

ファジイ推論による映像情報処理 (VIP) データベース

上野 義人

創価大学工学部

〒192 八王子市丹木町1-236

動画データベースの検索システムとして、動画をシーン、カット、コマと階層的に分割し、最小コマにおける内容的な代表フレームをヒストグラムによる最頻フレームを抽出して、コマの意味づけをする。この代表フレームをイメージ処理技法により、時空間特徴を抽出して、このコマのインデックスとする。質問処理は、インデックスの類似度検索として、ファジイ推論を利用して、類似度を求める。このような部分的な特徴情報の検索により、動画の内容検索を実行する。この検索結果をプレゼンテーションマネージャにより、動画の編集を行ない、動画をインタラクティブに再生する。この特徴パターンによるファジイ推論の類似検索手法を動画データベース管理手法に適用するシステムについて、考察した。

Visual Information Processing (VIP) Database System by Fuzzy Reasoning

Yoshito Ueno

SOKA University

1-236 Tangi-cho Hachioji-shi Tokyo 192

The visual information database system has a structure of executing the queries by the user interactively which is introducing the knowledge base using fuzzy reasoning. An extracting method of temporal and spacial characteristics of visual information adopts the method of making an index of video frame's characteristics of the representative video frame using image processing methodologies. The queries by the user is utilizing this index and executing the retrieving of the similarity of a video image. This retrieved representative video frame is linking to the actual visual information which is modified and edited by the user and then displayed to the user by the presentation manager or sent to the other client through the network.

1. まえがき

最近マルチメディアが脚光を浴び、ビデオデマンドサービスが実用化段階に入ってきた。このマルチメディアは、映像、音声、テキストなどの複数メディアを有機的に結合したものであり、テレビ、コンピュータ、通信などのメディアが融合したものと考えられるが、まだ、その実体は漠然としている。しかし、マルチメディアの主体は、映像メディアであり、この映像情報を自由に選択し、操作できる映像情報処理(VIP)システムを提案してきた。(1)

この映像情報処理(VIP)システムの目標は、人間の創造性を高め、人間相互のコミュニケーションにおける相互理解を深め、さらに、プレゼンテーション・ツールとして、高い効果を発揮できるように、個人が映像情報を自由に操作できる環境を構築することである。このため、多数の変化に富んだ映像情報から、個人が選択し、利用する手段として、ラベル付けされたビデオテープよりも使い易く、編集作業が容易な映像情報データベースが必要である。

複雑な内容を持つ映像情報データベースの検索手法には、必要とする映像のフィルタリング操作とブラウジング操作の繰り返しであり、映像データのインデックス付けにより、ユーザの質問処理が実行される。

この映像情報の質問処理として、映像情報オブジェクトから特徴を抽出し、ユーザの欲する映像情報とマッチングを採る方法がある。しかし、映像メディアの正確なマッチングは、きわめて困難なため、類似検索手法が広く用いられている。

この類似検索の一手法として、今回、ファジイ推論を用いた映像情報処理(VIP)データベースシステムについて検討したので、以下にその内容について述べる。(2)

2. オブジェクト指向による映像情報のモデル化

映像情報のモデル化について、オブジェクト指向の考え方を取り入れて、多くの研究成果が得られているが、空間的映像情報のモデル化は、画像情報のモデル化技法が利用でき、(3) 時間的映像情報のモデル化については、映像情報を連続オブジェクトとして取り扱い、映像情報そのものを複合オブジェクトとするモデル化手法がある。(4)

空間的映像情報の意味内容をオブジェクト指向分析ツールを用いて記述し、時間的映像情報のクラス、ノード間に動的リンクを設定する。この動的リンクにより、大量の映像情報を記憶している媒体にメッセージ通信を行ない、参照リンクで、メディアオブジェクトから映像情報を参照して、ユーザに表示する。(5)

今後、映像情報のファイルシステムとして、デジタル映像の記録が主流となり、コンピュータとのインタフェースも簡単になる。

また、映像情報オブジェクトを管理するオブジェクト指向データベース管理システムも、複雑かつ大規模なデータ構造の取り扱いも容易になり、映像情報などの可変長データのサポートも可能となる。

