

複数のカメラを用いた映像データベースにおける 位置情報等を用いた映像の再構成機能

何 書勉[†] 久保 広樹[‡] 横田 裕介[†] 上林 弥彦[†]

[†] 京都大学情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

[‡] 京都大学工学部 〒606-8501 京都市左京区吉田本町

E-mail: [†] {shumian,yy,yahiko}@db.soc.i.kyoto-u.ac.jp, [‡] hkubo@db.soc.i.kyoto-u.ac.jp

あらまし これまで我々はミーティングなどの状況を複数のカメラを用いて同時に撮影し、あとから任意の視点で再生することによって過去の映像空間のウォークスルーを実現するシステムを開発してきた。本稿では、撮影時にセンサ等を用いて記録した参加者の位置情報及び映像に記録された音声データなどを基に検索用のインデックスを生成し、動き回る参加者や会話が行われているグループを選択するための検索機能について述べる。これによって大量の映像データから必要な部分のみを選択し映像を再構成することが可能になる。この機能を用いることで、例えば撮影データから特定の人物を追う映像を生成することや、興味のある会話が行われているグループの映像のみを取り出すといった利用が可能になると考えられる。

キーワード マルチメディア DB, 検索

Reorganizing Video Data using Location Information on a Video Database System with Multiple Cameras

Shumian HE[†] Hiroki KUBO[‡] Yusuke YOKOTA[†] Yahiko KAMBAYASHI[†]

[†] Department of Social Informatics, Graduate School of Informatics, Kyoto University

[‡] Faculty of Engineering, Kyoto University

Yoshidahonmachi, Sakyo-ku, Kyoto-shi, 606-8501 Japan

E-mail: [†] {shumian, yy, yahiko}@db.soc.i.kyoto-u.ac.jp, [‡] hkubo@db.soc.i.kyoto-u.ac.jp

Abstract We have been developed the system to create virtual space and to realize walk-through. The system records meetings or parties as multiple video stream data with omni-directional sensors and offers walkthrough of the past video space realized by replaying with arbitrary viewpoints later. In this paper, indexing functions based on location data of moving participants or voice data recorded in the video data are described and a retrieval function to select a part of video data of individual participants or a discussing group is realized. By using these functions, only necessary data among large quantities of video stream data can be selected and reorganized. It enables us to generate videos that trace a specific person or select only the video of the group in which interested conversation is performed.

Keyword Multimedia DB, Retrieval

1. はじめに

近年、技術進歩によってディスクなどの記憶容量は以前に比べて格段に大きくなり、またデータの圧縮技術の発展もあり、まずすべてのデータを蓄えておいてからあとで必要な部分だけを検出して利用するという方法が可能になりつつある。そこで我々は、データベースに出来るだけ多くの情報を保存し、あとでそこから必要となる情報のみを選び出す“store-everything select-later”原則に基づき、Retraxシステム[1]の開発を行ってきた。Retraxシステムとは、ミーティングの状況などを複数のカメラを用いて同時に撮影

し、あとから任意の視点で再生することによって過去の映像空間のウォークスルーを実現するものである。

とりあえずすべてのデータを保存するため、保存されたデータの中には再生時にあまり必要とされない部分がある。このような不要な部分を省き、映像を再構成することができればミーティングなどの状況を撮影して、あとでその様子を見るときなどには非常に効率のよい参照が可能となる。本研究はRetraxシステムにおいて撮影対象となっている人物をキーとした検索機能、およびその検索結果に基づいた動画像の再構成機能の実現を目標とするものである。この

機能を実現するために、撮影対象となる人物には位置情報を把握するためのデバイスをあらかじめ携帯してもらい、複数あるカメラのうち現在どのカメラが一番近いところにいるのかを映像データと同時に常に記録していく。こうして得られた位置データと、映像中の音声データなどから検索用のインデックスを作成する。このインデックスをもとにユーザが見たいと思う人物や場面、会話が行われているグループなどに注目した映像などの必要な部分だけを取り出した映像を再構成する機能を提案する。Retrax システムは撮影時には選択を行わずミーティングやイベント等の状況全体を複数の映像データとして保存するため、一般にデータ量は膨大なものになる。この映像データには利用者にとってあまり価値の無いような部分も含まれるため、いわば疎な映像データになっているということが出来る。利用者はシステム利用時に自分にとって必要な部分のみを選択しながら再生するため、映像データに対するインデックスを用意することは非常に重要である。またデータ量が膨大なため人手によるインデックス付けを行うことは現実的でないことから、ほぼ自動的にインデックスを生成する技術が必要になる。

本論文は以下のように構成される。2 章では複数カメラによる動画データベースシステムである Retrax システムについて説明する。3 章ではこのシステム上に構築する位置情報を用いた動画選択システムの構成について述べる。4 章で Bluetooth の通信に関する実験の結果について述べ、5 章でシステムの設計について述べる。6 章では関連研究について述べ、7 章でまとめと今後の課題について述べる。

