

被写体追跡による全方位映像のメタデータ生成

何 書勉[†] 田中 克己[†]

[†] 京都大学大学院情報学研究科社会情報学専攻, 京都市左京区吉田本町

E-mail: †{shumian,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 360度の視野角を持つ全方位映像から、特定の被写体に注目する画面を生成する際に、被写体と全方位カメラ間の距離と相対角度、再生時間などを示すメタデータが必要である。我々は、このような全方位映像における映像解析に基づいた被写体の追跡によるメタデータの自動生成手法を提案した。注目する被写体にキーワードを与えると、位置検出装置を用いることなく、システムは映像から被写体の移動軌跡を取得し、メタデータを生成する。このメタデータを用いて、全方位映像から特定の被写体の映像を検索することができる。本論文ではまず、全方位映像の時間的、空間的特性を述べ、全方位映像におけるフレーム、ショットの基本概念を定義し、さらに被写体追跡とメタデータ生成の手法を提案する。

キーワード マルチメディア処理、メタデータ管理、デジタルアーカイブ、全方位映像、ビデオ代数

Generation of Metadata for Omni-Directional Video Data by Tracking Moving Subject

Shumian HE[†] and Katsumi TANAKA[†]

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

Yoshidahommachi, Sakyo-ku, Kyoto 606-8501, Japan

E-mail: †{shumian,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract The metadata that indicates distance and relative between subjects and omni-directional cameras and playback time of video is necessary for generating the shots that show specific subjects from omni-directional video with the viewing angle of 360 degrees. We proposed the automatic generation technique of the metadata by tracking the subjects based on the image analysis of omni-directional video. When it gives the keywords to the subjects that are to be shown, the movement tracks of subject are acquired from the image, and the system generates the metadata without using positional detection devices. The image of specific subjects can be retrieved from omni-directional video by using this kind of metadata. In this paper, we describe the temporal and spatial characteristic of omni-directional video, and then define some basic concepts of the frame and the shot in omni-directional video. Finally, we propose proposes a technique of the subject tracking and the metadata generation.

Key words multimedia processing, metadata management, digital archives, omni-directional video, video algebra

1. はじめに

1826年に世界最初のカメラが誕生して以来、撮影設備は目覚ましい発達を見せた。静止写真から動く映像への進化は1895年頃に遂げたあと、人々はよりより広い範囲を撮影するために、広角レンズや魚眼レンズと次々と開発した。しかし、魚眼レンズの視野角は最大180程度で、その視野角を遥かに超えるレンズは20世紀70年代に登場した全方位レンズであった[6]。全方位レンズは全方位視覚センサとも呼ばれ、常に360度の視野角を持っている。全方位レンズを取り付けたカメラを全方位

カメラと呼ばれ、それを用いて撮影されたカメラ周囲の景色はドーナツ型の画像に写る。ドーナツ型の画像をパノラマ形式に展開して四角形の部分を切り出して表示すると、普通のカメラで撮られる映像と同じぐらいの効果が得られる。このような全方位カメラで撮影された360度の視野角を持つ映像のことを全方位映像と呼ぶ。全方位映像から特定の人物の映像をどのように検索するかは、我々の研究課題である。

映像データはいわゆる連続データ (contiguous data) で、その検索は文書データの検索と大きく異なっている。文書データの検索は、検索対象となる文字列がファイルのどのあたりにある

ということが比較的わかるが、映像データの場合、どの部分が目的シーンにあたるか決めるのは大変である。内容に基づく映像検索の多くは、フレーム、ショット、シーンといった映像データの構成単位にキーワードを与えることによって、検索キーワードによる検索を可能にしている。このような検索キーワードはメタデータという。メタデータは一般的には「データに関する（構造化された）データ」と定義されている [1].