3. 映像情報表現の階層化

映像情報は、時間的に連続した静止画として表現できる。この静止画の属性は、画像が持つ意味を表現する必要があるが、複雑な画像の場合、一義的に表現することが困難である。

また、画像の物理的属性として、テクスチャ、色、明度の強度、形状などがある。一枚の静止画から、色ヒストグラムや強度ヒストグラムによって、画像を識別することも可能である。また、形状属性として、

点、線、領域特性などがある。さらに、テクスチャ属性として、形状表面の粗さ、エッジ特性などがある。

ユーザが映像情報を検索するステップとして、映像全体のストーリーを表す属性を最初に検索する。

次に、その映像の中の特別のシーンを探す。また、その次に、そのシーン中のどのカットを見たいかを探す。この検索順序が自由に入れ替えることができる構造を持つ必要がある。

したがって、映像情報の属性として、物理的な特性を探すケースは少なく、映像が持っている意味によって探す場合が多く、映像の意味を特徴づける属性を定義する必要がある。

映像情報データベースは、現状の記憶媒体の記憶容量に比較して、映像情報の容量が膨大であり、アプリケーションの分野別にデータベースを構築する必要がある。

この分野の区分法の一つとして、図書分類で用いられる十進分類法がある。この分類法は、階層構造しており、大分類として、総記、哲学・宗教、歴史・地理、社会科学、自然科学、技術、産業、芸術、言語、文学に大別されている。

この大別された一分野は、例えば、芸術は、芸術・美術、彫刻、絵画、版画、写真、工芸、音楽、演劇、スポーツ・体育、諸芸・娯楽等に細分類される。この細分類された項目から、例えば、スポーツ・体育は、体育一般、体操・遊戯、陸上競技、球技、冬季競技、水上競技、戸外レクリエーション、釣り魚・遊猟、相撲・拳闘・競馬、武道等に細分割される。

さらに、各細項目は、例えば、相撲・拳闘・競馬は、相撲、レスリング、ボクシング、動物の闘技（闘牛、闘鶏、闘犬）、競馬、競輪・オートレース・競艇、その他に細分割される。

勿論、細分割した項目は、二カ所以上にまたがるものもある。

このように階層化された映像情報データベースを構築すれば、ユーザがどの分野の映像情報データベースをアクセスすればよいか、手がかりを掴むことができる。

しかし、実際の映像情報は、放送モードの形態が多く、分類法は、ユーザの利用頻度の高い分野から、階層的に分類したデータベースを構築する方が適している。

すなわち、インタラクティブ映画、スポーツ、教育、ドキュメンタリー、旅行、各種案内情報、イベント、ゲーム、ニュースなどに分類する。

一例として、富士山を背景とした旅行シーンの階層化例を図1に示す。

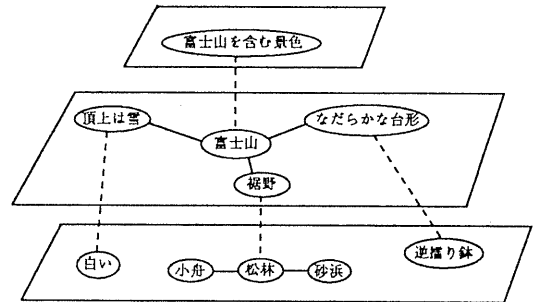


図1 階層化された映像情報セマンティックマップの例

4. 映像情報データベースの構築と検索技術

映像情報の編集の民主化を実現するためには、専門家集団が制作した映像情報を一方的に受け入れる放送システムから、ユーザ自身が自由に映像情報を選択したり、蓄積する能動的な映像情報システムを構築する必要がある。このため、柔軟で、かつ、ユーザが使いやすい映像情報データベースの構築が不可欠である。

時間的、空間的に広がった映像情報データベースを構築するためには、従来のデータベースと異なった方法を適用する必要がある。(6)

