1 3D仮想空間の利用

位置情報を用いた検索を行う場合、検索したい場所の緯経度を調べて手入力するのは煩わしい。また検索結果として抽出した動画の撮影場所を緯経度等の数値で表現しても、どの地点を指しているのか解りにくい。そこで本研究では3D仮想空間内に動画画像を割り付けることで、この問題を解決している。動画画像割り付けとは、実空間の位置座標と仮想空間の位置座標が対応している事を利用して、GPSで取得した撮影地点と対応する3D仮想空間の位置に動画の存在を表す静止画を表示する事である。これにより、動画を位置情報から視覚的に探索する事ができる。

本研究では、仮想空間にデジタルキャンパスを用いている。これは大学のキャンパスを3D仮想空間としてWeb上に構築し、シミュレーションや遠隔教育等を行なっている我々<sup>[11]</sup>の仮想空間である。このデジタルキャンパスは実空間を再現する形で構築されており、図5のように3D仮想空間と実空間が対になっている。従って、実空間のGPS値を基に、仮想空間内の同じ位置に利用者のアバターを立たせる事も可能である。

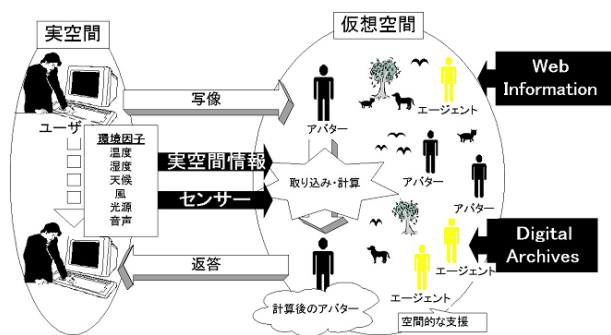


図5 デジタルキャンパスの概要図

そのため利用者に遠隔地から現実空間の任意の地点を移動しているようにみせる事が可能となる。

本研究では、デジタルキャンパスのこれらの機能を位置を用いた動画検索に利用している。しかし、この3D仮想空間は矩形ごとに3次元CGとテキストチャで構成されている。そのためGPSで取得した緯経度を基に3D仮想空間内の対応する位置に動画画像を割り付けるには、緯経度を3D仮想空間の位置情報に変換する必要がある。そこで、仮想空間を構成する矩形の中から代表を1つ選び、その矩形の緯経度を調べて基準点としている。これにより全ての矩形に対して、基準点からのオフセット値を用いることで緯経度に変換できる。この変換を用いることで、仮想空間内で自分のアバターが立っている場所の位置座標から実空間の緯経度を求め、その値を検索キーとしてシーンを抽出することができる。

##### 4.2 注視点からの検索

前述したように撮影場所をキーとした検索は、GPSや仮想空間を用いて検索領域の中心座標を検索キーとして指定し、指定した場所から撮影されたシーンを抽出する。しかしシーンを検索する場合、「どこから撮影したか」ではなく「どこを撮影しているか」が重要になることが多い。例えば、ある建物Aが写っているシーンを、撮影時のカメラの位置に関係なく抽出したい場合が考えられる。これをカメラの位置と方向をキーとした検索で行うには、建物Aから少し離れたある一点を指定し、そこから建物Aの方向を写しているシーンを検索することになる。しかし、この方法では1回の検索で1方向からのシーンしか抽出できず、何回もカメラの位置を変えて建物Aの方向を写しているシーンを検索しなければならない。この作業は非常に煩わしく面倒である。また、この方法では建物Aが写っているシーンを全て検索出来るとも限らない。

そこで新たに被写体の位置（注視点）をキーとした検索を加えた。この検索は、以下のような流れで行う。

- (1) 検索する被写体（建物A）の位置と被写体からカメラまでの距離（検索領域）を検索者が指定
- (2) 指定された領域内から撮影した動画を抽出
- (3) 抽出された動画から、指定した被写体（建物A）を撮影しているシーンに絞り込む