普通のカメラを用いて撮影を行う際、カメラマンやカメラ制御システムは被写体の動きを追いながら撮影の角度やズームを調節し、撮影のタイミングを決める。すなわち、どのようなものに注目するかは撮影時にカメラを操作する人（カメラマン）、あるいは制御システムが決定するわけで、ビデオ撮影の場合も断続的に撮影を開始したり一時停止したりして、ショットというビデオ断片を生成する。撮影時に被写体の識別子となる ID をビデオ情報に付加すれば、被写体に対してインデックスが作れる。このような映像には、カメラワークのパンやズーム、カメラ運動のドリーやクレーン、それにカット・ディゾルブ・ワイプなどのショット切り替えなど編集効果を加え、ストーリー性を持つものが多い。一方、全方位映像の場合、監視システムなど長時間自動撮影での利用面が多く、同時に複数の人間が撮られ、注目する角度を変えれば、違うストーリーが繰り広げられるように、多くのイベントが同時に起きているのである。また、長時間に撮影された全方位ビデオから、特定の人物の部分を検索するには、その人物が写る映像の経過時間だけでなく、円形の映像上に現れるか角度や、被写体の大きさなど情報も必要である。全方位映像の被写体の位置情報が映像検索にとって、非常に大事である。しかし、被写体がずっと動いているため、角度と大きさは常に変わると考えられる。そこで、全方位映像のメタデータはどうであるべきかは非常に重要な課題になっている。

本研究では被写体が全方位映像内に現れるタイミングや位置などの情報を、映像解析の手法で検出する手法を提案している。

本研究では、全方位映像における問題点として次の 2 点に注目した。

- データ量が非常に膨大なもので、すべてを閲覧することは難しい。
- 全方位映像を閲覧するには、ドーナツ型の映像をパノラマ映像に展開する特殊なアプリケーションが必要で、ビデオの基本構成となるフレームは生成時に決定されるが、その選択肢は無数ある。

これらの全方位映像の性質に基づき、移動物体の移動軌跡を追跡してメタデータを生成する手法を提案する。本稿は以下のように構成される。2 章において本研究に関連する研究について述べ、3 章で全方位映像の特性とそのモデルについて述べる。4 章では全方位映像のメタデータとその生成の方法について提案する。まとめと今後の課題は 5 章で述べる。

2. 関連研究

2.1 ビデオインデックシングモデル

膨大なデジタル映像データの発生にともない、映像を問い合

わせたりネットワーク経由で配信したりする技術は注目されている。映像データは連続データであるため、意味のある単位やオブジェクトが映像データの時間、空間上に分散している。映像解析、音声認識、自然言語理解などの方法で映像データを分析しインデックスをつける研究は行われてきたが、それらの技術は特定の映像の断片にだけ有効であり、意味のある単位が至るところに現れるような映像に対しては難しいと考えられている。映像データのモデリングと構造化に対する研究の中で、時空間関連の基本となっているのは、Allen の時空間関連モデルである [2][3]。Allen は 2 つの時空間の間には 13 の時間関連があると示し、以後のビデオデータベースの研究に大きく影響している。例えば文献 [3] では時空間に基づくモデルがビデオデータのような、時間に依存したマルチメディアデータに用いられている。

オブジェクトビデオデータベース (OVID) はインスタンス主導のビデオデータモデルである [9]。OVID では、ある 1 組の近接する区間は意味のあるものとして定義されており、これをビデオオブジェクトとしている。継承可能な属性情報と継承不可能な属性情報がそれらのビデオオブジェクトに割り当てられ、継承可能な属性情報は区間包含関連にあるビデオオブジェクト間で共有される。

2.2 被写体位置検出

被写体となる人間を追跡する映像を生成するために、位置情報を用いたインデックスを生成する必要がある。移動する人間の位置情報を認識する方法は多数開発されており、Bluetooth や RFID タグなどのデバイスを使った方法、映像認識による人物の認識の方法などがある。

[13] で提案された方式では、床などの環境側に RFID タグを複数貼付して、ウェアラブル機器や携帯機器、人が操作するカートなどに RFID リーダを取り付ける。あらかじめ位置が判明しているタグを読みとることで、ユーザの位置を検出する。一方、複数の全方位カメラを用いて仮想空間を構築するアプリケーションでは、Bluetooth デバイスをカメラ付近に設置し、被写体となる人物にも Bluetooth デバイスを携帯させ、デバイスの通信領域の重なる状態に基づいて人物とカメラ間の相対位置を計算する [14]。これらの方法による位置検出の精度はデバイスの特性に依存しており、Bluetooth を用いた場合、人物の空間座標を求めることができず、大まかな範囲しか取得できないという大きな制限がある。RFID を使用する場合、検出される移動人物の座標は離散的で、全方位映像における円滑な追跡は難しいと考えられる。

映像認識で被写体となる人物の位置を検出する方法として複数台のカメラで被写体を撮影し、それぞれのカメラで捉えた被写体の角度に基づいて計算するといったステレオ視がある。

