

4H-4

# 鳥類図鑑Hyperbookにおける 個人の記憶に基づく動画像検索

根岸晃 田淵仁浩 村岡洋一  
早稲田大学 理工学部

## 1.はじめに

現在我々は、電子化図書システムHyperbook[1]を研究・開発中である。Hyperbookはマルチメディアデータベースであり、研究においては例として鳥類図鑑を作成している。

鳥類図鑑Hyperbookで動画像を扱う場合、従来の方法では問題がある。そこで本稿では、鳥の動画像を対象とし、利用者の記憶をてがかりとして検索する方法を提案する。

## 2.個人の記憶に基づく動画像の検索

検索の手がかりとして重要な情報を2.1で述べ、それに基づいて検索を行うときの従来の手法の問題点を2.2で述べる。

### 2.1 検索の手がかりとなる情報

動画像検索の流れを図1のように捉えた時、手がかりとなる情報は次の二つの条件を満たしている必要がある。

- a. 動画像を見て再生可能な記憶になる
- b. 場面を探すときキーとなる

実際の鳥の動画像の場合は、次の3つの情報が重要な手がかりとなる。

- a. 登場物
- b. 登場物の相対的位置関係
- c. 登場物の属性(鳥・枝などの向き,鳥の動作など)  
(ただしb.,c.は時間推移を含む)

これは、鳥の動画像を見た複数の人に、覚えていることを任意の方法で書いてもらい、その場面を理解し捜し出すためにキーとなるものを調べた結果である。

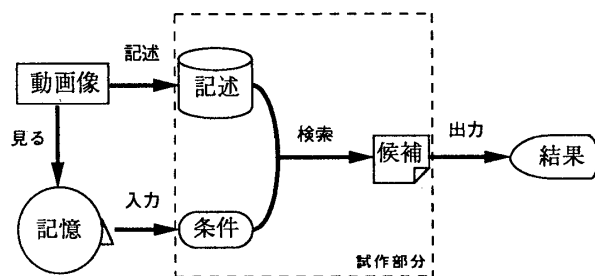


図1 動画像検索の流れ

### 2.2 従来の手法の問題点

動画像の検索方法としては、静止画と同等に扱う方式[2]や、物体の動きをキーとする方式[3]が提案されているが、鳥の動画像を検索するにはいくつかの問題がある。

Retrieval of motion picture on birds by personal memory  
in Hyperbook of birds  
A.Negishi M.Tabuchi Y.Muraoka  
Waseda University

前者の方法では、キーワードによる検索であるため、登場物の位置関係や時間推移をともなった動作などによる検索は行えない。

また後者の方法は、一般的な動画像の位置・色などの、表面的・物理的な因子から考え、その組合せで[動作]を捉えようとしている。しかし鳥の動画像では[餌を食べる]や[片足で立っている]など、単純な位置などの組合せでは記述することが困難な場合が出てくる。また、人間の目には明らかな奥行き方向の位置関係なども画像処理を利用して特定するのは現状ではまだ難しい。

## 3.鳥の動画像検索の実現方法

鳥の動画像検索の方針を3.1で述べ、動画像の記述形式検索手順をそれぞれ3.2,3.3に示す。

### 3.1 鳥の動画像検索の方針

2.1で述べた情報をもとに検索できるようにするには、覚えている範囲で検索条件を構成できなくてはならない。また、必ずしも正しい記憶とは限らないため、検索条件と2次データとを比較して類似した場面を検索する。なお、全くつながりのない複数の場面は、1つの検索条件では扱わない。検索条件の記述においては、以下の方針をとる。

#### (1) 登場物・動作など

対象となる動画像は分野が特定されているため、登場物や動作は、あらかじめ決めておいた候補から、覚えている情報を選択すればよい。

#### (2) 登場物の位置関係

記憶においては登場物同士の絶対的な距離や角度は曖昧となるため、どちらが右にあるかなどの相対的な位置関係のみを検索に扱う。

#### (3) 時間推移

位置関係や動作などには時間推移がある。しかし、時間的な長さは記憶では曖昧となるため、順序関係のみを検索に扱うことにする。

### 3.2 動画像の記述

動画像を扱うときに、カット(カメラの切り替わりから次の切り替わりまで)を時間的な単位とする。これは内容的な切り替わりがカメラの切り替わりとなることが多い点と、単位を決定するときの一意性を考慮したためである。以下にカットの記述内容を示す。

#### (1) 登場物

各カットにおける登場物に対して、そのカット内で有効なIDを割り振る。

#### (2) 登場物の相対的位置関係

画面に対して横(右)・縦(上)・奥行き(手前)の3方向にx軸・y軸・z軸を割り当てる(括弧内が正の向き)。そのカットの登場物の個数をn個とすると、ある瞬間の登場物の相対的位置関係は3つの $n \times n$ 行列 $(x_i)_j, (y_i)_j, (z_i)_j$ となる。行列 $(x_i)_j$ の要素は次の値をとる。

$$x_{ij} = \begin{cases} -1 & (\text{登場物}i\text{に対して登場物}j\text{は左}) \\ 0 & (\text{登場物}i\text{と登場物}j\text{は横方向でほぼ同位置}) \\ +1 & (\text{登場物}i\text{に対して登場物}j\text{は右}) \\ \times & (\text{定義不可}) \end{cases}$$