2. Retrax システム

2.1. システム概要

われわれが現在開発を進める Retrax システムは、全方位センサを用いて撮影した動画による仮想空間の構築を目標としている。本プロジェクトの目標は過去のデータにアクセスすることで過去と現在のコミュニケーションを可能なものにする事である。

本研究では映像データを取るために全方位センサ (Omni-directional Sensor) を用いている。全方位センサを DV カメラに取り付けることで、単一のカメラで 360 度の視野角を持つ映像を獲得することができる。全方位センサを用いて得られたドーナツ型の画像をパノラマ状の画像に変換することで、360 度視野角のある画像および映像を利用することが可能になる。得られた全方位画像の例を図 1 に示す。本システムで現在使用している全方位センサは現大阪大学の石黒浩教授によって開発されたものである[2]。

本システムでは、仮想空間内におけるユーザの移動と視点の変更や、通常再生・早送り・逆再生などの再生モードの切り替えなどを支援するユーザインターフェース(図 2)を備えている。



図 1 ドーナツ画像とパノラマ画像

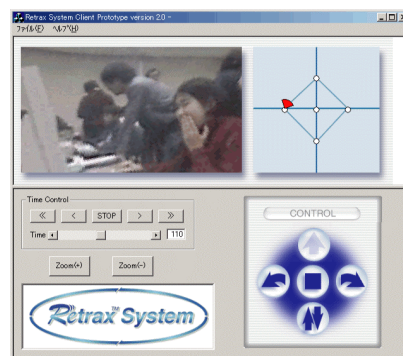


図 2 ユーザインターフェース

実際の運用例としては 2002 年 11 月 22 日に御所南小学校で行われたワークショップの撮影を行った(図 3)。パソコンを学習する生徒達の様子とそれを指導する先生の様子を記録している。ここで得られたデータから仮想空間を構築し、その仮想空間内をウォークスルーすることで過去に行われたワークショップをあとから体験することが可能になる。



図 3 撮影風景

本システムでは全てのデータを記録している。データ量の増大に伴いその中から必要な情報を検索し利用することがますます重要になってくる。有効な検索手法の構築はシステム全体の利用価値を大きく増大させるものである。そこで撮影対象者にセンサ等を取り付けておくことで、撮影時に映像データだけでなく各個人の位置データも記録しておく。この位置データと映像に記録された音声データなどを基に検索用のインデックスを作成することで、動き回る人物や注目したい場面を選択するための検索機能の実現を

目指している。これによって大量の映像データの中から必要な部分のみを選択し映像を再構成することが可能になると考える。この機能を用いた応用例としては、例えば運動会の撮影データから特定の子供だけを追いかける映像を生成したり、あるイベントの撮影データから興味のある会話が行われているグループのみを取り出した映像を生成したりするといった利用を想定している。

2.2. システム構成

2.2.1. 記録システム

全方位カメラ 本システムではある状況全体をすべて記録し、後から仮想空間をウォークスルーすることであたかも当時その場にいたかのような体験ができることを目標としている。そのためには会場全体を撮影し、参加者の行動や会話を全て記録する必要がある。そこで前述の全方位センサを DV カメラのレンズ部分に取り付け、カメラは三脚で固定する。これによってカメラの周囲 360 度視野角のある映像を撮影することが可能である。またマイクは無指向性のものを使用し、全ての方向からの音声を記録する。

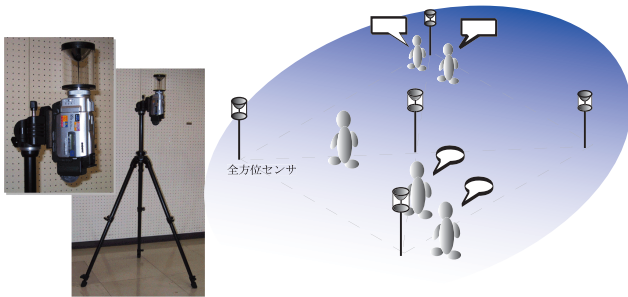


図 4 カメラの配置例

カメラの配置 イベント等の参加者は常に会場の一箇所に集中して活動するわけではなく、同時に複数の場所それぞれで独立に作業や会話が行われることがある。これらも全て記録する必要があるがカメラ一台では撮影できる範囲に限界がある。そこで対象となる空間にこの全方位センサを取り付けた DV カメラを複数台設置することで全体を記録するようにする。現在のシステムでは、正方形に四台のカメラを設置しその中心に一台設置するという合計五台のカメラで全体の状況を撮影している(図 4)。

空間全体を一様に撮影可能であるのでこの配置を本システムの標準配置とする。解像度の問題からカメラ間の距離が長すぎると再生時に遠くの様子が分からなくなってしまう。そこで標準配置での中心のカメラと外側のカメラ間の距離は 5m としている。ここでそれぞれの全方位センサとカメラの組み合わせを一つのノードと表現することになると各ノードで得られるデータは、ビデオ 1 ストリームと音声 1 ストリームである。標準配置の場合ノード数は 5 で、システム全体としてはビデオ、音声ともに 5 ストリームが記録される。しかし撮影の状況によってはこの標準配置では撮影対象となる人物の活動の妨げになるなどして撮影が

