

多視点全方位映像データの代数的モデリング

何 書 勉[†] 田 中 克 己[†]

複数台のビデオカメラを用いて撮影された、いわゆる多視点映像データの検索に関する研究が多くなされてきた。普通のビデオカメラを全方位視覚センサで置き換えることによって、単一のカメラで360度の視野角を持つ映像を撮ることができる。複数の全方位センサを空間内に置くことによって、多視点全方位映像を生成することになり、空間内の至るところの状況を記録することが可能になる。しかし、多視点全方位映像の空間的自由度が多く、このような映像を検索するために、既存のアノテーションモデルと映像代数モデルが対応できない。我々は、そのような多視点全方位映像の検索を実現するために、全方位映像の特性に基づき、独自の時空間アノテーションモデルと映像代数モデルを提案した。本論文では、多視点全方位映像のメカニズムを紹介し、このような映像を検索するためのアノテーションモデルと映像代数モデルを説明する。

Modeling for Multiple Perspective Omni-Directional Video Data

SHUMIAN HE[†] and KATSUMI TANAKA[†]

Many researches on retrieval of multiple perspective video data which is recorded by multiple video cameras have been done. Replacing a general video camera with an omni-directional sensor makes it possible to record images with 360-degree view-angle by using an only camera. By setting up multiple omni-directional sensor in a space, we can record situation of everywhere in a certain space as multiple perspective omni-directional video data. Because of the characteristics of multiple perspective omni-directional video data, it is difficult to be adapted by annotation models and video algebra for normal video data. In order to realize retrieval of multiple perspective omni-directional video data, we proposed a spatio-temporal annotation model and video algebra model. In this paper, we introduce the mechanism of these models.

1. はじめに

デジタル映像技術の進歩により、大量なビデオデータを生成することがより簡単になってきた。映像を編集するための手法と、検索する手法が多く提案されている。監視システムやスポーツ、コンサートの中継システムのようなアプリケーションでは、複数のカメラを用いて映像を取得している。このような撮影方式で取得された映像データは、相互に時間同期した複数のカメラで撮影されたビデオデータで、多視点映像データと呼ばれる。

前述のアプリケーションで使われるビデオカメラは、一定の視野角を持つもので、パノラマ撮影とする際に、「パン」という撮影方法を用いて被写体を追跡したり、捕捉対象を選択したりする。一台のカメラでパノラマ撮影をする方法として、1970年代から全方位センサ

による撮影方式が研究されてきた⁸⁾。本研究は、石黒浩氏（現在大阪大学工学研究科知能機能創成工学専攻教授）が開発した全方位センサを使っている⁶⁾⁷⁾。このような複数の全方位センサを用いて、多視点全方位映像を取得し、空間内の映像状況を取得し、人間の追跡や仮想空間などの構築に活用できる。これまでの研究では、会議やパーティの状況を多視点全方位映像として保存するアプリケーション Retrax システムのプロトタイプを開発してきた。

このような複数の全方位センサを用いて記録した映像データを、我々は多視点全方位映像データと呼ぶ。多視点映像の特徴として、連続メディア、同期メディア、不特定のカメラのデータ処理などが挙げられているが、多視点全方位映像の場合、同じ被写体を写す複数の映像ストリームが存在し、それぞれのストリームが違う角度と大きさで被写体を捕らえる。

また、多視点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

[†] 京都大学大学院情報学研究所社会情報学専攻
Department of Social Informatics, Graduate School of Informatics, Kyoto University

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、フレームから切り出した部分をすべて閲覧することは不可能である。

本研究は全方位映像の性質に基づき、視点選択、映像呈示のための多視点全方位映像のモデルを提案する。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章で多視点全方位映像について述べる。4章では多視点全方位映像の基本的構造を定義し、さらに代数モデリングについて述べる。まとめと今後の課題は5章で述べる。

2. 関連研究

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¹⁾²⁾。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²⁾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¹¹⁾。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。

代数ビデオモデル¹⁰⁾¹²⁾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する汎用性のあるアノテーションをしなければならない。本研究の出発点もこの代数ビデオモデルである。

