

実世界情報提供システムInfoGlassの提案

吉田忠城 森下慎次 種茂文之 片岡良治

{yosh,morisita,tanemo,kataoka}@syrinx.dq.isl.ntt.jp

NTT情報通信研究所

〒238-03 横須賀市武1-2356

ユーザが現実の世界の空間的な広がりをリアルに意識しながら、欲しい情報を直観的に検索できるハイパーメディアシステム「InfoGlass」を提案する。現実の世界の空間的な広がりをコンピュータを介してリアルタイムにユーザへ提供する単純かつ有効な方法は、ビデオカメラで現実の世界を撮影し、その映像を提供することである。ユーザがこのような形で提供される映像をいわばインデックスとして利用しながら、映像中の被写体を指示してその関連情報をデータベースから取り出せれば、マルチメディア時代のデータベースシステムが構築できる。本稿では、「InfoGlass」の実現に向けての我々のアプローチを示す。

Proposal of the Hypermedia System to Access Real World Information
- InfoGlass -

T. YOSHIDA, S. MORISHITA, F. TANEMO and R. KATAOKA

{yosh,morisita,tanemo,kataoka}@syrinx.dq.isl.ntt.jp

NTT Information and Communication Systems Laboratories

1-2356, Take, Yokosuka-shi, Kanagawa, 238-03 Japan

This paper proposes the hypermedia system called InfoGlass, which allows users to get information on the real world interactively. It is important that users can be aware of expanse of the real world and retrieve the information about the objects of their interest intuitively. One of the simplest and most efficient ways to present real world to users through computers is to display the scenery using video cameras. InfoGlass aims to allow users to retrieve the information about the objects in the real world by pointing their images on the computer display. This paper describes our approach to construct InfoGlass.

1. はじめに

これまでのデータベースシステムは、現実の世界に含まれる情報の論理的な関係を容易に表現して管理できる器という位置付けにあり、例えばキーワードを与えるだけで欲しい情報を容易にアクセスできることが重要とされてきた。データベースは現実の世界と一対一に対応する情報を含むものの、情報が現実の世界で持っていた空間的な位置という概念は重要視されず、データベースに格納された途端に失われてしまっていた。つまり、欲しい情報を容易に検索できるように格納した従来のデータベースシステムでは、情報が現実の世界で持っていた空間的な位置という概念を喪失した世界しか提供できず、キーワードでしか検索ができないため、欲しい情報がかえって見付けずらくなってしまうという矛盾を含んでいたのである。このことは情報をデータベース中に埋もれさせる原因になっていると考えられる。

マルチメディア時代のデータベースシステムでは、このような矛盾を解消し、ユーザが現実の世界の空間的な広がりを実際に意識しながら、欲しい情報をデータベースから直観的に検索できることが重要である。

一方、実時間の映像や音が人々に与える影響は大きい。アメリカでは、警察と犯人のカーチェイスが始まると放送中の番組を中断してまでその模様を放送するテレビ局がある。今起きている事実についての映像や音が人々の興味を惹き付けているのである。従ってこれからのデータベースでは、予め蓄積されたデータに限らず時々刻々と入力されるデータについても同様に扱えることも重要である。

現実の世界の空間的な広がりを実際に意識しながらリアルタイムにユーザへ提供する単純かつ有効な方法は、ビデオカメラで現実の世界を撮影し、その映像を提供することである。ユーザがこのような形で提供される映像をいわばインデックスとして利用しながら、映像中の被写体をハイパーメディア的に指示してその関連情報をデータベースから取り出せれば、前述の要求を満たしたマルチメディア時代のデータベースシステムが構築できる。

本稿では、このようなカメラに写る被写体に関する情報をリアルタイムに検索・表示するシステム「InfoGlass」の実現に向けての我々のアプローチを示す。

2. 実世界情報提供システムInfoGlass

2.1 システムの概要

InfoGlassのコンセプトを図1に示す。InfoGlassは、実世界とそれに関連する情報を含むデータベースを、ビデオカメラから実時間に入力される映像（実時間映像）を介してリンクするシステムである。ユーザは、カメラを操作して空間を見渡し、提供される実時間映像中の被写体を指示することにより、関連する被写体情報を検索することができる。