すなわち、映像情報をモデル化し、シーン、カットを自動抽出し、属性をつけて、インデックス化する。

このような技術は、広く研究されているが、もう少し視点を変えて、システム全体を総合的に検討する必要がある。

現在の映像情報源であるテレビ放送やVTRなどは、制作者側の効率を重視したもので、ユーザ側の便利さについては、全く考慮されていない。

現在の標準テレビ方式でもデータチャンネルを活用することができるが、その主な目的は、文字放送、静止画放送、FAX放送、ゲームソフト放送等ユーザが受動的に視聴するためのデータ伝送である。

この224-1760 kb容量のデータ伝送容量を有効に活用すれば、映像情報にシーンやカットの属性情報を挿入して、映像情報データベースの情報アクセスに簡単に利用できる。

また、将来、テレビ放送をデジタル化する際、映像情報データベースを構築することを想定して、送り手側に必要な情報付加機能を追加すればよい。例えば、デジタルビデオ符号化列のヘッダ部分にデータベースアクセスのための付加情報を挿入する。

ユーザ自身がビデオ一体型カメラを用いて、映像情報を発信して、双方向型映像情報システムを構築するためには、テレビカメラの機能もインテリジェント化する必要がある。そのためには、撮影開始時に映像情報にラベル付けするための手書き入力または音声入力など、グラフィカルユーザインタフェースの機能を装備する必要がある。また、カメラ操作のズームやパンに相当した付加情報を挿入して、映像情報データベース構築に便利な機能を追加することは、技術的には、あまり困難ではない。

さらに、VTRは、現状でも、タイムコードを挿入することができるエキストラクト

ラック機能を持っている。

以上のような映像情報取得環境が提供されると、ユーザ自身が主体性を持って映像情報データベースの構築が可能になり、映像情報のフィルタリングやインタラクティブ性に富んだ映像情報ネットワークが構築できる。図2に映像情報データベース構築の効率化について、各種の手法を示す。

TV + コンピュータ = 新しい映像端末

映像ソース	送り手側	受け手側
放送テレビ (将来デジタル化)	キーワード付与 (シナリオから抽出) カットの指示 (シナリオから抽出)	アクセスツールの簡略化 映像情報フィルタリングの簡易化 カットの自動抽出…不要 シーンの自動抽出…不要
カメラ (デジタルカメラ)	ズーム、パンの番号付加 (ヘッダに挿入) 撮影開始番号付加 (アイコン利用GUI)	カットの自動抽出機構不要
VTR (デジタルVTR)	キーワード付与 カットの指示 (圧縮符号化フレームのヘッダに挿入)	アクセスツールの簡略化 映像情報のフィルタリングの容易化

図2 映像情報データベース構築の効率化方法

5. オブジェクト指向による映像情報オブジェクトの表現

映像情報データオブジェクトの動的振り舞いは、映像情報処理 (VIP) データベースシステムの応用分野を特定し、映像情報をモデル化して、コマの代表フレームをオブジェクトとする。この映像情報データの意味付けと内容表現を多重クラス階層でデータ表現する。

各オブジェクトは、関係の合成、汎化、抽象化等のプロセスをもち、かつ、サーバスタブをもって、外部からの要求信号を待機し、ネットワーク上の管理機構へ要求信号を送出する。また、各オブジェクトがノード間とのリンクをとるため、特徴抽出された概略画像オブジェクトとコマとのリンク、コマとシーンとのリンク、シーンとストーリーとのリンクなどを設定し、属性値を上方および下方継承する機構をもつ。

さらに、ファジイ推論による質問処理検索結果で選択されたオブジェクトを実際の映像情報ファイルにアクセスするため、参照リンクをもっている。この選択された映像情報オブジェクトを加工し、編集するための時空間操作オペレータを映像情報処理データベースマネージャにもっている。このような映像情報オブジェクトの動的な振る舞いを図3に示す。