(1)で検索したい被写体の位置Aと、検索領域の半径rを指定する(図6参照)。点Aは、GPSで検索者の現在地を指定する方法と、4.1節で述べた仮想空間内での現在地を指定する方

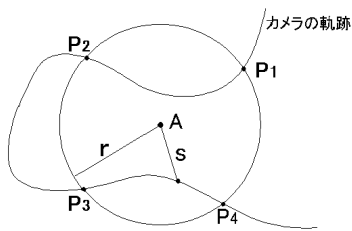


図6 2次元で表した検索範囲とカメラの軌跡

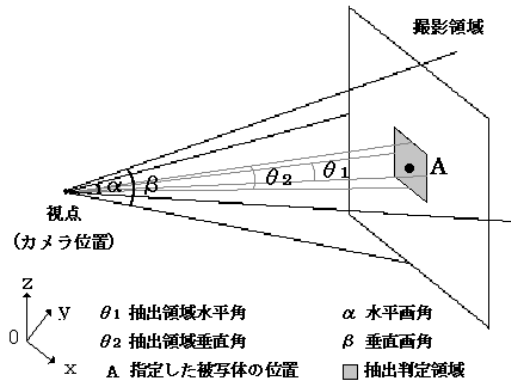


図7 抽出領域の指定

法がある。(2)は、検索者が指定した被写体の位置から撮影位置までの距離  $S$  を求め、撮影位置が指定された半径  $r$  の円内か調べる作業である。しかし、被写体の位置とカメラの位置は緯経度で表現されているが1秒の長さは場所によって異なるため、緯経度のままで距離  $S$  を計算することは出来ない。そこで2点の緯経度を平面直角座標に直してから  $S$  を求めている。この  $S$  と利用者が指定した  $r$  を比較することで、 $P_1 \sim P_2$  と  $P_3 \sim P_4$  のシーンを抽出することができる。

$P_1 \sim P_2$  間と  $P_3 \sim P_4$  間から撮影している映像の中から、求めている被写体が写っているシーンに絞り込むのが(3)の作業である。しかし「しっかり写っている」と判断する基準は、一意ではない。例えば、富士山のような数多く映像化されている地物を検索する場合は、地物をより真ん中で捉えているシーンを抽出したいと考えられ、逆にあまり映像化されていない地物を検索する場合は画面の端で少し写っているシーンでも価値があると考えられる。また、この基準は検索者によっても変わってくるであろう。そこで、カメラ画角の中心を基準として検索する被写体が何度以内に入っているシーンを抽出するかを、検索者が図7の垂直角  $\theta_2$  と水平角  $\theta_1$  を入力することで抽出領域を指定する。視点の平面直角座標を  $(x_1, y_1, z_1)$ 、注視点の平面直角座標を  $(x_2, y_2, z_2)$  とすると(3)の計算式は以下ようになる。

$$\text{Arctan} \frac{y_2 - y_1}{x_2 - x_1} - \frac{1}{2} \quad A \quad \text{Arctan} \frac{y_2 - y_1}{x_2 - x_1} + \frac{1}{2} \quad (3)$$

$$\text{Arctan} \frac{z_2 - z_1}{s} - \frac{2}{2} \quad P \quad \text{Arctan} \frac{z_2 - z_1}{s} + \frac{2}{2} \quad (4)$$

(但し、 $\theta_1$ 、 $\theta_2$  とする)

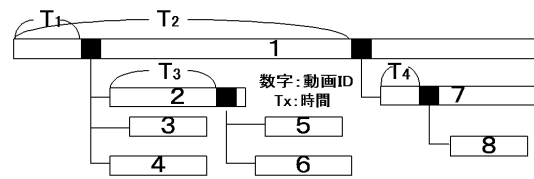


図8 動画コミュニケーションのモデル

ここで  $A$  はカメラの方位角、 $P$  はカメラの仰角である。この(3),(4)式を満たしているシーンを抽出することで、指定した位置が写っているシーンを角度に関係なく抽出することができる。

