

動画像データベースにおける状況概念の検索

細田 祐一† 吉高 淳夫‡ 平川 正人‡ 市川 忠男‡

†広島大学大学院工学研究科

‡広島大学工学部

〒739 東広島市鏡山 1-4-1

近年、動画像データを容易に扱うことができる環境が整いつつあり、それに伴い動画像データベースへの要求が高まっている。動画像には複数の物体が存在し、それぞれが時間の経過とともに動いているが、それらは互いに関係し合っていることがある。本研究では、チーム単位で対戦が行われる球技の試合映像に含まれる物体間の時間的、空間的な相関関係に基づく意味概念を状況概念として定義し、それによって所望のシーンを検索する手法を提案する。本手法では、球技において重要なボールの動きに注目し、ボールと複数のプレーヤーとの間に存在する時間的、空間的な相関関係を抽象的に表現する。それにより、効率的な検索処理が可能となる。

Video Database Retrieval by Spatio-Temporal Correlations among Objects

Yuichi Hosoda† Atsuo Yoshitaka‡ Masahito Hirakawa‡ Tadao Ichikawa‡

†Graduate School of Hiroshima University

‡Faculty of Engineering, Hiroshima University

1-4-1 Kagamiyama, Higashi-Hiroshima 739, Japan

In recent years, because of the progress of hardware capacity, it has become easy to handle video data with computers. With response of the progress, more advanced video database techniques for storing, managing, or retrieving video data, have been required. As video data implies objects and their motion, they are often related with each other. In this paper, we propose a retrieval method by specifying spatio-temporal correlations among objects observed in the video data of a ball game played with teams. Our method concentrates on abstract representation of the spatio-temporal correlations among a ball and some players. It provides an efficient retrieval.

1 はじめに

近年、コンピュータハードウェアの技術的進歩に伴い、画像、音声、動画といったいわゆるマルチメディアデータを容易に扱うことができる環境が整いつつあり、マルチメディアデータベースへの要求が高まっている。マルチメディアデータの中でも動画データは、時間的、空間的な関係を含んでおり、その多量な情報量のためにソフトウェア技術、特にデータベース技術に関する問題が山積している。最近では、動画データに関する研究が盛んに行われており、検索手法についての研究報告も増えてきている。

従来の動画検索手法には、動画に現れる物体に関する特徴、あるいは画像の特徴など、検索対象となる要素間に依存関係がない独立した特徴を指定することにより所望のシーンを検索するものが多い[1][2][3][4]。動画ではさまざまな物体が現れ、それらが互いに関係し合っている場合が多いが、[1]-[4]には動画の特徴である複数物体間の時間的、空間的な相関関係を含む問い合わせを評価する枠組みが用意されていない。動画の内容に基づく検索を行う場合、特にこの時間的、空間的な相関関係の処理が重要である。

一方、そのような相関関係に注目した内容検索手法として[5][6]が報告されている。[5]は、動画に現れる複数物体の状態変化をプリミティブな現象の時間的な関係として表現し、検索を行っている。[6]は、空間的な関係と時間的な関係を n 項から成る関係子で表現することにより、時間的、空間的な要素を持つデータのモデル化を行っている。しかし、これらの手法では、関係表現の抽象度が低いため、ユーザは時間的、空間的な相関関係に対してその構成内容をモデルで用意された関係表現のレベルで詳細に記述しなければならない。それに対して、人間は物事に対して大域的な、すなわち抽象度の高いレベルで解釈、あるいは判断をする。したがって、時間的、空間的な相関関係に対しても、抽象度の高い表現記述による問い合わせ方法が有効であると考えられる。

[7]では、抽象的な表現による問い合わせに対してセマンティックネットワークに基づくデータモデルを用いた動画検索手法が提案されている。このセマンティックネットワークは、抽象的な事象の意味内容から具体的なレベルの制約条件を推論するための機構であり、これにより動画データから抽出さ