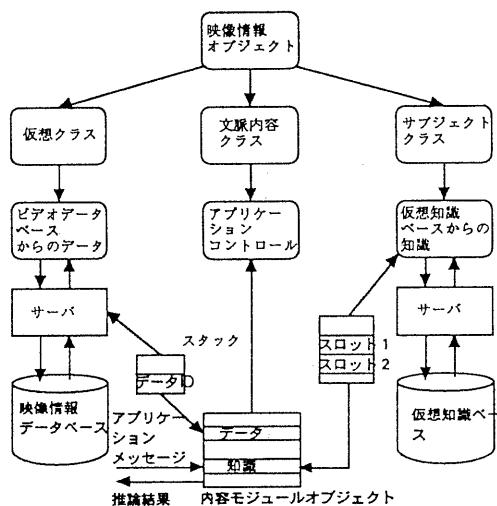


図3 映像情報データオブジェクトの動的振り舞い

この際、重要なことは、映像情報オブジェクトのセマンティックスを正確に表現できるかどうかという問題がある。映像情報オブジェクトの物理的特徴は、画像処理技法により、自動的に特徴抽出することが可能であるが、一フレームの映像中の他の対象物との相対関係（前後、左右等の空間的な位置関係）、複数の対象物間の意味的な動的対応関係（動き、意図などの作用関係）等を正確に表現することが困難である。

したがって、映像情報オブジェクトの表現を曖昧のまま表現し、ファジイ推論により、曖昧さの定量的な不確実差を求めて、近似的な映像情報オブジェクトの意味的な表現を行う。

6. ファジイ推論知識ベースによる質問処理

6.1) ファジイ推論

ファジイ推論による概略画像の類似度検索のブロックを図4に示す。

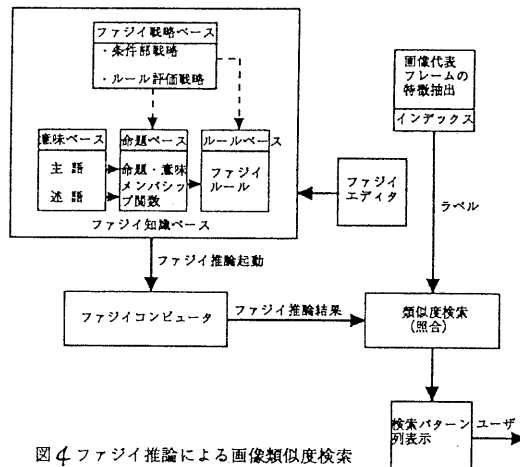


図4 ファジイ推論による画像類似度検索

ファジイ理論を画像の類似検索に適用するためには、人間が認識している概念や意味の曖昧さをファジイ集合で定義し、このファジイ命題を用いて、曖昧さを含む画像の特徴的形状および色彩に対する経験則群をif-thenルールとして記述し、このルールベースに事実をあてはめ、近似的な推論結果を得て、画像の類似度検索を行なうことができる。さらに、ファジイ命題の述語修飾演算子を拡張し、基本的な経験に基づく推論に加えて、状況変化の傾向を加味した推論結果を得る。(7)

ファジイ集合の抽象化にオブジェクト指向の考え方を取り込み、ファジイルールの記述を効率化するため、エキスパートシステムとの融合をはかる。また、ファジイチップやファジイ推論を用いた知識データベースなどが開発されているので、これらを採用して、画像の類似度検索のハードウェア化を行なう。

ファジイ・エディタは、ファジイ推論の知識ベース専用のエディタでユーザは、対話的にファジイ知識ベースを構築する。

すなわち、以下に示す、ファジイルールの作成、メンバシップ関数の定義、ファジイ戦略の設定などの手順にしたがって実行する。

(1) ファジイルールの作成

ユーザは、経験に基づく数種類のルールメニューからルールを選択し、if-then型のファジイルールを作成する。この入力情報に基づき、ファジイ命題、ファジイルール、ファジイ意味などをファジイメモリに格納する。

(2) メンバシップ関数の定義

ユーザは、作成されたファジイ命題に対応したメンバシップ関数の型、定義域の範囲、分割数、座標などを指定し、ファジイメモリに格納する。

(3) ファジイ戦略の設定

条件部戦略は、ファジイルール条件部に複数のファジイ命題が記述されているため、ファジイ命題の適合度の中から最適なルール適合度に基づき、対応する結論部ファジイ命題のメンバシップ関数をどのような方法で変形するかを示す。