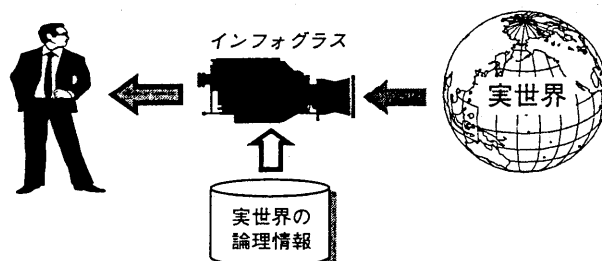


図1 InfoGlassのコンセプト

InfoGlassの主要な機能には以下のものがある。

(1) カメラ操作機能

ユーザは実時間映像を写し出すカメラを操作することが可能である。また、予め用意されたシナリオに従って、システムが操作することも可能である。

(2) 被写体ラベリング機能

被写体に対するユーザの興味を引き出す（ナビゲーション）ために、画面中の被写体についての名称を被写体にオーバーラップさせて表示する。

(3) 被写体選択機能

映像中の被写体をハイパーメディアシステムのアンカとする機能である。被写体は、ある程度長い時間で一定の位置を保持する静的な被写体（静オブジェクト）と、常にその位置や形状を変える動的な被写体（動オブジェクト）より構成されるが、これら両方のオブジェクトが選択可能である。

(4) 情報提供機能

ユーザの映像中の被写体への指示を検出し、予め定義されているシナリオに従って、データベース中の被写体に関する情報をユーザに提供する。

(5) シナリオ定義機能

情報提供機能での動作を定義可能にする機能である。この機能により、映像を変えずに音を変えたり、データベース中のマルチメディアコンテンツを呼び出したり、ハイパーメディアシステムとして様々な振る舞いが定義可能になる。

2. 2 技術課題

実時間映像中の被写体と、それに関連するデータベース中の情報（被写体情報）の対応付けを的確に行なうためには、以下の3つの動的な要素をどのようにシステムで管理するかが課題となる。

(1) カメラの操作

InfoGlassでは、ユーザが興味のある被写体へカメラを向け指示するだけで、その被写体情報を直観的に検索できるシステムを目指すため、ビデオカメラをユーザの意図で自由に動かせなければならない。しかし、カメラの向きや表示倍率を変化させたり、カメラの位置を動かしたりすれば、たとえ被写体の実世界で静止していてもその映像中での表示位置は変化する。InfoGlass実現のためには、ビデオカメラの動きを考慮して、実時間映像中の被写体とその被写体情報を対応付ける技術が必要である。また、InfoGlassでは、複数のカメラを用いてユーザに映像を提供する。このためカメラの位置変動に関しても考慮する必要がある。

(2) 被写体の動き

ビデオカメラの向き、表示倍率、位置のすべてが固定されていても、被写体が動いていればその映像中での表示位置は変化する。静止している被写体に限らず、あらゆる被写体の情報をもInfoGlassを通してアクセスできるようにするためには、上述したカメラの動きに加えて被写体自体の動きも考慮して、実時間映像中の被写体とその被写体情報を対応付ける技術が必要となる。

(3) ユーザのアクション

ある被写体に関してユーザに提示すべき被写体情報は常に一定とは限らない。例えば、ある人が遠くから自動車指して「あれ何?」と聞いたときは、その車の名前を知りたいと考えるのが自然であるが、同じ自動車を近くから指して「これ何?」と聞いたとき、指した先にGPS(Global Positioning System)用のアンテナがあったなら、その人はそのアンテナについて知りたいと考えるのが自然である。このように同じ被写体でも、その見え方の粒度に応じてユーザに提示すべき情報は変化する。InfoGlassでは、ビ