[16] はパン・チルト・ズームカメラ (FV-PZT カメラ) [17] を使い、背景差分によって対象検出・追跡するシステムが提案されている。視点固定型パン・チルトカメラを用いて移動対象の追跡システムを実現する場合は、予め撮影しておいた全方位全天空パノラマ画像から、現在のカメラの視線方向に対応した画像を生成し、それとカメラから撮られた画像との差を取ること



図1 全方位センサと DV カメラ

によって移動対象を見つける。

実時間人間追跡システム [11] は映像認識による人物の位置を検出することが可能である。複数台のカメラ [4] [5] が違う方向から同時に同じ人物を捕らえ、人物の三次元的な空間座標をリアルタイムに計算できる。そして、複数の人物の位置推定を行うために、3 眼ステレオ視 [12] を拡張した N 眼ステレオ視が開発された。

これらの方法は、いずれも被写体が実空間内の三次元位置情報を取得を目的としている。一方、全方位映像の場合、全方位カメラを中心とした全方位映像空間が考えられ、実空間内の位置情報より、全方位映像空間内の独自の位置情報を用いたほうが考えられる。

2.3 ビデオ代数

代数ビデオモデル [8] [10] は映像データの階層化に基づいている。しかし、代数ビデオモデルでは、簡単な階層化と違って、同じ映像データ間との継承関係を記述できる。結合 (union)、共通集合 (intersection)、連結 (concatenation) などの代数演算で映像インターバルの関係を示す。階層関係の親ノードはその子ノードの文脈関係を表す。この階層関係では、同じ映像データから、違う文脈関係をもつ映像の呈示を行うことができる。代数ビデオモデルの研究もまたクエリーモデルよりアノテーションモデルに基づいている。どのようなクエリーにも対応できる汎用性のあるアノテーションをしなければならない。本研究もこの代数ビデオモデルに基づき、全方位映像のフレーム、ショットを定義するかを独自の視点から考えている。

3. 全方位映像の再生

3.1 全方位映像の特徴

全方位映像は全方位センサ (図 1) と DV カメラのセットで記録される。全方位センサを利用して撮影された画像や映像は、ドーナツ型をしているため、使用するアプリケーション上でパノラマ形式に変換する必要がある (図 2)。もとの画像や映像はドーナツ型をしている都合上、ドーナツの中心に近くなるほど展開後該当する部分の解像度が低くなる。また、遠くにある物体を写すために、展開後の画像の一部を拡大してズーム処理をしているが、解像度の都合上、全方位センサから 10 メートル

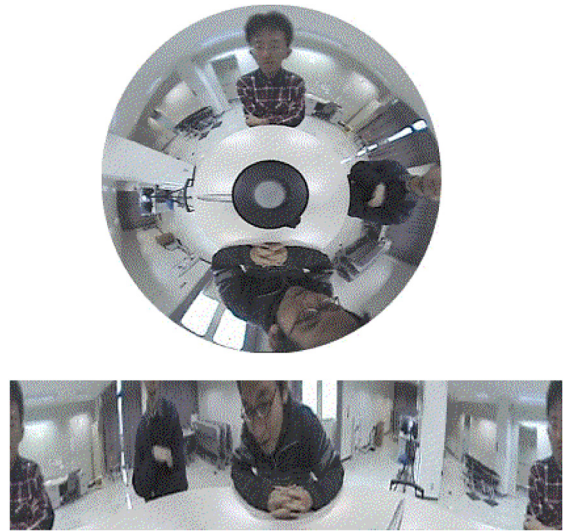


図2 全方位センサで記録したドーナツ型映像と展開後のパノラマ映像

ル以上離れた物体は鮮明に映らないのが現状である。全方位センサの映像をハイビジョンカメラで記録することで改善できると考えられるため、以下の議論では解像度による撮影範囲の制限を考慮しない。

全方位映像の特徴を普通の映像と比べながら、特徴を以下のようにまとめる。

- 放送用と家庭用のビデオカメラ (以下、普通のカメラと呼ぶ) で記録した映像では、各フレームの縦横比はほとんど 3:4 で、ワイドテレビやハイビジョンの場合は 9:16 となる。一方、全方位のパノラマ映像の縦横比は、2:9 である。全方位映像を利用するために、パノラマ型の映像から適切な大きさの長方形部分を切り出す必要がある。