多視点映像を検索するための研究も多く行われてい

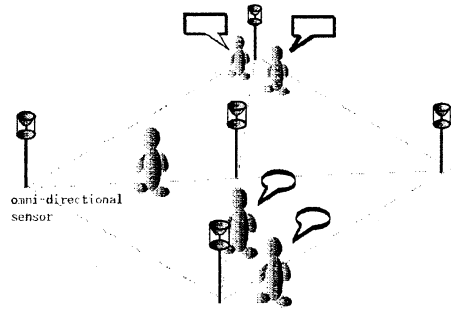


図1 全方位センサを空間内での設置例

る。多数のカメラで撮影され相互に時間同期した多視点映像を用いたサービスが文献で紹介されている¹³⁾。また、カメラメタファによる多視点映像の検索方式は提案されている³⁾⁴⁾⁵⁾。この検索方式では、ユーザが慣れ親しんだカメラ操作を用いて興味ある被写体に関する問い合わせを形成し、被写体を“よりよく写す”他のカメラの映像区間を抽出する。検索においては、カメラから得られた撮影範囲や距離計から得られる被写体距離を用いて、注目する被写体の存在領域を求める。一方、本研究は複数の全方位映像の特性から、被写体とカメラの相対位置を把握できるため、ドーナツ映像をパノラマ型に展開する際に取得した被写体の拡大倍率や円周角などの属性から、カメラの切り替えるタイミングと切り替えた後の角度とズームを求めるモデルを提案している。

3. 多視点全方位映像の性質

全方位センサ記録する映像は全方位映像で、複数個の全方位センサを使って記録した映像は多視点全方位映像である。多視点全方位映像の性質を説明するために、まずアプリケーションの例を示す。

3.1 アプリケーションの例：Retrax システム

ここでは、まず本研究で開発した複数個の全方位センサを用いた撮影アプリケーション Retrax⁹⁾ について紹介する。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 (図1)、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2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。

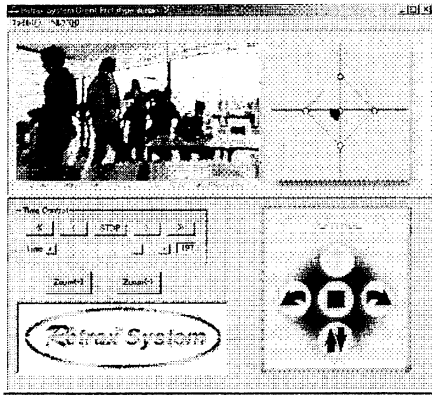


図2 Retrax のユーザインタフェース

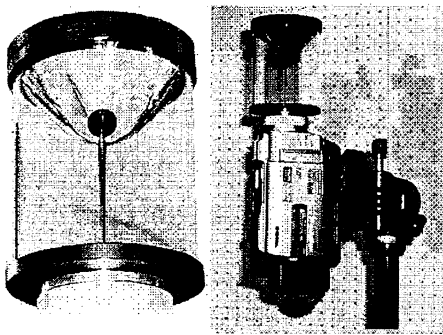


図3 全方位センサとDVカメラ

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¹³⁾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。これらの要素のモデリングについては4章で述べる。

3.2 多視点全方位映像の特徴

全方位映像は全方位センサ(図3)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4)。もとの画像や映像はドーナツ型を

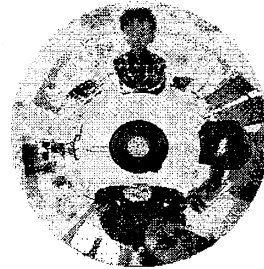


図4 全方位センサで記録したドーナツ型映像と展開後のパノラマ映像

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く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りに押し切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。

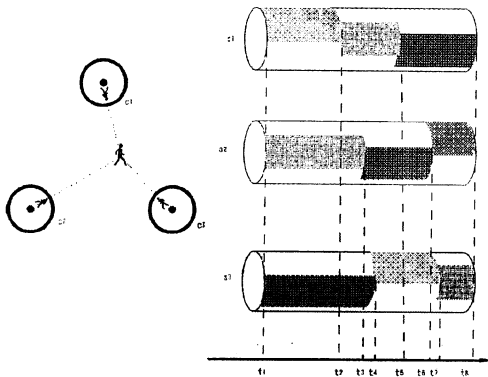


図5 被写体は複数の全方位映像に映る

- 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まる。

3.3 多視点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。多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。図5で示されたように、同じ人間の動きが複数の全方位センサで記録される場合、それぞれの全方位映像の中に映る人間の向き、大きさなどの属性が異なるため、映像ごとにアノテーションを行うことは非現実である。そこで、我々が考えたアノテーションの内容は、各全方位センサで記録した全方位映像に対するものではなく、複数個の全方位センサで記録した空間そのものを視野に入れて、オブジェクトの辞空間状態を記述するものとする。アノテーションの要素として、ある時間間隔で各オブジェクトの ID (識別子)、空間内にある三次元空間座標を時間順で記録する。また、人間などの移動する物体に関して、方向性もあるため、その「向き」の要素もアノテーションに取り入れるべきである。