この注視点からの検索は、被写体の位置座標と被写体からカメラまでの距離を検索キーに用いているが、被写体の大きさは検索キーに含まれていない。しかし、実際に抽出したい被写体は1点とは限らず、前述した富士山のような広範囲に及ぶ大きい地物や、1つのビルだけでなく複数のビル群全体を検索したい場合も多いと考えられる。このような広範囲の地物を検索する場合は、図6の半径  $r$  を大きくし、被写体から撮影位置までの距離を長くすることでカメラ画角に入る範囲が広くなり、広範囲の地物が写っている動画はある程度抽出できるようになる。

## 5. 位置依存の動画像コミュニケーション

現在、Web を用いて他者に意見を述べたり一緒に議論したりする場合、意見をメールで送ったり掲示板に書き込むといった方法を用いるが、これらの方法では雰囲気などを相手に伝えることはできない。そこで、2章で動画利用例の1つとして挙げた、動画を利用して不特定多数の人々と意見交換ができれば、この問題を解決することができる。本節では、ここまで述べてきた動画索引情報と検索方法を用いて、この「動画を利用した不特定多数との時間非同期の双方向コミュニケーション」を実現する手法について述べる。

### 5.1 位置依存の動画コミュニケーションとは

動画コミュニケーションとは、視聴した動画に返答動画を連携させていくことで複数の動画をあたかも一連の会話のように見せる手法である。この手法はインターネット掲示板やメーリングリスト等と、ある種の類似性を持っているといえる。例えばインターネット掲示板やメーリングリストは、まずスレッド作成者がなんらかの文章を発信し、それに対して他者が自分の意見を文章で発信する形態になっている。この動画コミュニケーションも同様に最初に発信した人の動画に対して他者が自分の意見を動画で発信する形態をとっている。逆に異なる点は、掲示板やメーリングリスト等が一つのソースを引用して複数のファイルを増やしてコミュニケーションを行なっていく事に対し、本手法は一つのソースを単一のファイルで何度も再利用する、つまり引用に元のソースを用いて有効活用している。動画コミュニケーションを図8のようなツリー構造としてモデル化する。

### 5.2 動画の連結情報構成

本システムは動画で図8のようなツリーを構成するために各動画に固有の連結情報を持たせている。ここで図8の動画連結

表 2 動画固有の連結情報

動画 ID	前接続情報	後接続情報
1	NULL	2:T <sub>1</sub> ,3:T <sub>1</sub> ,4:T <sub>1</sub> ,7:T <sub>2</sub>
2	1:T <sub>1</sub>	5:T <sub>3</sub> ,6:T <sub>3</sub>
3	1:T <sub>1</sub>	NULL
4	1:T <sub>1</sub>	NULL
5	2:T <sub>3</sub>	NULL
6	2:T <sub>3</sub>	NULL
7	1:T <sub>2</sub>	8:T <sub>4</sub>
8	7:T <sub>4</sub>	NULL

表 3 時空間と本システムの関係

	空間同期	空間非同期
時間同期	通常の対面型の会話	TV 電話やネットミーティング等
時間非同期	本システムの現実空間利用	本システムの仮想空間利用

情報を表 2 に記す．固有の連結情報とは，その動画がどこの動画のどの時間から繋がっているかという前接続情報と動画のどこの時間にどの動画へと繋がるかという後接続連結情報から成り立つ．また，返答動画の再生は最初から開始する事を想定しているため連結後の開始時間情報は含まない．

これらの連結情報は図 3 にある「MPEG-7.xml」の中に記述されている．

### 5.3 対面型の時空間非同期コミュニケーション

動画を用いたコミュニケーションの利点は，相手の顔や背景等を視覚的にとらえられる事ができる点である．つまり，動画を用いたコミュニケーションは通常の会話のように顔を見合わせた形での会話となる事から対面型のコミュニケーションを行っていると言える．