困難である場合が考えられる。将来的にはカメラとセンサの配置状況によらないシステム構築を目指している。これを実現するためにはカメラ及びセンサの配置情報が必要となる。他のカメラとの相対的な位置関係を正確に記録したカメラの配置情報をそれぞれのストリームのメタデータとして記録する。再生時にはそのメタデータをもとに再生システム側で補正することでカメラの配置状況によらないシステムの運用が可能となり、撮影できるコンテンツも増えることが期待される。

2.2.2. 再生システム

記録システムを用いて記録した映像や音声を利用して仮想空間を構築する。本システムはその仮想空間内のウォークスルーを支援するためのユーザインタフェースを備えている。

仮想空間内における再生支援 本システムでは複数カメラで撮影された映像を用いた仮想空間内のウォークスルーが可能である。この仮想空間内における利用者の移動や視点の移動、再生・早送り・巻き戻し・逆再生・停止などの再生支援を実現するユーザインタフェースを備えている。また仮想空間内の利用者の距離と各カメラの位置に応じて、音声を調節してミックスした音声データを生成しより臨場感のあるウォークスルーを実現している。

ハイパーメディア 本システム内には、時間的・空間的ハイパーリンク機能[3]が備えられており、それぞれ事前に指定しておいた時間的・空間的シーンの表示・ジャンプなどの機能を実現している。ユーザは空間のオブジェクトにハイパーリンクを貼ることやブックマークの追加を行うことで、再生時に空間的・時間的移動を行うことが可能である。これらの機能によって共有の仮想空間を個人個人でカスタマイズでき、プライベートな空間の構築が可能になる。

3. 動画像選択システムの構成

複数のカメラを用いてミーティングなどのある一つの状況全体を撮影するとき、撮影対象者がどこにいるのかという位置情報も映像データと同時に記録することが出来れば、後で再生する際にその対象者を最もよくとらえているカメラを選択して再生することが可能になる。本章ではこのような位置情報等を用いた動画像選択システムについて述べる。

3.1. 検索機能の必要性

本システムは撮影時には選択を行わずミーティングやイベント等の状況全体を複数の映像データとして保存するため、一般にデータ量は膨大なものになる。その中である利用者にとって重要な部分としては自分が映っている場面や、興味のある人物や会話が写っている場面などが挙げられる。ある利用者が見たいと考える部分は全体の中の限られた一部である。これらの映像データには利用者にとってあまり必要で無い部分も含まれていて、いわば疎な映像データになっている。そこで効率の良い検索を実現するため、イン

デックスが必要となる。インデックスには空間を区切るインデックスと時間を区切るインデックスの二種類を考える。前者は対象人物の位置情報を基にしたインデックスであり、対象人物の位置がわかればその場所に最も近いカメラを選択して再生することができる。後者は音声情報をベースとしたインデックスで、会話が行われている場面を再生することができる。またデータ量が膨大なため人手によるインデックス付けを行うことは現実的でないことから、ほぼ自動的にインデックスを生成する技術が必要になる。

3.2. 位置情報の獲得

位置情報を用いたインデックスの作成が必要である。そのためセンサを用いた位置情報の取得方法について検討する。センサは、被撮影者が存在する現実空間の情報を感知し、その情報を計算機に取り込むものである。このセンサにより、現実空間と仮想空間の三次元位置を関連付けることができる。

位置認識の方法として画像処理による点追跡、超音波や磁気による空間位置センサ、RFIDを用いた方法、Bluetoothを用いた方法などを検討した。位置情報の取得方法としてRFIDを用いる方法[6]や撮影された映像データをもとにした画像認識による人物の認識の方法[7]がある。本研究で想定しているのは、多数の人が会話をしていたり、数 m ~数十 m の範囲で動き回ったりしているという状況である。したがってある一つのシーンに多数の人物が映ることがある。画像認識を用いた場合、多数の人が同時に存在した時に、人と人が重なり合ったりすれ違ったりしたときなどに各個人を識別することが困難になると考えられる。一方通信デバイスを使う場合、対象人物にあらかじめIDを与えたデバイスを携帯させることで、個人の識別が容易に実現できるという利点がある。また、想定される移動範囲をカバー可能なデバイスが必要である。RFIDの場合、本研究で想定する状況ではIDをつけたタグを床面に広範囲に敷くのにコストがかかり、またPDA接続型のデバイスでは通信距離が数cmと短いため位置情報を得るためには利用者の動作に制限をつける必要がある。本研究ではなるべくシステムの付加を減らし、利用者の動きの制限が少ない方法で利用者の位置を知りそのデータを利用することを目的とするのでRFIDは適当でないと考えた。そこで本研究では位置情報を獲得する手段としてBluetooth[4][5]を用いた手法を提案する。Bluetoothは2.4GHzの周波数帯の無線を利用し、最大約10mの範囲内の機器と1Mbpsの速度で音声・データ両方の通信を可能にする無線技術である。さらに範囲内の機器ならば最大七つまでの機器と同時に通信することが可能である。また、赤外線通信と異なり指向性を持たず、遮蔽物があっても通信できる。仕様はエリクソン、ノキア、インテル、東芝、IBMの5社によって1998年に結成された業界団体「Bluetooth SIG(Special Interest Group)」によって進められている。