デオカメラのズーム操作が被写体の見え方の粒度を変えるユーザのアクションとなる。このアクションに応じて適切な被写体情報へナビゲーションする技術が必要となる。

2. 3 アプリケーションイメージ

我々の研究室をInfoGlassを通してアクセスするアプリケーション「ビデオハイパールーム」のイメージを図2に示す。我々の研究室には数多くの研究者が在籍し、彼らが使う机や書架、ワークステーション、プリンタ等、数多くのものが配置されている。そして、これらの被写体はそれぞれに個別の論理情報を持っている。例えば、各机はそこを自席とする人の名前を論理情報として持つし、ワークステーションには論理的なホスト名が与えられているのが一般的である。ビデオハイパールームは、研究室という実世界に付帯する論理情報をカメラの目を通してユーザがアクセスできるアプリケーションである。ビデオハイパールームでは、被写体である机やプリンタなどの静物体や、部屋の中を歩き回る人物の位置情報が管理されている。従って、書架にあるマニュアルをクリックすることにより電子図書館の形でマニュアルを読むことができるし、ワークステーションをクリックすることによりそのワークステーションの管理者やOSの種別などの情報を取得することができる。同様に人物をクリックすることにより、その人の電話番号やE-Mailなどの情報を検索することも可能である。

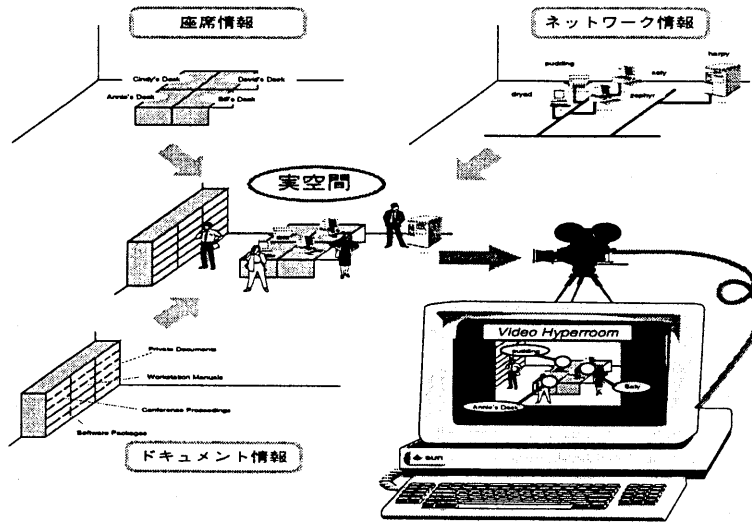


図2 ビデオハイパールーム

2. 4 関連研究

InfoGlassの機能の一つである被写体選択機能を実現するためには、映像中の被写体をアンカとすることが必要である。特に常に位置を変える動オブジェクトについては管理が難しい。ハイパーメディアシステム「雅」[1]では、MOL(Moving Object Locus)データモデルにより、画像中の被写体(動オブジェクト)の位置管理を行っており、被写体をアンカとするハイパーメディアシステムを実現している。また、我々もビデオリアリティ[2]において、動オブジェクトを時空間管理する一手法を示した。ただし、これらのシステムは蓄積映像を前提としており、動オブジェクトの位置情報の抽出もリアルタイムではなくオフラインで手動または半自動で行なわれている。従って、実時間を対象とするInfoGlassへの適用は難しい。

GOLS(Graphical Object Language System)[3]は、蓄積映像だけでなくテレビなどのライブ映像をもハイパーメディアシステムに取り込んだシステムである。GOLSでは、ライブ映像を従来の蓄積型データベースに対してストリーム型データベースと呼び、これらのストリーム(チャンネル)にリンクを張ることでよりハイパーメディアシステムを実現している。GOLSで提唱するデータモデルは、映像データベースで多く用いられるフレームの集合や物理シーン、論理シーンなどに加え、登場物体や任意の区画のセグメントを含んだモデルである。しかしながら、その対象はストリームとして入力される画像であり、映像の枠を越えた物体間の空間的管理はされていない。つまり、GOLSにおける実世界は映像空間がすべてであり、映像枠の外に存在するオブジェクトについてはリンクを張ることは不可能なのである。

一方、IMPACT[4]は、映像の実時間ブラウジングを目的に映像を複数の論理カットに分割し、カットされた複数の映像中の同一オブジェクト間にリンクを張るシステムである。IMPACTでは、例えば映像中の人物などのオブジェクトを自動抽出するが、抽出された人物間のリンクは半自動である。すなわち、映像中の一場面において対象人物の頭・服などの色情報をシステムに登録する。システムは抽出した人物のうち、登録された色のコンビネーションを持つ人物を同一人物とみなしてリンクを構築する。従って、色情報の少ない単色系の人物では誤認識が多く、また同色系の複数の人物の判別が難しい。IMPACTの方式では、たとえ映像中の人物が同一だと判別できたとしても、その人物が誰なのかを認識することはできない。このため、色情報の登録の際にその人物の関連情報とのリンクを構築することが必要であり、これによりカット間で自動的にリンクを構築することが可能になる。しかしながら、色情報の登録には映像が静止している状態が必要であり、実世界の空間に適用するには、オブジェクトを一時的に静止させる等の制約を課さなければならないため問題が大きい。