- 普通のカメラで撮影するとき、連続した画面のひとつぎりのことでビデオカメラの録画ボタンを押してから切るまでを 1 カット (ショット) といい、意味のあるカットが集まって出来たひとつの場面のことをシーンという。一方、全方位センサで全方位映像を記録する場合、機材設置の都合上、ビデオカメラの録画ボタンを頻繁に押ししたり切ったりすることはなく、通常撮影開始から終了まで撮りっぱなし状態である。そこで、全方位映像のショットとシーンという概念は撮影する際に存在せず、パノラマ映像から適切で意味のある部分を切り出す際に生じることになる。

- 普通のカメラの撮影では、被写体の遠近と角度に応じてズームをしたりパンをしたりする必要がある。一方、全方位映像では、ズームの拡大縮小はパノラマ映像切り出した部分の大きさとその拡大倍率で決まり、パンの操作も切り出した部分の相対位置で決められる。

3.2 全方位映像のモデリング

カメラは現実世界の景色と状況を撮影する。そこで被写体の位置情報は (x, y, z) といった三次元座標で表すことができ、位置情報検出システムの多くは、このような三次元座標を使っている (図 3)。一方、全方位カメラの視野は、カメラを中心とする円形で、全方位映像の中の被写体と全方位カメラ間の相対位

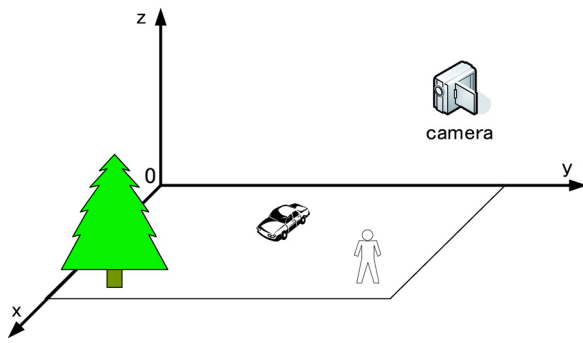


図3 三次元座標系

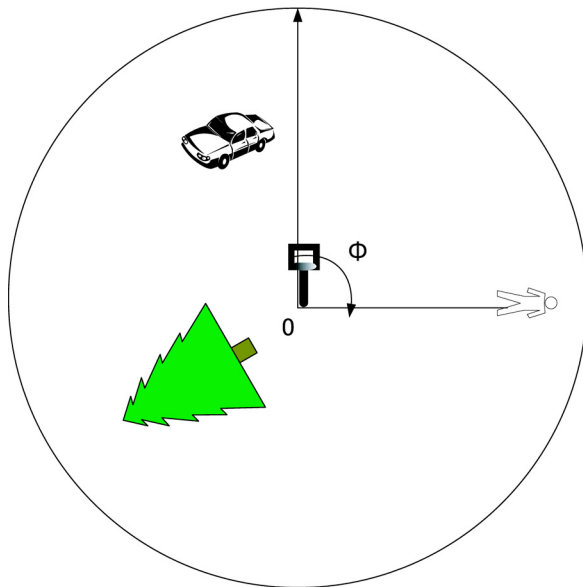


図4 全方位映像の極座標系

置は、円周角 (ϕ) および全方位カメラと被写体間の距離 (l) の極座標系で表すことができる (図4)。本研究では、全方位映像における位置情報は極座標系に基づいている。

普通のカメラで撮影された映像では、フレームは撮影した際に生成される。全方位映像では、フレームは映像を再生する際に動的に生成される。図5は全方位映像からフレームの生成を示す。上の円はドーナツ型の全方位画像を示し、その下の長方形は展開されたパノラマ型の画像を示す。ドーナツ型の画像の中の扇形部分はパノラマ型画像の長方形部分に展開される。

全方位センサ c_i で記録した全方位映像から取得するフレームは以下のようにビデオ代数で表す：

$$\text{create_frame } time(\phi, l), (\text{zoom}, \theta)$$

各要素が以下のように表される：

ϕ ：フレームの中心点と全方位ドーナツ型の映像中心点となす円周角の角度。展開後のパノラマ型映像では、 ϕ はフレームの中心点の x 軸座標とされる。

l ：フレームの中心点と全方位映像の中心点間の直線距離。展開後のパノラマ型映像では、 l はフレームの中心点の y 軸座標とされる。また、ドーナツ型映像の半径を L とすると、 $0 < l < L$ となる。