ただし、行列の要素値は記述者の判断により定める。「×(定義不可)」は、例えば棒状物(枝,幹など)同士でどちらが右かを言及できない場合などに用いる。行列 $(y_{ij}), (z_{ij})$ の要素も同様に定める。

このような瞬間の相対的位置関係を時間推移に沿って並べた列を記述する。

(3) 登場物の属性

登場物の属性は、動画像上に視覚的に現れる情報に限る。これは、登場物の種類により異なる。例として、鳥の属性の向きと動作について述べる。

鳥の向きは、頭(クチバシ)と胴体(背骨)の2つの矢印で表現し、それぞれ3つの軸方向において{-1,0,+1}のうちのひとつをとるように記述する。

鳥の動作を表す言葉を、文献[4]に基づいて、

- a. 基本的運動(状態)
  - [立っている],[飛んでいる]などの鳥の基本的な一状態
- b. 各部の動作
  - [首をのばす],[クチバシを開く]などの部分的な動作
- c. 合目的行動
  - [餌を探す(摂食のため)],[踊る(求愛のため)]などある目的をともなった行動

の3つに分類した。さらにこれを階層的に分類し、各々の言葉に一意の番号を割り当てる。その番号によって、カット内の鳥の動作を記述する。

鳥の向きとこれら3種の動作も相対的位置関係と同様に、時間推移を記述列で表現する(記述列は、向き・3種の動作の4本を独立に作る)。

(4) 記述列の時間的対応

複数の記述列の時間的対応をとるために、記述者が任意に定めた瞬間における各記述項目の対応を記述する(図2)。

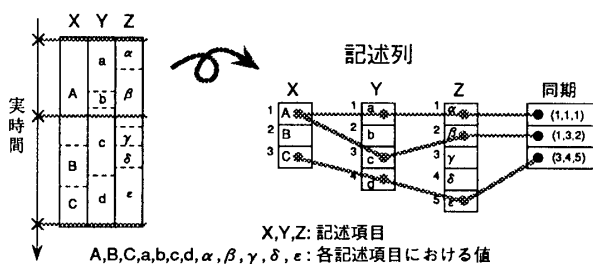


図2 時間同期の記述

3.3 検索手順

検索は、利用者の検索条件を3.2の記述形式に翻訳し、翻訳された条件と、動画像の2次データとを比較して行なう。比較の際に、動画像の1カットをあらかじめ記述した2次データ(以降、データ)と、検索条件(以降、条件)との距離を計算し、得られた距離値がある値以下のカットの集合を検索結果とする。カット間の距離は記述列の直積距離としている。以下に距離の計算手順を示す。

(1) 登場物のマッチング

データと条件の登場物の対応がとれるかどうかを調べる。この時、同じ登場物同士だけでなく、例えば「鳥」

と「雛」も対応するようにする。マッチングがとれなければそのカットは探しているものとは違うと判断し(距離=∞)、とれば、考えられる全ての対応の場合について(2)の比較を行い、最も小さな距離を返す。

(2) 位置関係・属性の比較

チャンク毎に、a,bを繰り返し、それぞれにおいて算出した距離の総和を返す。

a. 位置関係条件が最もよく一致する(つまり距離が小さい)位置関係データとの距離を返す。この距離は、3.2(2)で述べた行列の対応する要素ごとに差をとり(ただしどちらかが×(定義不可)の場合は0とする)、その絶対値の総和により算出する。

b. a.で決定した位置関係データと時間的に対応しうる属性データを調べ、最もよく属性条件と一致する(つまり距離が小さい)データとの距離を返す。この距離は、鳥の向きでは、頭と胴体の矢印それぞれについて条件とデータのなす角を用いる。また、鳥の動作では、条件とデータの相違点の個数を用いる。ただし、条件とデータのどちらかが無記述の場合は除く。

4. 実験・評価

3.で述べた方法に基づいて動画像検索の核部分(図1の点線内部)を試作し、提案した動画像検索法を検証するために実験を行った。実験では、9種類の鳥、全22カットの動画像(5分弱)を被験者に提示し、登場物、登場物の位置関係、鳥の向き・動作、およびこれらの時間推移を、記憶している範囲で記入してもらった。これをもとに、試作したプログラムにより距離を算出すると、3分の2以上の場合目的のカットが検索条件の近傍第3位までに含まれるという良好な結果を得ている。

5. おわりに

本稿では、利用者の記憶に基づいて動画像を検索する方式を提案した。提案した方式では、登場物やその属性(動作など時間推移を含む)に限られているような動画像において、利用者の記憶をもとにして動画像を探すことが可能となる。ただし、

- a. 登場物を勘違いすると、目的のカットが「登場物のマッチング」の段階で蹴られてしまうことがある。
- b. 距離の計算において、位置関係や各属性の間の重み付けを決める指針がまだない。
- c. カット間のつながりを考慮していないため、複数のカットにまたがる検索条件に対応できない。
- d. 背景,群れといった不定形の広がりを持つ登場物を十分に扱えない。

といった問題点があり、その解決が今後の研究課題である。

【参考文献】

[1]田淵,村岡:"電子化図書構想とデータモデリング",情処35回全大6Bb-5,pp.363-364,1987.  
 [2]柴田:"連想記憶可能な画像データベース",信学技報IE88-117  
 [3]安部,石橋,笠原:"物体の動きをキーとした動画像検索法",信学技報IE88-4  
 [4]中村:"野鳥の図鑑水の鳥①",pp180-207,保育社