れる具体情報とのマッチングが行われている。しかし、さまざまな事象についてのセマンティックネットワークの構築は容易ではない。

本研究では、動画に含まれる物体間の時間的、空間的な相関関係に対して抽象的な関係表現による指定を可能にし、その特徴に基づく所望のシーンを検索する手法を提案する。また、動画の種類を限定することにより、その動画の特性を生かした問い合わせ処理の枠組みも提案する。

2 状況概念

チーム単位で対戦が行われるスポーツの試合を撮影した映像には、複数のプレーヤーが存在し、時間的、空間的な相関関係に基づくプレーが多く見られる。このようなスポーツの映像では、一般的に得点にかかわるようなシーンが検索対象になりがちであるが、チーム分析、あるいは試合内容の分析という視点で考えると、複数のプレーヤーが関係し、攻撃あるいは守備を効果的にするためのプレーも検索対象に含まれると思われる。

本研究では、アプリケーション動画をチーム単位で対戦が行われる球技の試合映像に設定し、この動画に現れる物体間の時間的、空間的な相関関係に基づく意味概念を状況概念として定義する。

[8]では、サッカーの試合映像の解析が行われている。フィールドのラインの認識によりそれぞれのカメラショットがフィールド内のどの場所を撮影したものを判別し、またショット内の物体とその動きも検出することにより、シュートやコーナーキックといった場所への依存度が大きい内容解析を行なっている。しかし、この手法では、物体間の時間的、空間的な相関関係について考慮されていないため、複数のプレーヤーが関係するプレー、すなわち状況概念を含む問い合わせに対して検索を行うことができない。それに対して本研究では、チーム単位で対戦が行われる球技の試合映像としてサッカーの試合映像を実験的に扱い、状況概念に基づく所望のシーンを検索する枠組みを提案する。なお、サッカーにおける状況概念は、ボールと複数のプレーヤーの動きで構成されるプレーであり、例えば壁パス、スルーパス、オーバーラップ、スクリーンプレーなどが挙げられる [9]。

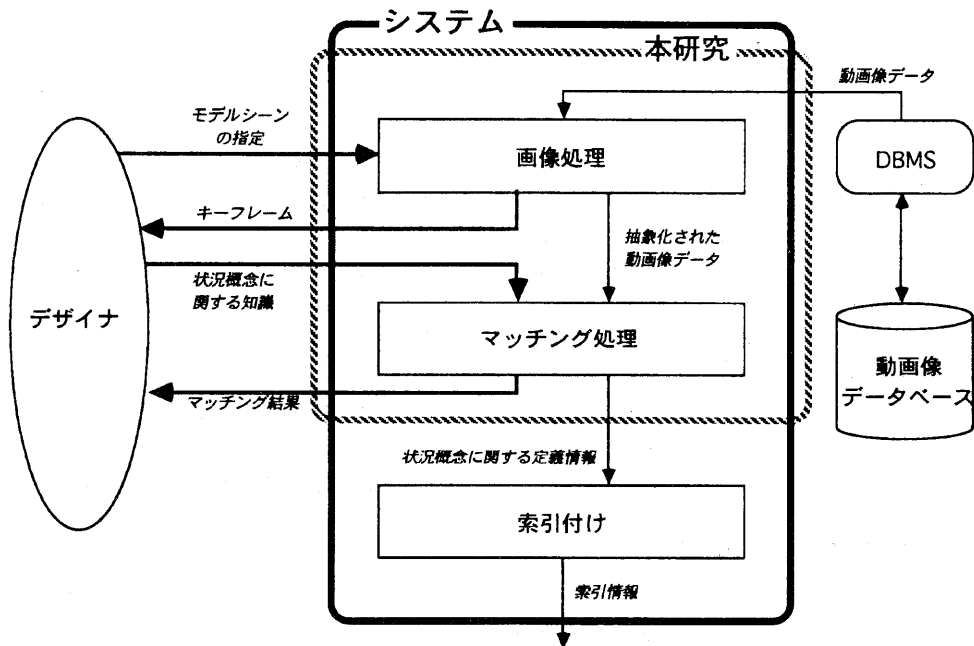


図1 システム構成

3 基本方針

本研究では、デザイナーが状況概念に関する知識の定義付けをし、それに基づきシステムが動画データに対して索引付けを行う。ユーザは索引付けされた動画データに対して検索を行うものと想定する(図1参照)。なお、本研究では状況概念に関する知識の定義付けが正しく行えたかどうかを確認するためにデザイナーが行う検索に注目し、以後それを単に検索と呼ぶことにする。