ここでBluetoothのデバイス探索には十数秒かかるので位置情報はだまかにしか分からないという精度の問題がある。しかし本システムのユーザの多くは特定の人物の移動ではなく活動に注目すると思われる。この場合移動の様子は追跡できないとしても、その後の活動の状況を把握することができれば精度の問題の影響は少ないと考えた。

3.2.1. ユーザ位置決定手法

Bluetoothの通信範囲の特性を利用した位置決定方法について考える。本システムで設置されるカメラ付近にBluetoothデバイスを設置し、参加者にもBluetoothデバイスを携帯させる。参加者の位置によって、会場に設置されたデバイスのうち携帯しているデバイスが通信可能なものと不可能なものがあり、その通信状況は履歴として時刻とともに記録される。このデータから参加者がいる時刻にどの範囲にいたのかを決定することができる。

まず単純な場合としてBluetoothデバイスが2つ配置されている場合で考える。簡単のため参加者はこの2つのBluetoothデバイスを結ぶ直線上を移動するものとする。ここでBluetoothデバイスの通信可能範囲の閾値を r とする。このとき2つのBluetoothデバイス間の距離を $2r$ 以下とすれば中央付近は両方のBluetoothデバイスから通信可能な領域となる(図5)。

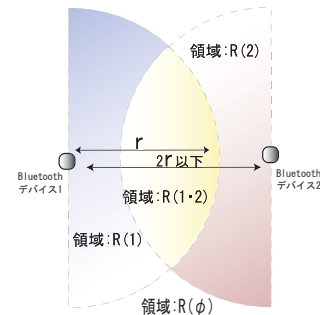


図5 デバイスが2つの場合

上記のように配置することで全体をデバイス1(Bt1)のみと通信する領域、デバイス2(Bt2)のみと通信する領域、デバイス1・2両方と通信する領域の3つの領域に分けることができる。これらの領域をそれぞれ $R(1)$ 、 $R(2)$ 、 $R(1,2)$ と表すものとする。より広範囲を移動する可能性を考慮して、Bt1ともBt2とも通信しない領域を $R(\phi)$ とする。カメラおよびBluetoothデバイスが3つ以上になったとしてもカメラとデバイスの配置を換えることでそれぞれ対応することができる。こうして位置情報を得ることでユーザは以下のような(ID, 時刻, 位置)のパラメータを持つ。

- ID: ユーザ個別のID
- 時刻: 撮影開始からの時刻
- 位置: ユーザのいる領域

これらのデータを撮影された映像データに関連付けることで再生時に検索のキーとして利用できる。例えばコンテン

ツの利用者が特定の人物に注目したいと考えた場合、その人物の ID と時刻を指定することで、その時刻に対象人物が映っている場面を選択して再生することができる。

3.2.2. 映像の構成機能

映像と同時に求めた位置情報を利用して、長時間の複数映像の中からユーザが注目したいと考える場面を選択して構成する方法について述べる。あるイベントを撮影した場合、参加者は時間とともに移動して別の場所で作業をするということがありえる。こういった場合でも参加者の位置情報(ID,時刻,位置)を参照することである時刻におよそどこにいたのかを知ることができる。これによって時間の流れにつれて対象人物が作業をしていた場面を知ることが可能になる。

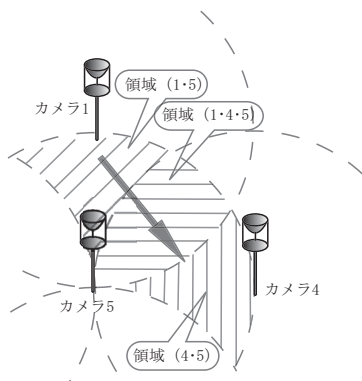


図 6 移動の例

例えば図 6 での矢印のように対象人物が移動したとする。このときははじめはカメラ 1 で撮影された映像を再生して、対象人物の移動に伴いカメラ 2 で撮影された映像に切り替えることで対象人物の活動がより良く撮影されているであろうカメラによる映像の再生が可能になる。またこの際中間地点でどちらのカメラで撮影された映像を再生するかという問題があるが、この場合は対象人物の移動の履歴を利用することで対応する。この例のように $R(1,2) \Rightarrow R(1,2,3) \Rightarrow R(2,3)$ と移動が行われていけば対象人物がカメラ 1 側からカメラ 2 側へと移動しているのであろうということが推定できる。この場合カメラ 1 からの映像では対象人物の後方からの様子を撮影していて、カメラ 2 からの映像では対象人物を前方からの様子を撮影したものであると考えられる。映像として必要とされるのは対象をなるべく正面からとらえた映像であるので、中間地点では正面が映っている状況の映像を再生するものとする。この例の場合では領域 $R(1,2,3)$ の地点ではカメラ 2 で撮影された映像を再生する。