4.1 節で述べたように本研究では動画の検索インターフェースに 3D 仮想共有空間を用いている．この仮想空間は Web 上に構築しているため遠隔地から，実空間にいるようなアクセスが可能であるため利用者の位置にとらわれない空間非同期のコミュニケーションを可能としている．更に，動画を蓄積しているため時間にとらわれない時間非同期のコミュニケーションも可能である．

表 3 に時空間同期の関係図を記す．また，あえて空間を同期させる事で，山頂やスポーツ観戦地の盛り上がり等，遠隔地では味わえきれない雰囲気動画を共有する地域特有のコミュニケーションも可能となる．

## 6. プロトタイプシステム

本システムで用いる動画は，図 9 のようにワゴンに乗せた PC に GPS や電子方位計を接続して位置情報を自動的に取得しながら撮影している．本節では，このように取得したデータを蓄積し検索するためのプロトタイプシステムについて述べる．

### 6.1 システム概要

本研究を実証するためのプロトタイプシステムについて述べる．本システムは Web を利用しているためインタフェース



図 9 動画と位置情報の同時蓄積の様子

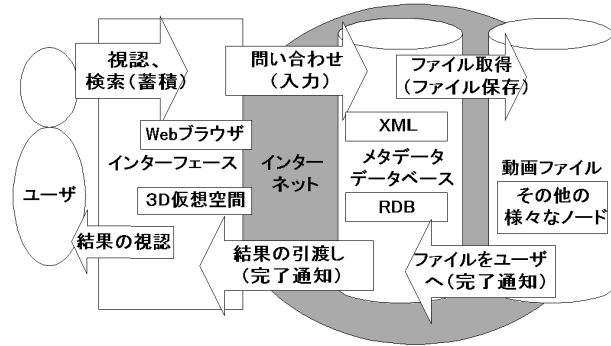


図 10 システム構成図

に Web ブラウザを用いている．更に 3DML を用いてブラウザ上に 3D 仮想空間のキャンパスを構築している．これらのインターフェースを用いてインターネット上にある XML やデータベースにアクセスを行う．XML には 3.1 節で述べたように動画メタ情報やコンテンツ情報等が記述されており，データベースには RDB として PostgreSQL を用いて 3.2 節のように矩形ごとに動画 ID を管理している．

本システムは，これらの情報を用いて検索を行い，結果をインタフェースに返す．さらに検索した動画を再生したい場合，検索結果に付けられているリンクを利用して検索インターフェースから様々なサーバやピアといったノード内に存在する動画へアクセスする事ができる．この構成を図 10 に記す．

### 6.2 システム

#### 6.2.1 動画情報蓄積機能

本システムに蓄積する動画情報は，図 11 のインターフェースから入力する．入力された動画情報は，JAVA Script で処理され XML とデータベースに格納される．このとき，3.2 節で述べた位置情報と動画の時間同期も，動画ファイルの作成日時とファイルの長さを利用することで XML 生成時に自動的にを行っている．この XML 生成インターフェースによって，ほぼ全ての索引情報が MPEG-7 等で XML に記述されるが，位置のオブジェクト化情報は RDB に格納される．これを図示すると図 12 のようになる．

#### 6.2.2 検索機能

検索は位置情報を用いた検索と，位置情報を用いない検索に大きく分けることができる．図 13 にそれぞれの検索の流れを示す．位置情報を用いる場合の特徴は，RDB にアクセスして