x ：フレームの左上の点の x 座標。

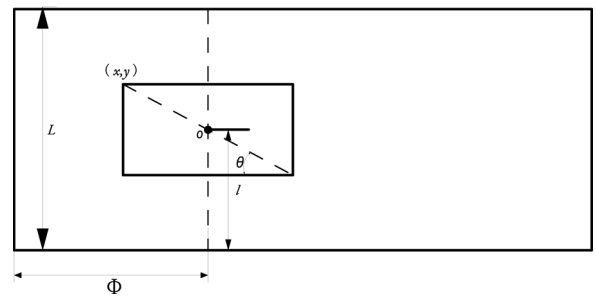
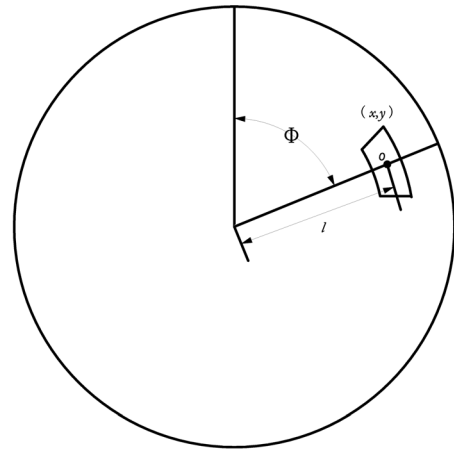


図5 全方位映像フレームの生成

y ：フレームの左上の点の y 座標。

$time$ ：フレームが全方位映像の中の時間刻印。

$zoom$ ：ズーム。フレームの大きさを意味する。

θ ：フレームの縦横比。

撮影する際に全方位センサのズームは固定であるため、映像を再生する際に、ズーム z はフレームの大きさと反比例すると考えられる：

$$\text{zoom} = \frac{1}{\sqrt{(x-\phi)^2 + (y-l)^2}}$$

すなわち、再生時の表示領域の大きさが一定の場合、全方位映像から切り出したフレームが大きいほど、ズームが小さくなる。

また、フレームの縦横比はフレームの対角線と底辺がなす角度 θ を以下のように決める：

$$\theta = \left| \frac{x-\phi}{y-l} \right|$$

360度の視野角を持つ全方位映像には、同時に複数の被写体が多くなる。複数の被写体の中から一部を呈示することを全方位映像の再生という。全方位映像は全方位カメラで撮影され、プログラムでドーナツ型の映像をパノラマ形式に展開してから、その中から四角形の部分を切り出して表示される。

ビデオ撮影の基本手法として、パン、チルト、ズームなどが挙げられる。全方位映像の場合、これらの手法はパノラマ画像から切り出す際にデジタル的に実現される。

- パン・チルト 通常、パン (pan) とはカメラを左右に振ることをいう。逆にカメラを上下に振ることをチルト (tilt) という。全方位映像の場合、パノラマ映像から四角形のフレームを

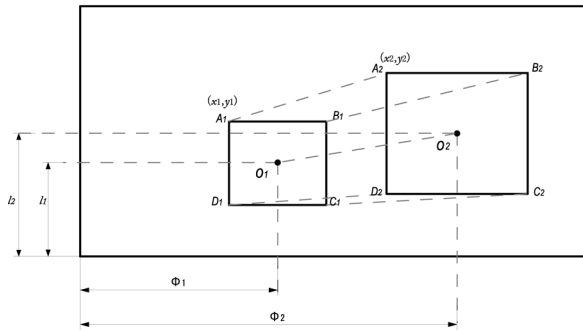


図6 視点移動によるフレーム選択範囲の変化

切り出す際に、 θ , l のパラメータを変えることによって、パン・チルトの効果を作り出すことができる。

- ズーム 全方位映像から切り出したフレームをアプリケーション上の領域に表示される。表示領域の大きさが一定であるとする、フレームの大きさが変わる場合、表示領域いっぱいに表示するために、フレーム自体を拡大したり縮小したりする必要がある。

全方位映像におけるパン・チルトやズームを変えることによって、映像を再生するときいろいろな視野角や視点を得ることができるようになる (図6)。

3.3 アプリケーションの例：Retrax システム

ここでは、我々が従来の研究で開発した複数の全方位センサを用いた撮影アプリケーション Retrax [7] について紹介する。Retrax システムでは、視点の切り替えやパン、ズームの制御は、すべて全方位映像の視聴者の操作によって行われる。全方位センサを取り付けた DV カメラを空間内に正方形に四角とその中心点に、合計5個設置することで、ミーティングなどの状況を三次元的に保存する。保存された三次元空間内を移動しながら再生することが可能になる。図7は Retrax システムのユーザインタフェースを示す。視聴者はユーザインタフェースを通して、自分が仮想空間内で前進・後退・回転などの動きを制御し、その方向と位置に応じて、見える映像が変化する。