3.3. 音声情報の利用

センサを用いることで得た位置情報だけでなく、映像に記録された音声データにも注目し、これを基に検索用のインデックスの作成を考える。[8]では映像に含まれる音情報から、リズムのみでない音楽と音声重なった場合においても各々を独立に検出する方法を提案し、応用例として空

間転回型の映像ブラウザを提案している。本研究では音声の解析は行わず音声データ中の無音区間にのみ注目する。各カメラでは映像を撮影すると同時にカメラに取り付けたマイクから音声も記録している。その各音声データについて、無音区間を検出する。会話などの連続する音声データの中の無音区間は会話と会話の切れ目であると考えられる。さらに長時間の無音区間があればその場所付近では会話は行われていないことがわかり、その部分の映像データは必要とされない可能性が高い。ミーティングやイベントなどの状況を想定した場合、重要なのは各人物がどういった発言をしたのかということである。発話ごとに区切られた部分がデータとして重要であり、そこに索引付けする。

また、これら複数のデータを比較すればある時刻に会話が行われていた場所と会話が行われていなかった場所のどちらかに分類することができる。そこである時刻に会話が行われていた場所を再生して見たいという検索が考えられる。同時に複数の場所の音声データに会話が存在した場合、音量の比較や位置情報を利用して再生する場面を選択できると考える。また同程度の音量を持つ音声情報が複数あった場合、それだけでは選択できないので、位置情報を参照してどのカメラ付近に会話しているグループがいるかおよそその位置を知ることができる。これにより見たいグループを選択して再生することが可能になる。音声情報から現在会話が行われている場所が分かれば、位置情報を用いたインデクシングに加えることでより対象人物の活動を追跡した映像再生が可能になる。

4. Bluetooth の通信実験

Bluetooth を用いた対象者の位置情報の検出についてその検出された値がどの程度有効なものであるのかを検証するために実験を行った。

4.1. 概要

複数のデバイス間の通信としては最も単純な 2 つのデバイス間での通信状態を調べた。実験状況は以下である。2 つの Bluetooth デバイス(Bt1,Bt2)を配置する。デバイス間の距離は 12m と 15m の 2 通りで実験した。Bluetooth デバイスを持って 2 つのデバイス間を進む。このデバイスから探索のクエリを投げる。一方のデバイス側(Bt1)から開始し、その場で探索を行い、探索が終了したら Bt2 側へ 1m 進む。記録するデータは(位置, 検出したデバイス, 探索時間)である。

4.2. 結果

Bluetooth デバイス間の距離を 12m, 15m にした場合それぞれ 3 回ずつ計測を行った。Bluetooth の通信範囲はその場の状況によって若干変化するが、一般的には約 10m であり今回の実験では Bluetooth の通信範囲の閾値 r を 10m であると想定している。実験の結果、表 1 のようになった。

表 1 Bluetooth 通信実験結果

距離	1 回目		2 回目		3 回目	
	デバイス	時間	デバイス	時間	デバイス	時間
1m	Bt1	11	Bt1	11	Bt1	19
2m	Bt1	6	Bt1	31	Bt1	0
3m	Bt1	13	Bt1	1	Bt1	21
4m	Bt1	11	Bt1,Bt2	49	Bt1,Bt2	33
5m	Bt1	33	Bt1,Bt2	29	Bt1,Bt2	20
6m	Bt1,Bt2	32	Bt1,Bt2	14	Bt1,Bt2	21
7m	Bt1,Bt2	31	Bt2	10	Bt1,Bt2	10
8m	Bt1,Bt2	28	Bt1,Bt2	20	Bt2	23
9m	Bt2	11	Bt1,Bt2	24	Bt2	6
10m	Bt2	0	Bt1,Bt2	36	Bt2	12
11m	Bt2	10	Bt2	18	Bt2	14

距離	1 回目		2 回目		3 回目	
	デバイス	時間	デバイス	時間	デバイス	時間
1m	Bt1	48	Bt1,Bt2	31	Bt1	11
2m	Bt1	11	Bt1	19	Bt1,Bt2	31
3m	Bt1,Bt2	12	Bt1,Bt2	26	Bt1,Bt2	23
4m	Bt1,Bt2	9	Bt1,Bt2	29	Bt1,Bt2	26
5m	Bt1,Bt2	14	Bt1,Bt2	13	Bt1,Bt2	14
6m	Bt2	30	Bt1,Bt2	12	Bt1,Bt2	25
7m	Bt2	11	Bt1,Bt2	24	Bt1,Bt2	18
8m	Bt2	0	Bt1,Bt2	52	Bt1,Bt2	12
9m	Bt2	18	Bt1,Bt2	31	Bt1,Bt2	31
10m	Bt2	26	Bt2	12	--	10
11m	Bt2	4	Bt1,Bt2	17	Bt2	11
12m	Bt2	21	Bt1,Bt2	2	Bt2	4
13m	Bt1,Bt2	37	Bt2	11	Bt2	1
14m	Bt2	10	Bt2	20	Bt2	11