膨大な情報量を持つ動画を扱う際、検索を行う度にかかる処理時間の長さを無視することはできない。また、時間的、空間的な相関関係を含む状況概念を構成するには、依存関係のない特徴の指定に比べて多くの情報を必要とするため、そのような状況概念に対して知識の定義付けを行うのはデザイナーにとって困難である。そこで本研究では、以下のアプローチによりそのような問題の解消を図る。

- 検索処理を効率的に行うために動画データの抽象化を行う。
- 知識の定義付けを行う際のデザイナーにかかる負担を軽減させるためにモデルシーンを利用した定義方法を探る。

動画データは、画像処理を施すことにより、キーフレーム単位のボールとプレイヤーの位置情報に抽象化される。ここでキーフレームとは、ボールがあるプレイヤーに接触しているすべてのフレームとする。動画データをあらかじめキーフレーム単位のボールとプレイヤーの位置情報に変換しておけば、検索のたびに画像処理を行う必要がなく、検索処理にかかる時間を短縮することができる。

デザイナーは状況概念に関する知識の定義付けを行うためにモデルシーンを利用する。ここでモデルシーンとは、画像中に所望の状況概念が確認される典型的なシーンである。システムは、デザイナーが指定したモデルシーンに対して画像処理を施し、状況概念を構成するために必要なキーフレームやそれぞれの物体に関する情報を取得する。このため、デザイナーは各キーフレームに対して状況概念を構成するプレイヤー間の空間的な相関関係を記述するだけでなく、複雑な状況概念の定義付けを容易に行うことができる。

以上のような方針に基づき、本手法による問い合わせ処理の流れを以下に示す。

- (1) デザイナーは所望の状況概念が含まれるモデルシーンをシステムに指定する。

- (2) システムはモデルシーンから状況概念を構成するために必要な情報を抽出し、モデルシーンにおけるキーフレームをデザイナーに提示する。
- (3) デザイナーは提示された各キーフレームに対して状況概念に関する知識を記述する。
- (4) システムはモデルシーンから得られた情報とデザイナーによって定義付けられた知識に基づいて状況概念を構成し、キーフレーム単位のデータに変換された動画像データとのマッチングを行う。
- (5) システムはマッチング処理の結果をデザイナーに出力する。

4 キーフレームの検出

キーフレームの検出は、動画像データとモデルシーンの両方に対して行う。画像処理により抽出されたそれぞれのフレームにおける各プレイヤーとボールの位置情報とその追跡情報に従って、ボールとあるプレイヤーが接触しているフレームがキーフレームとして検出される。

ここで、画像上では接触していても実際には接触していない場合、すなわちプレイヤーの前後をボールが通り抜ける場合も考えられる。このような問題に対しては、次のような規則に従って対処する。画像上で接触しているフレームの前後数フレーム目におけるプレイヤーからボールへの向きを比較し、それらの向きが互いにほぼ逆であり、画像上で接触しているフレームにおけるプレイヤーとボールの接触点がプレイヤー領域の上半分の部分に位置している場合は接触していないとみなし、それ以外を接触しているとみなす。

また、プレイヤーがドリブルなどにより一時的にボールと離れる場合、すなわちボールとの接触を終了したプレイヤーと次にボールとの接触を開始するプレイヤーが同じである場合は、一時的に離れているすべてのフレームもキーフレームとして扱うことにする。これは、ボールを所持しているとみなされるプレイヤーが長時間に渡って同じであるとき、他のプレイヤーとの関係が変化することを考慮するためのものである。なお、それぞれのキーフレームにおいてボールに接触しているプレイヤーをそのキーフレームにおけるボール所持プレイヤーとする。