Retrax システムは、5つの視点を持つ全方位映像を利用している。視聴者の行動に応じて、適切な全方位映像とそこから表示する部分を自動的に選択して利用者に呈示するのが特徴である。

Retrax システムでは、空間内の同じ人間や物体などの被写体は同時に複数の全方位センサに映り、全方位映像として記録される。各時点被写体の移動方向や各全方位センサの間の相対位置によって、それぞれの全方位センサで被写体はどの程度よく撮影されているかを示す捕捉状態度 [15] を計算する必要がある。ここでは、捕捉状態度の要素は被写体の大きさ、角度、時間、被写体数となっている。

大きさ

全方位のパノラマ映像から切り出したフレームの大きさはそのフレームにより呈示される範囲を決める。視聴者は被写体を

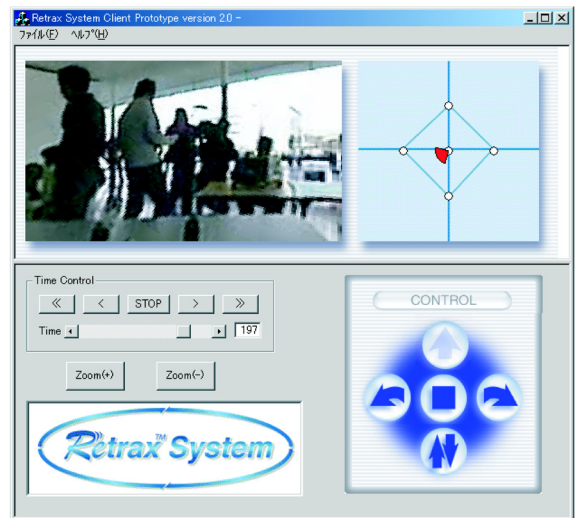


図7 Retrax のユーザインタフェース

大きく写している映像や広い範囲を撮影している映像を望む場合がある。

角度

フレームに映る被写体の角度は、全方位センサの位置だけに依存する。視聴者が自分と逆の方向から撮られた映像や様々な方向から撮影された映像を望む場合が考えられる。角度の変化は全方位センサの切り替えによって実現する。

時間

視聴者は複数の短いビデオデータの集合をイベントのハイライトとして望むかもしれない。また、一つの被写体を長い時間追っている映像を望むかもしれない。

被写体数

一本の全方位映像の中で、複数個の被写体を映っている可能性がある。しかし、前述の大きさの要素によって、一フレームの中で表示できる被写体の数は違うため、計算する必要がある。

3.4 全方位映像のアノテーション

映像データは、フレーム、カット、シーン、シーケンスの階層構造として表せる。視点全方位映像の場合、フレームは映像再生時に全方位映像から切り出して生成される。また、同じ被写体が複数個の全方位映像に写りうるため、全方位映像によって、被写体が写り出す方向が違ってくる。Retrax システムのような応用例では、被写体を見る場面や角度を視聴者に選択させるため、シナリオといったものを事前編集することはできない。実空間における三次元座標系を用いたアノテーションモデルを使って、全方位映像のアノテーションを以下のような形で指定することができる：