4.3. 考察

結果をみると確かに通信する状況は Bt1 のみ、両方、Bt2 のみと移り変わっている。しかしその精度は一定ではなく 2m 程度のずれが生じる。これは 2 通りの実験でデバイス Bt1 を後退させることで 12m から 15m へと変更したことによってデバイス Bt1 が周囲から受ける影響が変化したのではないかと考えられる。常に一定の値では無いという状況は実際の計測を行う場合は望ましくない。そこで Bluetooth デバイスを覆い隠すことで電波を弱め通信距離を短くすることを試みた。通信距離の短縮はより精度の高い領域設定を可能にし、またカメラ間の距離の短縮はより鮮明な映像の撮影を可能にする。実際にアルミホイルで覆うと通信距離が短くなり、多量に覆うと通信ができなくなった。しかし覆う量と通信距離の減少にはっきりとした相関関係を見出すことはできず、通信距離を正確に制御することは難しい。ある程度どんな状況であっても対応できるシステムにすべきであり位置情報の信頼性を高めることは今後の課題である。

今回の実験では探索に要した時間の平均は、それぞれ 17.5 秒、18.5 秒であった。ここで人が日常的に歩く速度を時速約 4km と考えれば、18 秒あれば大体 20m の距離を歩行可能である。これでは一つの地点で探索が完了して次の探索が終わる前には既にまったく別の場所まで移動してしまうことが起こり、各領域 R を通り過ぎてしまう。対象人物が

長時間移動を続ける場合は途中で領域を通り過ぎ、位置情報に漏れができてしまう。このため現段階のモデルでは対象となる人物の動きそのものを追跡するようなコンテンツの作成は困難である。しかしミーティングやイベント等の状況を考えると対象人物が常に動き回っているということは無く、また必要な情報は動いている場面そのものよりもある領域で止まって活動している状況のほうが多いと考えられる。実際あるタイミングで動作があり、その後場所を換えてまたしばらくそこで会話や作業が行われる場合、重要なのは会話などの活動内容であり移動の途中経過ではない。したがって人の動作ではなく作業を追跡することを考えた場合、この位置データを用いたインデックスにより必要なデータを選択することは可能であると考えられる。

5. 動画像選択システムの設計

まず位置情報と映像データの関連付け手法について述べ、次に再生機能の設計について述べる。

5.1. 位置情報と映像データの関連付け

ある特定の人物が常に一つのカメラからのみ撮影されていることはなく、多くの場合同時に複数のカメラで撮影されている。この中から利用者の要求に合う映像を再生するためには、どのカメラで撮影された映像を選択・再生するかが重要である。

以下で、再生するカメラを位置情報から決定する手法について説明する。撮影された人物を A,B とし、複数のカメラをそれぞれ CAM1,CAM2,... とする。

5.1.1. カメラ選択決定表

3.2.2 節で述べたように、対象人物の位置情報と履歴情報から選択するカメラを決定することができる。デバイスの配置によって全領域が n 個の領域に分割されたとしても、領域から領域への移動の組み合わせは高々 n^2 通りである。これらの組み合わせに対して 3.2.2 節のようにして選択するカメラを決定する。同時に、選択されたカメラが再生する方向も決定できる。あらかじめこれらのデータを表にして保存しておくことで、この表を参照して再生するカメラを決定できる(表 2)。

表 2 移動履歴と選択するカメラの対応表

移動前	移動後	選択するカメラ	再生方向
⋮	⋮	⋮	⋮
R(1,3)	R(1,2,3)	CAM2	315°
R(1,2,3)	R(2,3)	CAM2	315°
R(1,2)	R(1,3)	CAM1	45°
⋮	⋮	⋮	⋮

Bluetooth の通信状況はログとして記録しており、3.2.2 節で述べたようにこのログから各人物に対して(ID, 時刻, 位置)という履歴データが得られる。このデータと先ほどの表から、ある人物をある時間帯に最もよくとらえたカメラが決まる。これにより各人物に対して、時刻と再生するカ

メラおよび再生の方向が1対1で対応する(図7).