MPI-Video[5]は、複数のカメラ映像から最適なカメラ映像をユーザに提供したり、オブジェクトの情報をユーザがインタラクティブに検索可能なシステムである。空間の3次元モデルをシステム内に保持しており、空間情報を用いて複数のカメラに写っている動オブジェクトが同一かどうかの判別をリアルタイムに行う。また、映像中の動オブジェクトをリアルタイムに追跡する機能を持つ。しかしながら、現在のところ動オブジェクトが何であるかの認識機能は備えておらず、動オブジェクトとその関連情報との間にリンクを構築するためには、IMPACT同様、キーフレームにおいて登録が必要である。空間の3次元モデルをシステム内に保持しているので、一度マニュアルでリンクを構築してしまえば、その動オブジェクトが全カメラの映像から外に出してしまうまではリンクを保持できる。従って、バレーコートなどの限られた範囲で、複数台のカメラがその空間をすべて撮影するような状況では有効である。しかしながら、この場合にも人手を介するので、やはり登録の際にはオブジェクトを静止させなければならない問題が残る。

3. 技術課題へのアプローチ

3. 1 InfoGlassのアーキテクチャ

InfoGlass実現のためには、被写体である静オブジェクトと動オブジェクトを管理する必要がある。被写体の位置を管理するための空間範囲を観察空間、ユーザが見ることのできる空間範囲を視野空間と呼ぶ。観察空間には視野の広いカメラを用い、視野空間用にはユーザが操作可能なカメラを用いる。観察空間用のカメラを別に用いることで、ユーザの視野空間操作に左右されず広い観察空間から情報を収集することが可能である。また、複数の視野空間カメラに対応することが可能になる。

図3にシステムアーキテクチャを示す。視野空間におけるカメラ制御部は、ユーザインタフェース部からの命令に従いカメラの操作を行う。ユーザインタフェース部は、ユーザに実空間や被写体の情報を提供するインタフェース部分であり、映像中の被写体のラベリング機能や被写体選択部を含んでいる。

一方、静オブジェクト管理部は静オブジェクトの位置情報と、静オブジェクトの関連情報をデータベースとして管理する。動オブジェクト管理部は、動オブジェクトについての管理を行う。動オブジェクト検出部は、観察空間中のオブジェクトの位置をリアルタイムに検出する。リンク構築部は、動オブジェクト管理部で検出されたオブジェクトと、データベース内に予め蓄積してある動オブジェクトの関連情報とをリンクするモジュールである。オブジェクトとその関連情報のリンキングにはオブジェクトの認識が不可欠であるが、現在の認識技術では完全に自動で行なうことはできない。これに関する我々のアプローチを3.3で述べる。

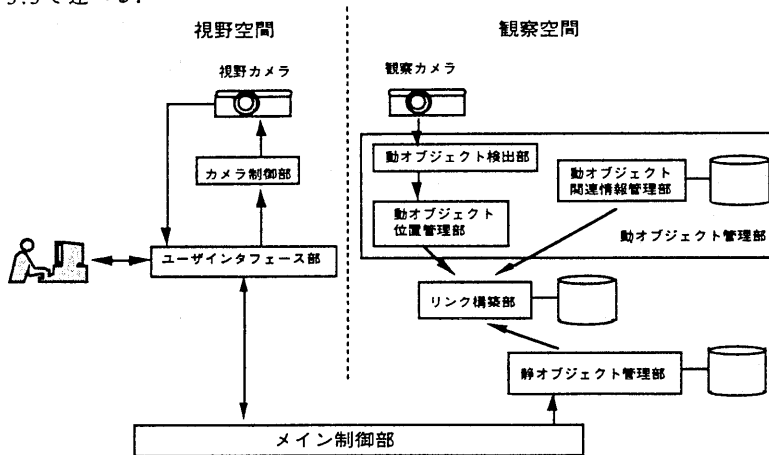


図3 InfoGlassのアーキテクチャ

3.2 カメラ操作への対応

被写体の表示位置がビデオカメラのパン、チルト、ズームといった操作によって移動しても、被写体と被写体のラベルを映像中で常にオーバーラップさせて表示することが必要である。このカメラ操作への対応について静オブジェクトを対象に検討を行った。画面イメージを図4に示す。