モデルシーンから検出されたキーフレームはデザイナーに提示され、デザイナーは各キーフレームに対して状況概念に関する知識の記述を行う。しかし、提示されたキーフレームの中に時間的に隣り合うキーフレーム群が存在する場合、それらにそれぞれ同じ内容を記述することはデザイナーにとって無駄な作業である。そこで、時間的に隣り合うキーフレーム群が存在すれば、そのキーフレーム群の中から状況概念を構成すべきキーフレームのみをデザイナーが任意に選択できるものとする。

5 状況概念の表現形式

5.1 状況概念の構成に必要な情報

システムは、状況概念を構成するためにデザイナーが指定したモデルシーンに対して画像処理を行い、以下のような情報を取得する。

- 各プレイヤーとボールの位置
- チーム区別を含む、各プレイヤーを表すラベル
- キーフレームとそのキーフレームにおけるボール所持プレイヤーを表すラベル

またデザイナーは、システムから提示されたモデルシーンの各キーフレームに対して、

- プレイヤー間の空間的な相関関係

を記述する。

システムは、画像処理により抽出した情報とデザイナーにより定義付けられた知識を組み合わせることで状況概念を構成する。

5.2 関係記述子

本研究では、プレイヤー間の空間的な相関関係を記述するためのプリミティブな関係記述子 (RDF: Relation Description Factor) を以下のように定義する (ただし、 A, B, C はそれぞれプレイヤーのラベルを表し、各 RDF 中のプレイヤーのラベルのうち少なくとも1つがいずれかのキーフレームにおけるボール所持プレイヤーを表すものとする)。

$near(A, B)$: あるボール所持プレイヤー A と別のプレイヤー B との間の距離が近いことを表す。 $near$ が指定されると、システムはデザイ

ナが近いと認めた AB 間の距離の値を保持する。ただし、この値は撮影時のカメラのズーム条件を考慮すると相対的な値であるため、さらにその値を A と B の領域の y 方向の長さの平均で割った値（絶対距離と呼ぶことにする）を保持するものとする。マッチングにおいてこの条件を満たすのは、モデルシーンの A と B に相当する動画像データのプレーヤー A' と B' との間の絶対距離が AB 間の絶対距離に許容量を加えた値より小さい場合とする（図 2 参照）。

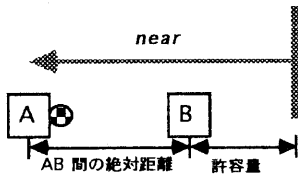


図 2 $near(A, B)$

$far(A, B)$: あるボール所持プレーヤー A と別のプレーヤー B との間の距離が離れていることを表す。 far が指定される場合も、システムはデザイナーが離れていると認めた AB 間の絶対距離の値を保持する。マッチングにおいてこの条件を満たすのは、 $A'B'$ 間の絶対距離が AB 間の絶対距離から許容量を引いた値より大きい場合とする（図 3 参照）。

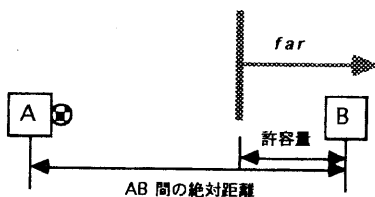


図 3 $far(A, B)$

$x_order(A, B)$: プレーヤー A とプレーヤー B との間の x 軸方向の順序関係を表す。これは、プレーヤー間の順序関係が時間的に変化することを表現するためのもので、複数のキーフレーム間で同じプレーヤーに対して指定することが前提条件である。システムはこの AB 間の x 軸方向の順序関係を保持する。マッチ

ングにおいてこの条件を満たすのは、 $A'B'$ 間の x 軸方向の順序関係が AB 間の x 軸方向の順序関係と同じ場合とする（図 4 参照）。なお、画面上での攻める向きの違いによる順序関係の対称性を考慮して順序関係が逆のときもこの条件を満たすものとするが、その場合他のキーフレームにおいて指定されたすべての x_order の順序関係が逆であるときに限る。