ファジイ推論による映像情報の類似検索の例を図5に示す。この類似検索手順は、以下の順序で実行する。

(1) ファジイフレームモデルとして、代表フレームを大きさ、概略形状、色彩情報、位置、接続、包含関係、相対関係などを木構造で記述する。

(2) 大きさ、形、色、位置などのタスクごとに、ファジイ推論を実行する。

(3) ファジイプロダクションルールの記述として、if-thenルールベース、ファジイ知識ベースにより、多数のユーザの知識を獲得したもので記述する。

(4) 類似検索部では、ユーザの入力印象語に対し、知識ベースを用いたファジイ推論を適用する。

(5) 代表フレーム類似度検索結果に相当

した候補代表フレームを表示するファジイマッチングとして、min-max重心法によるファジイ推論を実行する。

(6) ユーザに表示するために、メソッドの操作により、実際の映像情報を表示し、プレゼンテーションマネージャによる加工、編集を実施する。

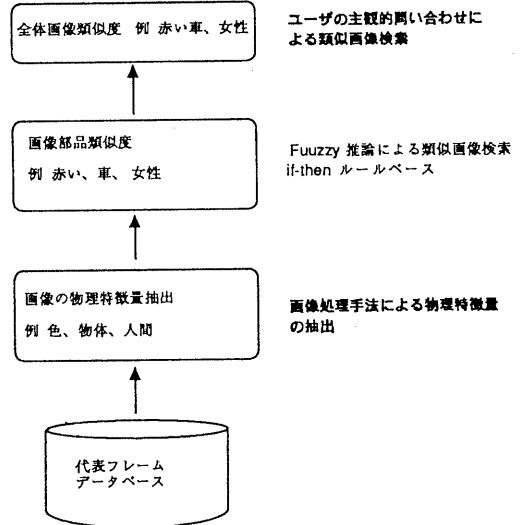


図5 Fuzzy推論による類似画像検索

このようなファジイ推論のみでは、類似度検出速度が遅いため、ファジイ・ニューロを結合した方法により、推論実行の高速化が図れる。

6.2) ファジイ、ニューロの融合

ニューラルネットとファジイとを論理演算によるハイブリッド結合により、ファジイ、ニューロの融合が可能である。

ニューラルネットワークの学習機能を導入したファジイルールの設定、ファジイ集合のメンバシップ値を入出力するニューラルネットワーク、一部をニューラルネットワーク化したファジイルール、一部をファジイ化したニューラルネットワークなどがある。

図6にニューラルネット駆動型ファジイ推論のブロック図を示す。これは、前件部

のファジイ集合と後件部の入出力関係をバックプロパゲーションモデルによって、補正する方法である。

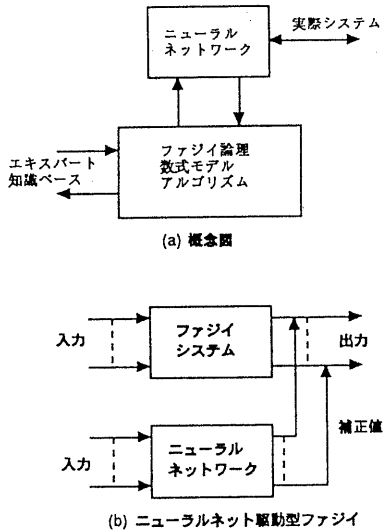


図6 ニューラルネット駆動型ファジイ推論

明確なプロダクションルールを記述できる場合には、if-then ルールを構築できるが、映像情報のように、記述できない対象の場合、ニューラルネット駆動型ファジイモデルにより、ルールのメンバシップ関数の形状を決定することができる。