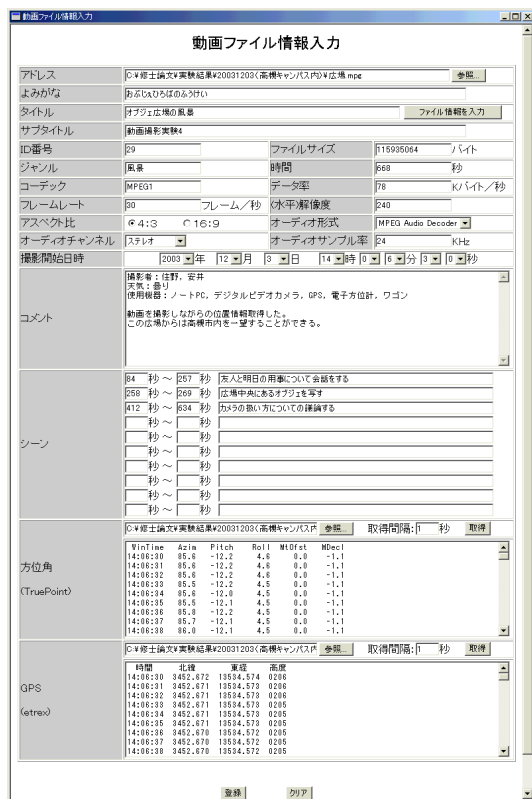


図 11 XML 生成インターフェース

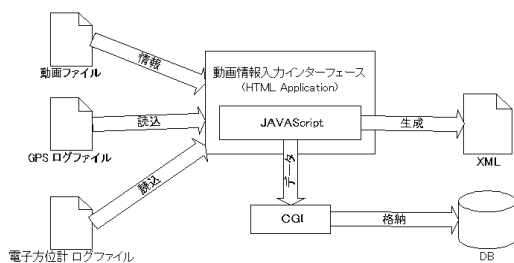


図 12 動画情報の蓄積処理

いることである。これは XML を用いて検索する前に RDB に格納されている矩形ごとの ID 情報を呼び出す事で、検索対象となる XML ファイルを絞り込んでいるためである。位置情報を用いた検索の場合、利用者は仮想空間上を移動し任意の地点から検索する。利用者が検索を行った後、システムは検索結果が静止画として仮想空間上に表示する。また、仮想空間の 1 ブロック内に複数の動画が存在した場合、その地点に表示された静止画をクリックする事で新規ウィンドウが開き、そのウィンドウに同地点の検索結果を表示している。この検索の様子を図 14 に示す。

これに対し、位置情報を用いない検索の特徴として 3D 仮想空間を用いていない点を挙げることができる。これは位置情報で絞り込みを行わない場合、抽出される動画は非常に広範囲に散らばっている可能性が高く、これを 3D 仮想空間内で表示すると仮想空間内を移動するという利用者の手間が増えるからである。そこで位置情報を用いない場合は、HTML に抽出された動画を一覧表の形で出力している。

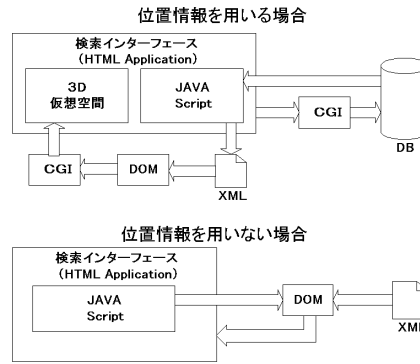


図 13 動画情報の検索処理

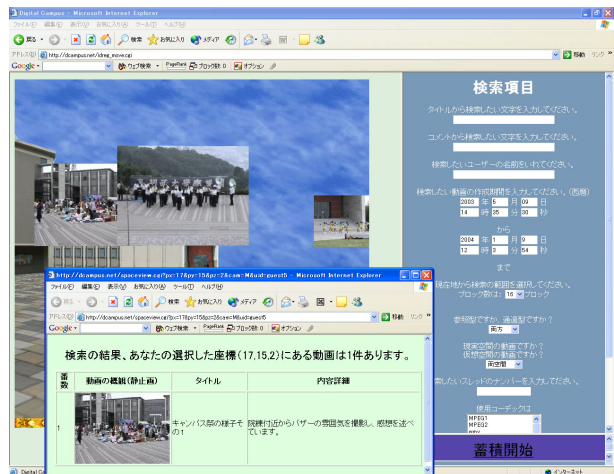


図 14 3D 仮想空間を用いた動画検索

### 6.2.3 コミュニケーション機能

動画に対して意見を返す場合、まず 6.2.2 の機能を用いて意見したい動画を抽出する。次に抽出した動画の返答ボタンを押すことで図 11 の XML 生成インターフェースが開き、6.2.1 と同様に情報を入力し登録ボタンをクリックすると自動的に動画連結情報もインデキシングされる。