$y_order(A, B)$: プレーヤー A とプレーヤー B との間の y 軸方向の順序関係を表す。 x_order と同様に、複数のキーフレーム間で同じプレーヤーに対して指定することが前提条件である。マッチングにおいてこの条件を満たすのは、 $A'B'$ 間の y 軸方向の順序関係が AB 間の y 軸方向の順序関係と同じ場合とする（図 4 参照）。同様に、この順序関係が逆の場合でも、他のキーフレームにおいて指定された y_order の順序関係がすべて逆であればこの条件を満たすものとする。

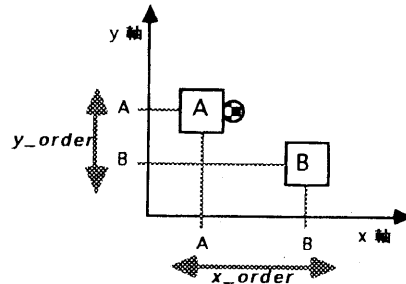


図 4 $x_order(A, B)$ と $y_order(A, B)$

$between(A, B, C)$: プレーヤー A がプレーヤー B とプレーヤー C の間に位置していることを表す。ここで、 A が B と C の間に位置しているとは、線分 BC を直径とする円の内部に点 A が存在しているものとする。 $between$ は、 x_order や y_order と異なり、 x 軸や y 軸のような固定された方向軸に依存しない。点 B と点 C の位置により決まる直線 BC がこの場合の方向軸となり、その方向軸上での順序関係を表すものである。マッチングにおいてこの条件を満たすのは、 A' が B' と C' の間に位置している場合とする（図 5 参照）。

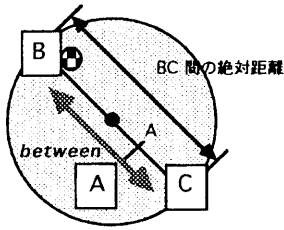


図5 between (A, B, C)

ϕ : そのキーフレーム内のプレーヤー間に、状況概念を構成すべき関係が存在しないことを表す。
 ϕ が指定された場合、マッチングにおける評価の対象はそのキーフレームにおけるボール所持プレーヤーのラベルのみである。

このような関係記述子を用いることにより、プレーヤー間の空間的な相関関係を抽象的に表現することができる。なお、同じキーフレームに対して複数の RDF を指定しなければならない場合は、それらに論理積 (AND) を用いて記述するものとする。

5.3 状況概念の構成

以上のような情報を基に、状況概念 C を以下のような形式で表現することができる。

$$C = S(f_{k_0}) \rightarrow S(f_{k_1}) \rightarrow \dots \rightarrow S(f_{k_{n-1}})$$

$$S(f_k) = \langle P(f_k), R(f_k) \rangle$$

ここで、 f_k はデザイナーにより選択されたキーフレーム、 $P(f_k)$ はキーフレーム f_k におけるボール所持プレーヤー、 $R(f_k)$ はキーフレーム f_k における RDF、 $S(f_k)$ は $R(f_k)$ と $P(f_k)$ から成る組、 n はデザイナーにより選択されたキーフレームの数、 \rightarrow は $S(f_{k_i})$ から $S(f_{k_{i+1}})$ への時間的な変化を表す。このように、状況概念はそれぞれのキーフレームにおけるボール所持プレーヤーと RDF から成る組の時間的な変化として構成される。

例として、図6に示されるような壁パスという状況概念を上記のような形式に従って表現すると、以下ようになる。

$$\begin{aligned} C = & \langle A_1, \text{near}(A_1, B_1) \rangle \\ & \wedge x_order(A_1, B_1) \rangle \\ & \rightarrow \langle A_2, \phi \rangle \\ & \rightarrow \langle A_1, x_order(A_1, B_1) \rangle \end{aligned}$$

なお、 \wedge は RDF 間の論理積とする。

6 マッチング処理

動画データとデザイナーが指定した状況概念の構成データとのマッチング処理は、キーフレーム単位で行われる。キーフレーム単位に抽象化されたデータ間でマッチングを行うことにより、検索の度に画像処理を行う必要がなく、処理時間の短縮が期待できる。