アノテーションモデルは以下のようになる：

$$a(t, o.id, x, y, z, d)$$

それぞれの要素は次のとおりである：

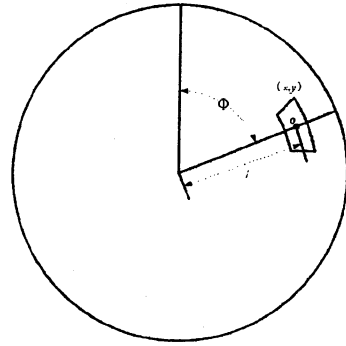
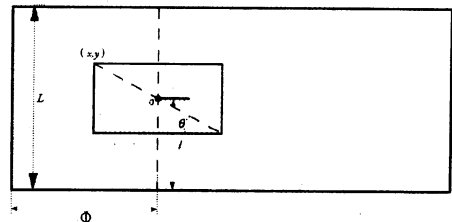


図6 全方位映像フレームの生成



- t : アノテーションの対象となる時間。
- $o.id$: オブジェクトの ID。
- 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。

4. 多視点全方位映像の代数モデリング

前章において、多視点全方位映像の応用例、および普通の多視点映像と比較してわかった特徴について述べた。この章では、多視点全方位映像の基本構造とアノテーションモデルを提案する。さらにこれらの基本構造とアノテーションモデルに基づき、多視点全方位映像の代数モデルを提案する。

4.1 多視点全方位映像のデータ構造

本研究で提案する多視点全方位映像の基本データ構造はビデオデータとそのメタデータ (全方位センサの位置、注釈情報) を含んだビデオオブジェクト間における複数の意味的関連である¹³⁾。

ビデオオブジェクトはある全方位センサによって記録されたビデオデータとそのメタデータであると考えられ、このモデルでは、ビデオオブジェクト O_i は次の要素からなる $O_i = (F_i, A_i, K_i)$ で表される。

空間内に設置される全方位センサに位置は 2 次元座

標で表される。空間内で j 個 ($j > 1$) の全方位センサが設置される場合、各センサは c_i ($1 \leq i \leq j$) で表し、それぞれの座標は (x_i, y_i) と表す。

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6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映像から取得するフレームは以下のように表される：

$$f(\phi, l, x, y, t, c_i)$$

各要素か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。

y ：フレームの左上の点の y 座標。

t ：フレームが全方位映像の中の時刻印。

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：

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

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：

$$\theta = \tan^{-1} \left| \frac{x-\phi}{y-l} \right|$$

4.2 多視点全方位映像のモデリング

4.2.1 フレームの特徴抽出

図7は時間 t_1 のフレーム $f(\phi_1, l_1, x_1, y_1, t_1, c)$ から時間 t_2 のフレーム $f(\phi_2, l_2, x_2, y_2, t_2, c)$ へ視点の切り替えの例を示す。

この例では、全方位センサ c で記録した全方位映像から切り出した連続フレームの開始時刻を t_1 、終了時刻を t_2 とすると、連続フレームオブジェクトは以下のようになる：

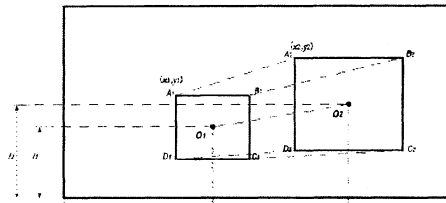


図7 視点移動によるフレーム選択範囲の変化

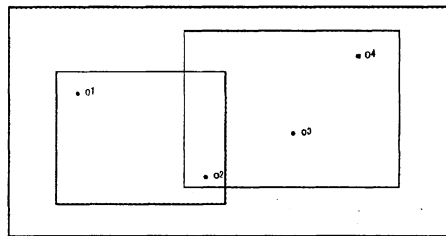


図8 被写体クラスターの特定

$$F(t_1, t_2, c) = f(\phi_1, l_1, x_1, y_1, t_1, c) \circ f(\phi_2, l_2, x_2, y_2, t_2, c)$$

$F(t_1, t_2, c)$ の各フレームは連続であり、時間 $t(t_1 < t < t_2)$ のフレームは以下のようになる：

$$f\left(\phi_1 + (\phi_2 - \phi_1) \frac{l}{(t_2 - t_1)}, l_1 + (l_2 - l_1) \frac{l}{(t_2 - t_1)}, x_1 + (x_2 - x_1) \frac{t}{(t_2 - t_1)}, y_1 + (y_2 - y_1) \frac{t}{(t_2 - t_1)}, t, c\right)$$