$$a(t, x, y, z, d)$$

それぞれの要素は次のとおりである：

t ：アノテーションの対象となる時間。

x ：オブジェクトが空間内の x 座標。

y ：オブジェクトが空間内の y 座標。

z ：オブジェクトが空間内の z 座標。

d ：オブジェクトが時間 t での向き ($0 \leq d < 360$)。特に向きがない場合、 $d = 0$ とする。

これを全方位映像の極座標系に変換すると、以下のような形になる：

$$a(t, \phi, l, d)$$

それぞれの要素は次のとおりである：

t ：アノテーションの対象となる時間。

ϕ ：オブジェクトが空間内の円周角。

l ：オブジェクトと全方位カメラ間の距離

d ：オブジェクトが時間 t での向き ($0 \leq d < 360$)。特に向きがない場合、 $d = 0$ とする。

4. メタデータの生成

4.1 全方位映像のメタデータ

全方位映像から特定の被写体の映像を検索して呈示するため、検索の対象となる被写体が全方位映像空間内の位置情報を

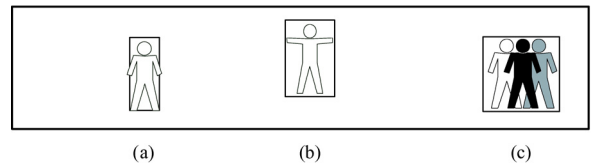


図8 全方位映像のオブジェクト



図9 被写体の映っていない背景画面

事前にメタデータとして記述しておく必要がある。この位置情報は、被写体が全方位映像に移る時刻と、その時刻に対応する位置座標および被写体の大きさから構成される。

被写体の移動する様子を全方位映像で記録すると、当然全方位映像に写る被写体も映像のフレームごとにその全方位映像空間内の位置と大きさを変えているのである。追跡映像を生成するために、被写体の位置の変化をなるべく正確に把握しなければならないが、毎秒 30 フレームもの大きなデータ量のビデオに対し、それぞれのフレームにおける被写体の位置情報を記述するのは効率が悪く、現実上不可能である。そこで、我々は被写体が全方位映像上に映る大きさと位置の変化の傾向が変わった情報だけをインデックスにする手法を考えた。

全方位映像に映る被写体はいろいろな形をしているが、その形状を詳細にメタデータに記述する必要がないと考えられる。全方位映像の場合、被写体を囲む最小の長方形を、オブジェクトと定義している。図8で示すように、(a)、(b)、(c)内の人間が被写体で、その人間を囲むそれぞれ最小の長方形がオブジェクトである。(c)内の複数の人間は各々被写体となるが、フレーム上に同時に表示される場合、複数の人間を囲む最小の長方形がオブジェクトとなる。

時間 t におけるオブジェクト o は、以下のように記述される：

$$o(t, x, y, x', y')$$

ただし、 (x, y) はオブジェクトとなる長方形の左上の点の2次元座標で、 (x', y') はその右下の点の2次元座標である。

4.2 映像解析による被写体の位置検出

オブジェクトの位置を特定するには、オブジェクトの四角形に囲まれる被写体の位置を把握する必要がある。さらに、被写体に ID をつけ、同一 ID の被写体を追跡していく。特定の被写体の位置情報を取得するには、まず画面上に被写体となる人間を選択し、システムが自動的にその人物の特徴を抽出して、映



図 10 被写体が映っている画面



図 11 背景差分によって抽出した被写体

像内のそれ以降のフレームをも解析することによってその移動軌跡と大きさの情報を取得していく。

映像解析の手順は、まず全方位カメラの設置された場所の、被写体がまったく存在しない場合の全方位映像を撮影し、それを背景として使う。それから被写体が映っている各フレームを背景と比較し、差分を求める。差分の映像にラベリングを行い、ユニークな ID をつける。さらに、被写体の領域を求め、それを囲む最小の長方形、すなわちオブジェクトの座標を求める。

背景差分やラベリングなどの技法は多くの文献で紹介されている [18][19][20]。全方位映像を再生する場合、ドーナツ型映像をパノラマ形式に展開する際にすでに大量な時間のかかる画像処理が行われると同時に、画質の劣化も非常に激しい。ここでは、カラー画像をグレースケール画像に変換し、全方位映像展開する際に生じたノイズを除去せずに認識を行うことで、処理コストの軽減を図っている。

図 9 は被写体の映っていないときにあらかじめ撮影した背景となる全方位映像の一部を示す。図 10 を背景と差分を求めると、図 10 のような被写体を抜き出すことができる。被写体指定と追跡のプロトタイプは開発中である。

5. まとめと今後の課題

本稿では、全方位映像の特性に基づいた被写体を追跡するためのメタデータの生成方法について提案した。ビデオ代数の観点から、全方位映像の基本単位となるフレームの生成を代数で定義し、さらに被写体とビデオ再生時に表示されるオブジェク

トの概念を定義した。オブジェクトの位置をパラメータに記述することで、被写体の追跡映像を生成することができると考えられる。

このような提案により、全方位センサから映像の検索・呈示ができるようになり、さらにさまざまなアプリケーションの開発が可能となる。

今回は単一の被写体の追跡について考案してきたが、複数の被写体を追跡する場合、オブジェクト間の位置関係が再生時のフレームの大きさや形状に影響すると思われる。複数の被写体を追跡する場合のメタデータの扱いは今後の課題のひとつである。また、メタデータの生成に関しては、映像解析のほかに、音声認識や自然言語解析などの手法でのアプローチも研究していきたいのである。さらに、システムの実装もこれから進める予定である。