マッチングアルゴリズムの概略を以下に示す。

```

for (i = 0; i < nv - n + 1; i++)
{
  cnt = 0;
  for (j = 0; j < n - 1; j++)
    if (P(i + j) から P(i + j + 1) への変化と
        P(j) から P(j + 1) への変化が等しい)
      if (j == 0)
        if (動画データから抽出された i 番目の
            キーフレーム内に R(0) を満たす
            プレーヤーが存在する)
          if (動画データから抽出された i +
              1 番目のキーフレーム内に R(1)
              を満たすプレーヤーが存在する)
            cnt++;
          else
            break;
        else
          break;
      else
        if (動画データから抽出された i + j
            + 1 番目のキーフレーム内に R(j
            + 1) を満たすプレーヤーが存在する)
          cnt++;
          else
            break;
        else
          break;
    else
      break;
  else
    break;
}

```

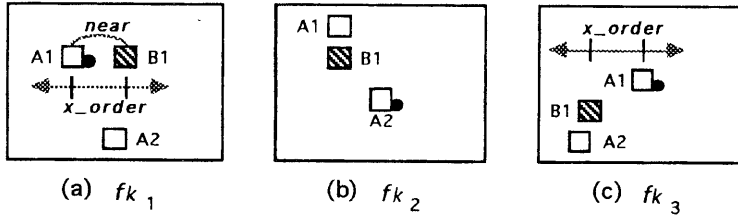


図6 壁パスを構成するキーフレームの例

```

if (cnt == n - 1)
    マッチシーン検出;
}

```

ここで、 i は動画データから抽出されたキーフレームの番号を表すパラメータ、 nv は動画データから抽出されたキーフレームの数、 j は状況概念を構成するキーフレームの番号を表すパラメータ、 n は状況概念を構成するキーフレームの数、 cnt はカウンタを示すものとする。

マッチングにおける処理内容は、主にボール所持プレイヤーの変化の確認と動画データのキーフレーム内に RDF を満たすプレイヤーが存在するかどうかの確認である。ボール所持プレイヤーの変化の確認は、ボール所持プレイヤーのラベルのパターンが一致しているかどうかにより判断される。ここでパターンとは、キーフレームの時間的な変化とともに移り変わるボール所持プレイヤーのラベルの履歴である。また RDF を満たすプレイヤーの検出処理では、デザイナーに指定された RDF に対してその評価条件 (5.2 節参照) を満たすプレイヤーがキーフレームごとに保持され、その中でプレイヤーのラベルの時間的な整合性がとれるものに絞られる。このような処理をくり返し、すべての条件が満たされるシーンがマッチシーンとしてデザイナーに出力される。

7 実験

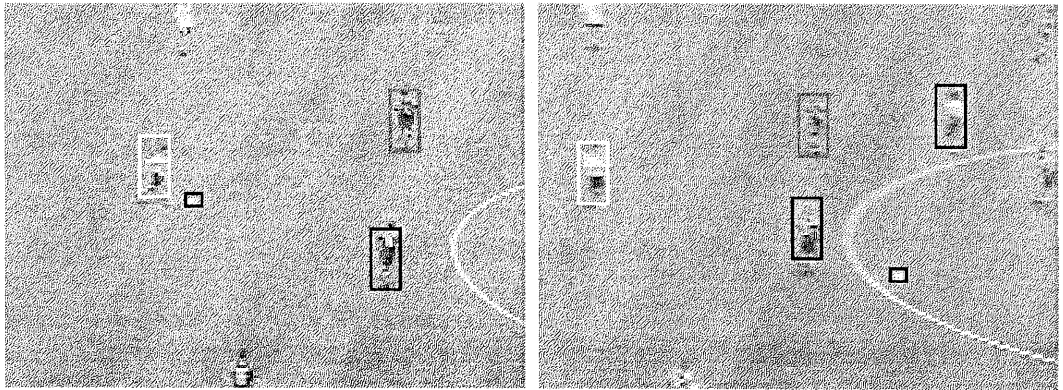
本研究では、画像処理における物体抽出処理、及びマッチング処理の実験を行った。実験で扱った動画は、フレームサイズが 160×120 ピクセル、フレームレートが 15 フレーム/秒、使用されている色数が 1677 万色中の 256 色のテレビで放映されたサッカーの試合映像である。