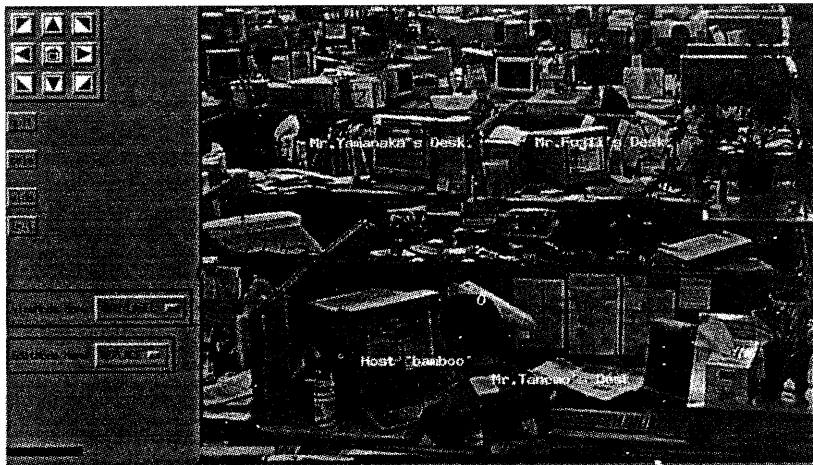


図4 静オブジェクトの情報検索

静オブジェクトの位置については、予めカメラ映像上で登録を行う。すなわち、直行座標系(x,y)を、

カメラ位置を原点とする球面座標系(頂角, 方位角)に変換し, またオブジェクトの大きさを, すべてのオブジェクトの表示サイズが一定の大きさになるようなズーム率に補正し, (頂角, 方位角, ズーム率)の3次元で管理する。

また, 映像中の被写体のラベルを映像中ですべてオーバーラップさせてしまうと, 被写体数が多い場合には情報過多になり, 有効なナビゲーションは実現できない。このため, (i)観察空間中の視野空間部分に存在する静オブジェクトに対する検索に加え, (ii)カメラのズーム率に応じたラベリング機能のための検索が可能となっている。

3. 3 被写体の動きへの対応

動オブジェクトは常に空間中で位置を変えるため, 静オブジェクトのように予め位置を定義しておくわけにはいかない。つまり, 動オブジェクトをリアルタイムに認識し, 対応するデータベース中の情報へリンクを構築する必要がある。動オブジェクトの認識には, 画像認識技術を利用するが, 現在の認識技術では, オブジェクトの位置をリアルタイムに検出できるものの, そのオブジェクトが何であるかを認識することはできない。IngoGlassでは, 静オブジェクトと動オブジェクトの位置的特性を用いて動オブジェクトの認識を行い, 動オブジェクトとその関連情報の間にリンクを構築する。図5はビデオハイバールームにおけるリンク方式の概念図である。以下にリンク構築手順を示す。

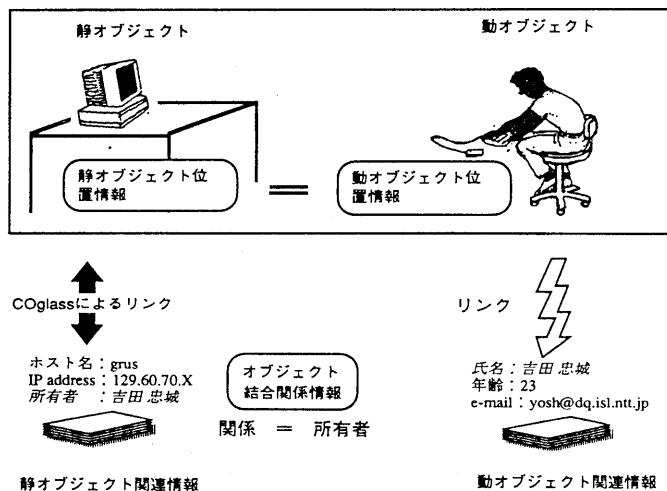


図5 リンク方式

(1) 静オブジェクト位置管理

コンピュータや席などの静オブジェクトの観察空間における位置情報, ホスト名やネットワークのIPアドレスなどの静オブジェクトに関連する情報は, 図3の静オブジェクト管理部により管理される。つまり, 静オブジェクトの位置が分かればホスト名などの情報を検索することが可能である。