4.2.2 複数被写体の同時表示

全方位映像において、複数の被写体が同時に映ることがある。図8の例では、パノラマ映像の中に、オブジェクトが同時に複数個存在する。ここで、オブジェクトの集合 O 、およびそれぞれのオブジェクトが全方位映像内の x, y 座標の集合 X, Y は以下のように表す：

$$O = \{o_1, o_2, \dots, o_n\}$$

$$X = \{x_1, x_2, \dots, x_n\}$$

$$Y = \{y_1, y_2, \dots, y_n\}$$

このような場合、いくつかの問題が考えられる：

- (1) フレームの大きさと縦横比が一定（ズーム z と θ が一定）の場合、指定されたオブジェクト o_i ($1 \leq i \leq n$) と同じフレームに入るオブジェクトはどれか。
- (2) フレームの大きさが可変、縦横比が一定の場合（ズーム z が可変、 θ が一定）の場合、指定さ

れた複数のオブジェクトを同じフレームの中で表示するためには、ズームの大きさは最小どれくらいにすればいいか。

上記の問題において、オブジェクト o_j が o_i と同じフレーム内に映るかどうかは計算は、*include* 関数で行う。

$$include(\theta, z, o_i, o_j) = \begin{cases} true & \text{if } |x_i - x_j| \leq \frac{z}{\sin \theta} \cos \theta \\ & \text{and } |y_i - y_j| \leq \frac{z}{\sin \theta} \sin \theta \\ false & \text{if } |x_i - x_j| > \frac{z}{\sin \theta} \cos \theta \\ & \text{or } |y_i - y_j| > \frac{z}{\sin \theta} \sin \theta \end{cases}$$

また、 o_i と同じフレーム内に映るオブジェクトの集合は、以下のように表される：

$$O' \subseteq O,$$

$$O' = \{o_j \mid include(\theta, zoom, o_i, o_j) = true, 1 \leq j \leq n\}$$

5. まとめと今後の課題

本稿では、複数の全方位センサを用いて撮影した多視点全方位映像の代数モデルを以下のように提案した。

- ドーナツ型の全方位映像をパノラマ方の映像に展開した際、被写体が円周角、円心との距離、ズームなどのパラメータに基づき、フレームを生成する。
- 複数の被写体が同時に全方位映像の中に存在する場合、一フレーム内に映るようにフレームを決定する。
- 被写体が複数の全方位映像に映る場合、それぞれの映像内の被写体の方向と距離に基づき最適な映像を選択する。

このような提案により、多視点全方位センサから映像の検索・呈示の基本ができ、これに基づき、さまざまなアプリケーションの開発が可能となる。

今後の課題としては、視聴者の要求に答えられる映像の再構成の手法や、複数の被写体の移動方向に基づいたクラスタリングなどをモデルに取り入れた時空間検索のアルゴリズムが考えられる。今後もこれらの課題を取り込んでいく予定である。

謝 辞

本研究の一部は、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。

参 考 文 献

- 1) Allen, J.F.: Maintaining Knowledge about Temporal Intervals, *Comm. ACM*, Vol.26, No.11, pp.832-843 (1983).
- 2) 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).
- 3) 中西吉洋, 廣瀬竜男, 秦 淑彦, 田中克己: カメラメタファーに基づく多視点映像の検索, 情報処理学会研究報告, Vol.2000, No.69, pp.215-222 (2000).
- 4) Hata, T., Hirose, T., Nakanishi, Y. and Tanaka, K.: Querying Multiple Perspective Video by Camera Metaphor, *Proc. 7th International Conference on Database Systems for Advanced Applications (DASFAA)*, Hong Kong, April 18-20 (2001).
- 5) 秦 淑彦, 廣瀬竜男, 中西吉洋, 田中克己: カメラメタファーによる多視点映像の検索, 情報処理学会論文誌: データベース, Vol.42, No.SIG4(TOD9), pp.14-26(2000).
- 6) Ishiguro, H., Sogo, T. and Ishida, T.: Human behavior recognition by a distributed vision system *Proc. DiCoMo Workshop*, pp.615-620 (1997).
- 7) Ishiguro, H.: Compact omnidirectional sensors and their applications M & E, Kougyou-Chosakai, March (1998).
- 8) Rees, D.W.: Panoramic television viewing system United States Patent No. 3505465 Apr. (1970).
- 9) 何 書勉, 横田裕介, 上林弥彦, 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).
- 10) 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).
- 11) 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.
- 12) 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.
- 13) 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.