本システムでは動画を視聴する際に動画連結情報を用いる事ができる。利用者が動画を視聴しているシーンの前後に連携動画が存在する場合、動画再生プレーヤーの下に視聴している連携動画のタイトルや詳細等の情報を表にして出力する。動画再生中に表示された動画連結情報の中から利用者が見たい動画を任意に選択する事で再生中であっても選択した動画が再生される。更に選択した動画の再生中も新たに連結された動画がある場合は同様に出力される。このように動画を連結させて順番に再生することでコミュニケーションが可能となる。このコミュニケーション機能の流れを図 15 に示す。

## 7. 関連研究

### 7.1 動画検索に関する研究

[1] はウェアラブル機器で撮影した長期間の体験映像を視点移動のパラメータを利用してショットごとに分割している。さらに分割したショットに対し、同期記録した撮影者の脳波で評価付けを行い、撮影者が興味を持ったシーンの抽出を実現して

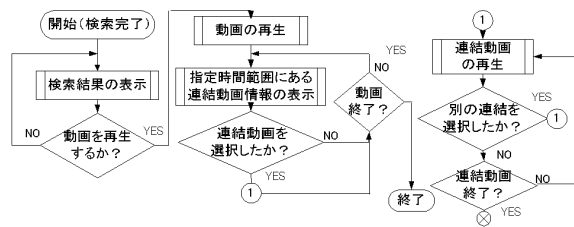


図 15 コミュニケーション機能による視聴の流れ

表 4 本研究と動画検索の関連研究との比較

	動画メタ情報	位置情報	注視点検索	3D 仮想空間
[2]	x		x	x
[3]	x			x
[4]	x		x	x
本研究				

いる。[4] はウェアブル機器で撮影した映像に、音声によるアノテーション情報や位置情報を MPEG-7 で索引付けしており、検索には対話的な手法を提案している。[2] は多視点映像の撮影範囲を時系列データとして蓄積し、そのデータを比較することでより良い映像を検索している。またキーワードの出現密度に注目した検索手法も提案している。[3] は GPS と電子方位計を用いて、デジタル写真を注視点の位置をキーとした検索で抽出しており、インターフェースには地図を利用している。[5] は断片的な内容記述と記述間の関連性に基づいて映像を記述し、自然言語によるキーワード列で検索を行っている。

本研究との違いは、コンテンツ情報の代りあるいは補助として位置情報を用いている点では本研究と共通しているが、インターフェースに 3D 仮想空間を用いている点で異なっている。また、位置情報だけでなく動画メタ情報も利用している点も他の研究と異なる点の 1 つである。

## 7.2 仮想空間を利用したコミュニケーション

本研究では、位置と方向等でインデキシングを行なった動画を実空間を再現した 3D 仮想空間に配置することで視覚的表現の出力を行い、過去の動画の視聴を容易にしている。さらにこれらの動画を用いて対面型時空間非同期コミュニケーションを行なう事を目的としている。[7] は特定の場所と時間でのみアクセス可能な仮想オブジェクトである SpaceTag を用いたシステムを作成し、あたかも張り紙を貼るように実空間に添付する事で時空間の壁を超えた情報アクセスを実現している。[8] では実環境を定点観測し、その結果のウェアネス情報を同一実環境に持続的に提示する Optical Stain というシステム構築し過去のコミュニケーションを実現している。[9] は遠隔勤務において、映像と映像表示部の近傍にメッセージ通信機能を持たせ、コミュニケーションの「場」としてメディア空間 e-office の構築を行いインフォーマルニーズの解決を狙っている。[10] は遠隔教育のコンセプトとしてインターネット上にバーチャルリアリティ技術を用いた学校としてバーチャルスクールを構築し、実際の学校のような利用者同士での時空間を共有を実現している。