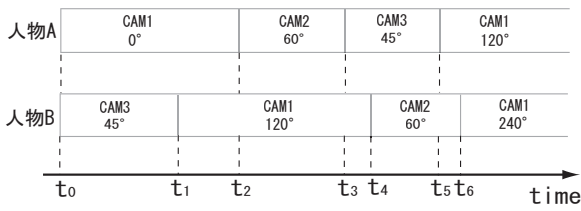


図7 人物とカメラの対応図

5.1.2. 映像データへのインデクシング

前節で得られた各人物に対する時刻とカメラの対応付けから、各カメラのストリームデータに対するインデクシングを行う(図8). 人物のIDをキーとし、再生の開始時刻と終了時刻でビデオデータを区切る. これによって人物IDで特徴付けられた断片的なビデオ区間が生成できる. ここで

$$I_i^j = (ID, s_i, e_i, d_i) \quad i=1,2,3,\dots$$

となる. ただし,

- ID: 人物のID
- s_i : 区間の開始時刻
- e_i : 区間の終了時刻
- d_i : 再生する方向

であり, j はカメラの番号を表すものとする.

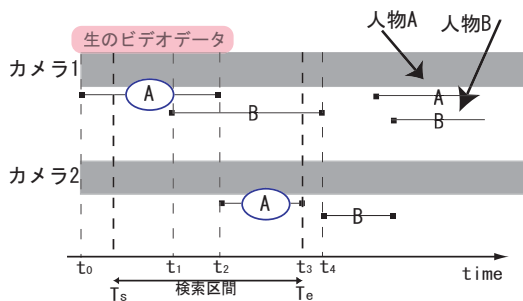


図8 インデクシングの例

5.2. システム設計

5.2.1. 検索機能

本節では5.1節においてインデクシングされた映像データへの問合せとストリームの獲得について述べる. システム利用者の要求に基づいた, 複数映像からの条件検索である. 検索におけるメタデータとパラメータは表3に示す.

表3 検索におけるメタデータとパラメータ

項目	概要
人物ID	イベント参加者固有の識別子
時間区間	映像の開始・終了時刻
再生方向	360度の映像データのどの方向を再生するか
再生モード	再生するモードの選択

検索機能は以下のように利用される.

- ① 利用者は再生方法(モード)を選択する
 - モード0:通常のウォークスルー再生
 - モード1:検索条件にマッチする場面から通常のウォークスルー再生
 - モード2:検索条件に合う映像を構成し連続的に再生
- ② 利用者は以下のメタデータを検索条件として検索要求する. 時刻の指定がない場合はデフォルトでイベントの開始・終了時刻とする. (人物ID,開始時刻,終了時刻)
- ③ システムは条件に該当する映像データを検索して以下のデータ集合を返す. (カメラID,開始時刻,終了時刻,再生する方向)
- ④ アプリケーションは検索の結果から映像ストリームを選択し,再生する. (カメラID,開始時刻,終了時刻,再生方向,再生モード)

図8において,例えば条件(人物A, T_s, T_e)で検索を行うと,丸で囲んだ(CAM1, T_s, T_2 , 方向), (CAM2, T_2, T_e , 方向)の映像区間が得られる.

またこれまで開発が行われてきたRSQL(Retrax SQL)は,我々が定義した分散映像データベースより映像と音声を取得するための質問言語である.例えば御所南で行われたイベントに参加していた山田さんの開始から30分以降の様子を見たいという場合,

```
SELECT VIDEO FROM GOSYOMINAMI
WHERE PERSON = YAMADA
AND TIME >= 30
```

のように指定すると,これを受けたシステムはインデックスを基にして

```
SELECT VIDEO FROM GOSYOMINAMI
WHERE CAM = 3 AND ANGLE = 45
AND 30 <= TIME <= 60
```

```
SELECT VIDEO FROM GOSYOMINAMI
WHERE CAM = 5 AND ANGLE = 60
AND 60 <= TIME <= 90
```

のようにして問合せを行う.

5.2.2. 再生・表示機能

ユーザからの要求に対しシステムの再生方法は3種類あり,ユーザの要求と検索の結果から適切な映像を再生する. クライアントシステムは図9のような状態を遷移する.

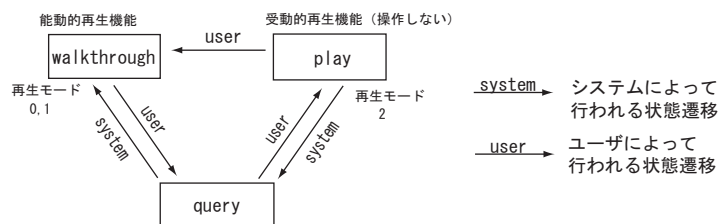


図9 状態遷移図

walkthrough 状態では仮想空間内を歩き回ることができ、play 状態では検索に基づいた映像の連続再生を見ることができる。各状態間はユーザの操作で切り替えることができる。例えば仮想空間内をウォークスルーしている最中に特定の人物に対して興味を持つことがあれば、検索条件を入力してその人物に注目した映像に切り替えるという利用方法が可能である。