謝 辞

本研究の一部は、《基盤 A》平成 16 年度科研費基盤研究 (A)(2) 「モバイル環境におけるコンテンツのマルチモーダル検索・呈示と放送コンテンツ生成」(課題番号: 14208036, 代表: 田中克己) および、21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」によるものです。ここに記して謝意を表すものとします。

文 献

- [1] R. Jain and A. Hampapur, Metadata in Video Databases. In ACM SIGMOD RECORD, 23(4):p27-33, Dec. 1994.
- [2] Allen, J.F.: Maintaining Knowledge about Temporal Intervals, Comm. ACM, Vol.26, No.11, pp.832-843 (1983).
- [3] Little, T.D.C. and Ghafoor, A.: Interval-Based Conceptual Models for Time-Dependent Multimedia Data, IEEE Trans. Knowledge and Data Engineering, Vol.5, No.4, pp.551-563 (1993).
- [4] Ishiguro, H., Sogo, T. and Ishida, T.: Human behavior recognition by a distributed vision system Proc. DiCoMo Workshop, pp.615-620 (1997).
- [5] Ishiguro, H.: Compact omnidirectional sensors and their applications M & E, Kougyou-Chosakai, March (1998).
- [6] Rees, D.W.: Panoramic television viewing system United States Patent No. 3505465 Apr. (1970).
- [7] 何 書勉, 横田裕介, 上林彦彦, 全方位センサによる分散ビデオ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), March (2002).
- [8] Tanaka, K., Tajima, K. and Sogo, T.: Algebraic Retrieval of Fragmentarily Indexed Video, Journal of New Generation Computing, Vol.18, No.4, オーム社, September (2000).
- [9] Oomoto, E. and Tanaka, K.: OIVD: Design and Implementation of a Video-Object Database System, IEEE Trans. on Knowledge and Data Engineering, 5, 4, pp.629-643, August 1993.
- [10] Pradhan, S., Tajima, K. and Tanaka, K.: Interval Glue Operations and Answer Filtering for Video Retrieval, IPSJ Transactions on Databases, 40, (SIG3), pp.80-90, February 1999.
- [11] 古川雅之, 神原宣雄, 港隆史, 石黒浩, View and Motion-based Aspect Model に基づく人間行動認識システム 日本ロボット学会第 20 回学術講演会予稿集, 2G16, 2002
- [12] 港隆史, 浅田稔, 選択的注意機構のための視覚運動学習による画像特徴の生成, 日本ロボット学会第 19 回学術講演会予稿集, pp.253-254, 2001.
- [13] 椎尾一郎, RFID を利用したユーザ位置検出システム, 情報処理学会研究会報告 00-HI-88, 2000.5.12, pp.45-50.
- [14] 何書勉, 久保広樹, 横田裕介, 上林彦彦: 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, 電子情報通信学会 第 14 回データ工学ワークショップ

(DEWS2003).

- [15] 中西 吉洋, 廣瀬 竜男, 田中 克己, 多視点映像データの概念モデルリンクと代表映像検索, 情報処理学会論文誌: データベース, Vol.43, No.SIG5-6 (TOD14), June, 2002.
- [16] 松山隆司 和田俊和 物部祐亮: 視点固定型パン, チルト, ズームカメラを用いた実時間対象検出, 追跡, 情報処理学会論文誌, (1999) .
- [17] 和田俊和, 浮田宗伯, 松山隆司: 視点固定型パンチルトズームカメラとその応用, 電子情報通信学会論文誌, Vol.J81-D-II, No.6, pp.1182-1193 (1998).
- [18] Zhang, H.J., Kankanhalli, A., and Smoliar, S.W., "Automatic Partitioning of Full-motion Video", Multimedia Systems Vol. 1, No. 1, pp. 10-28, (1993)
- [19] Nagasaka, A. and Tanaka, Y., "Automatic Video Indexing and Full-Video Search for Object Appearances", in Visual Database Systems II, Knuth, E., Wegner, L., Editors, Elsevier Science Publishers, pp. 113-127, (1992).
- [20] Zabih, R., Miller, J., and Mai, K., "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks", Proc. ACM Multimedia 95, San Francisco, CA, November, pp. 189-200, (1993).