まず画像処理の実験として、動画データに対してボールとプレイヤーの抽出処理を行うと、図7のような結果が得られる。 f_0 、 f_5 はそれぞれ 0 番目と 5 番目のフレームを表し、フレーム内のそれぞれの矩形は抽出されたボールの領域、あるいはプレイヤーの領域を示している。プレイヤー領域を表す矩形については、その追跡情報に基づいて同一プレイヤーは同色で表示されている。CPU が *Pentium 150 MHz* の PC 上で、10 フレームの動画データに対してそれぞれの抽出処理を実行したときの処理時間は、ボールの場合が平均 2.17 秒で、プレイヤーの場合が平均 2.88 秒であった。

マッチング処理の実験は、時間的に隣り合うキーフレーム群が存在しないという条件の下で、キーフレーム数の平均が 3.60 である 10 個のシミュレーションデータに対して行った。この実験では、実際に壁パスという状況概念が含まれるデータを 10 個のうち 2 つ用意し、壁パスに関する知識を定義付けたデータによる検索のシミュレーションを行い、壁パスを表す 2 つデータがマッチするという正確な結果が得られた。なお、この実験での処理時間は 0.0633 秒であった。

8 おわりに

動画に含まれる物体間の時間的、空間的な相関関係に注目し、このような特徴に基づいて所望のシーンを検索する手法について述べた。本手法では、状況概念を抽象的な関係表現で記述することができ、また動画データの抽象化によって検索処理時間の短縮が可能となる。さらに、状況概念に関する知識の定義付けにモデルシーンを利用することによって、構成に必要な多くの情報を容易にかつ効率良く入力することができるようになる。



(a) f_0

(b) f_5

図7 物体抽出処理の結果

実験では、シミュレーションデータによるマッチング処理を行い、本手法の有効性を検証した。現在、システムは MS-Windows 上で C 言語を用いて構築中であり、今後は、画像処理部分とマッチング処理部分を組み合わせ、自動化されたシステムに対して検索実験を行う予定である。

参考文献

- [1] A. Yoshitaka, Y. Hosoda, M. Yoshimitsu, M. Hirakawa, and T. Ichikawa, "VIOLONE: Video Retrieval by Motion Example", *Journal of Visual Languages and Computing*, Dec. 1996 (to appear).
- [2] T. Satou and M. Sakauchi, "Video Acquisition on Live Hypermedia", *Proc. of the International Conference on Multimedia Computing and Systems*, pp. 175-181, 1995.
- [3] T. C. T. Kuo and A. L. P. Chen, "A Content-Based Query Language for Video Databases", *Proc. of the International Conference on Multimedia Computing and Systems*, pp. 209-214, 1996.
- [4] M. Frickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content: The QBIC System", *IEEE Computer*, Vol. 28, No. 9, pp. 23-32, 1995.
- [5] 安部 伸治, 外村 佳伸, "状態の時間変化をキーとする動画検索法", *信学論*, Vol. J75-D-II, No. 3, pp. 512-519, 1992.
- [6] Y. F. Day, S. Dağtas, M. Iino, A. Khokhar, and A. Ghafoor, "Spatio-Temporal Modeling of Video Data for On-Line Object-Oriented Query Processing", *Proc. of the International Conference on Multimedia Computing and Systems*, pp. 98-105, 1995.
- [7] I. M. Walter, R. Sturm, P. C. Lockemann, and H.-H. Nagel, "A Semantic Network Based Deductive Database System for Image Sequence Evaluation", *Proc. of the IFIP 2nd Working Conference on Visual Database Systems*, pp. 251-277, 1991.
- [8] Y. Gong, L. T. Sin, C. H. Chuan, H. J. Zhang, and M. Sakauchi, "Automatic Parsing of TV Soccer Programs", *Proc. of the International Conference on Multimedia Computing and Systems*, pp. 167-174, 1995.
- [9] 田中 和久, *サッカーの戦術*, 不味堂出版, 1981.