動画検索における関連研究との比較を表 4 に、仮想空間を用いた動画コミュニケーションの関連研究との比較を表 5 に示す。

表 5 本研究とコミュニケーションの関連研究との比較

	動画の連携	時間非同期	位置情報	3D 仮想空間
[7]	x			
[8]	x		x	x
[9]		x		x
[10]	x	x	x	
本研究				

これらの表の各項目に対し、各研究が利用していると考えられる場合は、確実に利用しているとはいえないが近いと思われる場合、利用しているとは考え難い場合を x として比較した。

## 8. まとめと今後の課題

本稿では、動画メタ情報、コンテンツ情報、位置情報を利用して少ない手間で作成した動画にインデキシングを行うだけでなく、位置情報を用いてシーンを抽出する場合は、インターフェースに 3D 仮想空間を利用することで、あたかも実空間を探索するような感覚で動画を抽出できる手法を提案した。更に、本システムの応用例として索引情報に動画連結情報を追記することで、時空間非同期の対面型コミュニケーションが実現できることについて述べた。

今後、位置情報をクラスタリングしておくことで、検索時間を短縮させるといったことが必要である。

## 文 献

- [1] 相澤清晴, 石島健一郎, 椎名誠, “ウェアラブル映像の構造化と要約: 個人の主観を考慮した要約生成の試み”, 電子情報通信学会論文誌 D-II Vol. J86-D-II No.6 pp.807-815 (2003)
- [2] 中西吉洋, 廣瀬龍男, 田中克己, “多視点映像データの概念モデリングと代表映像の検索”, 情報処理学会論文誌: データベース Vol.43 No. SIG 5(TOD 14), pp.54-65 (2002)
- [3] 藤田秀之, 有川正俊, 岡村耕二, “注視点を考慮したデジタル写真の検索インターフェース”, 電子情報通信学会データ工学ワークショップ (DEWS2003) 6-P-03 (2003) 付録 (プログラム)
- [4] 兵清弘, 天笠俊之, 吉川正俊, 植村俊亮, “MPEG-7 を利用したウェアラブルカメラ映像の索引付け手法”, 電子情報通信学会データ工学ワークショップ (DEWS2002) C2-17
- [5] 是津耕司, 上原邦昭, 田中克己, “時刻印付オーサリンググラフによるビデオ映像のシーン検索”, 電子情報通信学会論文誌 Vol.39. No.4 pp.923-932 (1998)
- [6] 上原邦昭, 麻植周, 堀内直明, “ストーリーを考慮したビデオの内容記述モデル”, 電子情報通信学会論文誌 Vol.39. No.4 pp.943-953 (1998)
- [7] 森下健, 中尾恵, 垂水浩幸, 上林弥彦, “時空間限定オブジェクトシステム SpaceTag: プロトタイプシステムの設計と実装”, 電子情報通信学会論文誌 Vol.41 No.10 pp.2689-2697 (2000)
- [8] 白井良成, 大和田龍夫, 亀井剛次, 桑原和宏, “実環境指向のウェアネス情報とその提示手法”, 電子情報通信学会論文誌 Vol.43 No.12 pp.3653-3663 (2002)
- [9] 榊原憲, 加藤政美, 田處善久, 宮崎貴識, “メディア空間による分散勤務者のコミュニケーション支援システム「e-office」”, 情報処理学会論文誌 Vol.43 No.8 pp.2821-2831 (2002)
- [10] 白戸仁博, 佐々木整, 竹谷誠, “バーチャルリアリティ技術を用いた遠隔教育システムの開発と適用”, 電子情報通信学会論文誌 D-I Vol. J83-D-I No.6 pp.619-626 (2000)
- [11] R.Nishide, H.Hiura, S.Kitamura, M.Ohnishi, N.Morita, S.Ueshima, “Digital Campus: Experiencing the Lifestyle In the Realm of Virtual World”, ED-MEDIA2003 Vol.2003 No.1, pp.265-272 (2003).