6. 関連研究

複数のカメラを用いた研究は数多くなされている。金出らは2001年のSuper Bowlにおいて“Eye Vision”と呼ばれるシステムを開発した[11]。これは複数のカメラで一つの対象物をとらえることで再生時に自由な視点からの再生を可能にしたものである。これはfootballのスタジアムの周囲に30台以上のカメラを配置し、それらのカメラからの映像を切り替えることで自由な角度からの再生を実現したものである。

また[9]では競技場観戦以上の臨場感を実現するために自由視点3次元映像を合成する手法が提案されている。現実空間を多数のビデオカメラで撮影した映像を統合して、仮想化現実空間を構成し臨場感の高いコミュニケーションを可能にするシステムを開発している。多数の2次元画像と、それらを加工して得た3次元モデルをコンピュータビジョン技術とコンピュータグラフィクス技術で統合して、自由視点3次元映像スタジアムを開発する。

[10]は、多数のカメラで撮影された多視点映像の応用として、ユーザが興味を抱いた被写体をユーザが見ているシーンよりもより良く写している映像を検索・表示するものである。複数映像の内から一つの映像を再生し、その中で興味のある被写体があればそれに対する検索が可能であり、別のカメラで撮影されたシーンを見ることができる。

本研究との違いは、これらの研究では視点の切り替えによって臨場感あふれる映像の再生を可能にしているが仮想空間の構築を目的とはしていないことにある。またこれらは再生中の映像を見て検索対象を決定するものであるが、本研究は個々の対象人物に個別のID情報をつけることで特定の人物に注目した検索を実現しようとするものである。

7. まとめと今後の課題

本論文では、複数のカメラを用いた映像データベースにおいて、位置情報等を用いて映像を再構成する手法について提案した。Retraxシステムでは、複数の全方位センサとDVカメラを用いて撮影することで、パーティーやイベントなどの状況の映像や音声をすべて記録することが可能である。撮影されたデータのうちで必要な部分を選択し効率よく活用していくことは非常に有用である。Bluetoothを利用して得た各人物の位置情報や映像中の音声データを用いることで検索のインデックスを作成する。そのインデックスを基に再生する映像を切り替えて映像を再構成することで、ユーザは自分の見たい人を見ることができ、興味のある場

面を体験することができる。

しかし映像の圧縮や転送速度など技術的な課題もあり、解像度の向上などが必要である。また位置情報に関しても、Bluetoothの通信範囲は約10mの範囲ではあるが使用する場所やその場の状況によって影響を受け通信範囲が変化するなど精度が課題である。さらにデバイスを探査するのに時間がかかってしまうことも検討すべき課題の一つである。探索に約18秒かかってしまうという現状では対象人物の活動はそれぞれの移動後の地点で確認することはできても移動の様子や頻繁に動きつづける対象の場合は追跡することが出来ない。今後カメラやデバイスの配置の調整による対応と、Bluetooth以外の位置検出の方法なども検討していく予定である。

文 献

- [1] 何書勉, 横田裕介, 石黒浩, 上林弥彦, “全方位センサによる分散映像データベースのアーキテクチャとプロトコル設計,” 第13回データ工学ワークショップ (DEWS2002), 2002.
- [2] H.Ishiguro, “Development of Low-Cost Compact Omnidirectional Vision Sensors and their applications,” Int. Conf. Information systems, analysis and synthesis, pp.433-439, 1998.7.
- [3] 何書勉, 横田裕介, 上林弥彦, “分散映像データベースによる仮想空間における権限付きハイパーリンク機能,” 情報科学技術フォーラム (FIT2002)
- [4] The Official Bluetooth Wirelees Info Site, <<http://www.bluetooth.com/>>
- [5] 宮津和弘, テクノロジー解体新書 Bluetooth 技術解説ガイド, 河合一彦 (編), (社)株式会社リックテレコム, 東京, 2001.
- [6] 椎尾一郎, “RFIDを利用したユーザ位置検出システム,” 情報処理学会研究報告 00-HI-88, pp45-50, May, 2000.
- [7] 十河卓司, 石黒浩, モーハン M.トリベティ, “複数の全方位視覚センサによる実時間人間追跡システム,” 電子情報通信学会論文誌, vol.J83-D-II, No.12, pp.2567-2577, 2000.
- [8] 南憲一, 阿久津明人, 浜田洋, 外村佳伸, “音情報を用いた映像インデクシングとその応用,” 電子情報通信学会論文誌 J81-D-II, No.3, pp529-537, 1998.
- [9] 北原格, 大田友一, “大規模空間を対象とした自由視点映像生成のための3次元形状表現手法,” 日本バーチャルリアリティ学会, Vol.7, No.2, pp177-184, 2002.
- [10] 秦淑彦, 廣瀬竜男, 中西吉洋, 田中克巳, “カメラメタファによる多視点映像の検索,” 情報処理学会, Vol.42, No. SIG04-003, 2001.
- [11] Caranegie Mellon Goes to the Super Bowl, <<http://www.ri.cmu.edu/events/sb35/tksuperbowl.html>>