学習のための教師データと、ラフなファジイルール、メンバシップ関数を用いて、ニューロ・ファジイも構築できる。

このため、ファジイルールは、ニューラルネットワークに埋め込まれ、ファジイ推論は、ルールを一つずつチェックする必要がなく、推論の実行が高速化される。

このようなファジイ・ニューロの融合による推論実行により、映像情報データベースの質問処理を高速化できる。

現在、SUN WS上で、簡単なシミュレーションの構築を実施している。

7. 映像情報処理 (VIP) データベースシステムの構成

映像情報処理 (VIP) データベースシス

テムの構成を図7に示す。

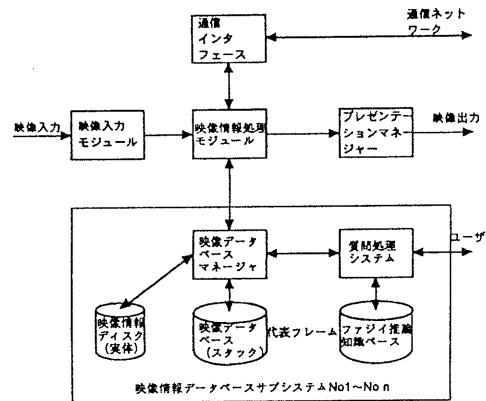


図7 映像情報処理 (VIP) データベースシステムブロック図

大量の映像データを効率よく圧縮可能な国際標準規格MPEGを採用して、デジタルビデオデータに変換する映像入力モジュール、圧縮された映像データを動画像処理し、画像のスムージング、強調、セグメンテーション、特徴抽出、カラーマップ、ヒストグラム等の処理を行なって、ズーム、回転、シフト、パンニング、スーパインポーズ、オーバーレイ等の操作を併せて行なう。

動き抽出として、動きのペースを抽出する。このような映像情報処理モジュールの機能として、以下に示すような、各種オペレータを追加する。

(1) 時間的制約を表現するオペレータ

equal : オブジェクトの同定

before : オブジェクトの前後関係

meets : オブジェクトの合成

sim(simultaneously) : オブジェクトの瞬時操作

starts : オブジェクトの開始

finishes : オブジェクトの終了

at : オブジェクトの一時停止

(2) 空間的配置オペレータ

left a : オブジェクトaを画面の左側に表示

right a : オブジェクトaを画面の右側に表示

bottom a : オブジェクトaをウィンドウの下

側に表示

up a: オブジェクトaをウインドウの上側に表示

a shownlb b: オブジェクトaをウインドウの左上角または、右下角に表示

arrange(a,b,c): オブジェクトaのウインドウとオブジェクトbのウインドウとの重なり部分を透視度cで表現し、c=0のとき、重なり部分が見えず、c=1のとき、重なり部分が見える。

映像情報データベースサブシステムの機能として、映像データの時間的前後関係の相関関係を決定する機能をもっている。また、映像情報を蓄積するアーキテクチャとして、格納場所のアロケーションやバージョン制御を行なう映像データベースマネージャをもっている。

映像情報のアクセス制御として、ファジィ・ニューロ融合による推論を用いた質問処理システムを採用した。このようにして獲得された映像情報のコマに対応した全体の実際の映像は、別個に設けられた映像情報ディスクファイルから、クライアント・サーバ通信技法によって、プレゼンテーションマネージャに供給される。このプレゼンテーションマネージャには、ユーザの要求に応じた編集機能をもっている。また、外部ユーザや映像データベースと通信するインタフェースをもっている。

8. おわりに

ファジィ推論による映像情報処理(VIP)データベースは、ただ一つの正確な推論結果が得られないため、数種類の概略画像を例示する必要がある。このため、その他の画像検索に比較して、余り大きな検索速度の向上は期待できない。

将来、自然言語処理技術の進展により、自然言語レベルによる質問処理検索が可能

となれば、ユーザインタフェースも飛躍的に向上する。

文献:

1. 上野: "映像情報処理(VIP)パーソナル映像情報データベースシステム" 1991年信学春季 全大、D-436
2. 上野: "ファジィ推論を用いた映像情報処理(VIP)データベース" 第48回情処全大、3M-07,1994
3. 加藤、栗田: "画像の内容検索" 情報処理、Vol.33, No5, pp466~477, 1992
4. 大本、田中: "マルチメディアデータベースにおける連続オブジェクトのモデリング" WOOC'92論文集、pp149~164, 1992
5. 藤川他: "分散型ハイパメディアシステムHarmonyにおける連続オブジェクトのモデリング" 信学誌 論文誌(D-I) pp473~483, 1993年7月
6. 坂内: "マルチメディアデータベースの展望" 信学技報IE94-10, pp1-8, 1994
7. 染谷、安信: "エキスパートシステム構築ツールES/KERNEL/Wのファジィ推論" 日本ファジィ学界誌、Vol.2, No.2, pp125-132, 1990