(2) 動オブジェクトから静オブジェクトの関連情報へのリンク

「動オブジェクトの位置が, ある一定時間Tだけ静オブジェクトに隣接する位置に留まった場合, その静オブジェクトの所有者は検出された動オブジェクトである。」といった規則を設けておくことにより, 動オブジェクト検出部により検出された動オブジェクトが静オブジェクトとリンクされ, さらに(1)により, 動オブジェクトは静オブジェクトを介して静オブジェクトの関連情報へリンクされる。

(3) 動オブジェクトの認識および動オブジェクト関連情報へのリンク

静オブジェクトの関連情報には、ホスト名などの情報のほか、その静オブジェクトの所有者など動オブジェクトと静オブジェクトとの論理的関係を示す情報(オブジェクト結合関係情報)も含んでいる。このオブジェクト結合関係情報から動オブジェクトの人物特定がなされる。そして動オブジェクトはその関連情報へリンクされる。

このようにInfoGlassにおける手法は、(i)特定の静オブジェクトと結合関係を持つ動オブジェクトを対象にし、(ii)静オブジェクトの位置情報と静オブジェクトの関連情報間に予めリンクを構築しておく、(iii)静オブジェクトと動オブジェクトの位置関係と、静オブジェクト関連情報中のオブジェクト結合関係情報を用いて動オブジェクトを自動認識する方式である。

4. おわりに

本稿では、実世界に存在するオブジェクトの関連情報を直観的に検索できるInfoGlassシステムについて、我々の提案を述べた。InfoGlassは、“直観的な検索とは、ビデオカメラから実時間に入力される映像から映像中の被写体に関連する情報を直接検索可能なことである”という観点から、映像中のオブジェクトとデータベース中のオブジェクトに関連する情報を、リアルタイムにハイパーリンクするシステムである。InfoGlass実現のためには、(i)カメラの操作、(ii)被写体の動き、(iii)ユーザのアクションの3つの動的な要素を考慮してオブジェクト管理を行なうことが課題である。

また、カメラの操作について静オブジェクトを対象に検討した結果と、被写体の動き(動オブジェクト)について検討した結果を示した。動オブジェクトの認識には、静オブジェクトと動オブジェクトの論理関係を示す「オブジェクト結合関係情報」がキーになる。本稿では「オブジェクト結合関係=所有者」という論理関係を用いたが、野球であればベースと選手、水泳であればコースと選手など実世界には多く存在するため有効な手法と考える。

今後、カメラの操作に関しては、(i)映像の歪を考慮した被写体の位置決定手法、(ii)カメラの位置変動を許容した位置管理手法が検討課題である。また、被写体の動きに関しては、実際にインプリメントを行い、どのくらいの信頼度が得られるかの評価が課題である。さらに、ユーザのアクションについても検討を進める。

参考文献

- [1] 高野,的場,原:"ビデオデータ中に現れる物体をノードとするハイパーメディア構成方式",7th symposium on Human Interface,pp.301-306,1991.10.
- [2] 片岡,紺谷,三浦,佐藤,井上:"実写映像を用いたマルチメディア情報システム「ビデオリアリティ」の構築",信学技報,IE94-123,DE94-78,1995.02.
- [3] 佐藤,坂内:"ライブ情報を扱うハイパーメディアシステムGOLSにおける認識システムの構成",信学技報IE94-12,pp.15-20,1994.
- [4] H.Ueda,et.al.:"Automatic Structure Visualization for Video Editing",INTERCHI'93,pp.137-141,1993.4.
- [5] P.H.Kelly,et.al.:"An Architecture for Multiple Perspective Interactive Video",VisComp Lab. Technical Report,VCL-95-103,1995.