

動画像内容検索方式の提案

6Q-4

渡邊 裕司 鈴鹿 豊明

日立ソフトウェアエンジニアリング（株）

1 はじめに

動画像は、時刻順に並んだ複数の静止画フレームの集合である。各フレームには、先頭から時刻順に番号が付けられており、この番号はフレーム番号と呼ばれている。再生を開始・終了するフレーム位置をフレーム番号で指定することで動画像の一部を再生できる。

単に動画像を再生するだけであれば、このフレーム番号によるアクセスだけでも十分である。しかし、動画像中に写っているもの（以下、動画オブジェクトと呼ぶ）を表示画面上でマウスを用いて指定し、その指定された動画オブジェクトに関連する情報を検索するというような機能を実現するためには、まず、マウスで指定した位置にどの動画オブジェクトが表示されているのかを知る仕組みが必要となる。しかも、ユーザは任意のフレーム位置で動画オブジェクトを指定する可能性があるので、すべてのフレームに対して、ユーザがマウスで指定した位置にどの動画オブジェクトが表示されているのかを知る必要がある。

動画像のすべてのフレームに対して、各動画オブジェクトが写っている領域を逐一入力することは、現実的に不可能であり、以前からさまざまな解決案が提案されている。例えば[1]では、

1. 動画オブジェクトの動きを代表するフレームを選択し、各フレーム上で動画オブジェクトが写っている領域を入力する。
2. その間のフレームに対しては直線で補間する。

という方法を採用しており、領域を入力する作業の工数を低減させている。しかし、直線補間を用いる方法では、複雑な動きをしている動画オブジェクトに対して、動きを忠実に再現することは難しい。

本稿では、複雑な動きをしている動画オブジェクトに対しても適用可能である、動画オブジェクトの動きのモデル化と、それを用いた検索手法について提案する。

Consideration on Moving Picture Content Search
Yuji Watanabe, Toyooki Suzuka
Hitachi Software Engineering Co., Ltd.

2 動きのモデル化

各動画オブジェクトの動きを、

1. 動画オブジェクトが写っている先頭のフレーム位置 f_H
2. 動画オブジェクトが写っている末尾のフレーム位置 f_T
3. 動画オブジェクトが写っている領域をフレーム番号 f を用いて表した関数 $R(f)$

という3種類のパラメータで表現する。

動画オブジェクトが写っている領域は、動画オブジェクトを囲む外接矩形で表現する。外接矩形を表現するには、左上・右下の頂点の座標がわかれば十分である。従って、上記3.の $R(f)$ は具体的には、外接矩形の左上の頂点の座標 $(Xmin(f), Ymin(f))$ と右下の頂点の座標 $(Xmax(f), Ymax(f))$ で構成する。動画オブジェクトの動きをモデル化する工程の模式図を図1に示す。

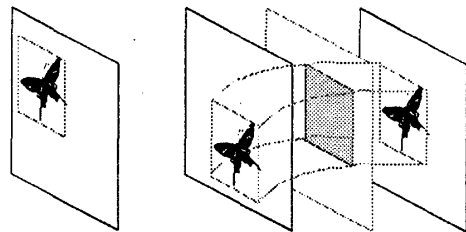


図1: 動きのモデル化

$R(f)$ を求めるには、 $R(f)$ が十分忠実に動画オブジェクトの動きを再現するようになるまで、以下の手順を繰り返す。

1. 動画オブジェクトの写っている幾つかのフレームを選択し、動画オブジェクトを囲む外接矩形を入力する。
選択したフレームの総数 N とすると、 i 番めのフレームのフレーム番号 f_i と、入力した外接矩形の左上の頂点の座標 $(Xmin_i, Ymin_i)$ 、右下の頂点の座標 $(Xmax_i, Ymax_i)$ が得られたことになる。
($i = 0, 1, \dots, N-1$)

2. $Xmin(f_i) = Xmin_i (i = 0, 1, \dots, N-1)$ を満足する3次の自然スプライン関数[2]を決定し、それを $Xmin(f)$ とする。 $Ymin(f), Xmax(f), Ymax(f)$ についても、同様に決定する。

また、 f_H と f_T については、それぞれ、最終的な f_0 と f_{N-1} で近似する。

3 動画オブジェクトの検索

第2節では、動画オブジェクトの動きをモデル化し、そのパラメータ $f_H, f_T, R(f)$ を求める方法を示した。本節では、それらのパラメータを利用して、ユーザがマウスで指定した位置に表示されている動画オブジェクトを決定する方法について説明する。

ユーザが、フレーム番号 F のフレーム上の (X, Y) という位置を指定したとする。以下の手順を実行して、その位置に表示されている動画オブジェクトを決定する。

1. ユーザが指定したフレーム F に写っている動画オブジェクトをすべて選び出す。
2. 選び出された動画オブジェクトのうち、フレーム F 上での表示領域 $R(F)$ が、ユーザが指定した位置を含んでいるものを、ユーザが指定した動画オブジェクトとみなす。

図2は、各動画オブジェクトが写っているフレーム範囲を表したグラフの例である。図中の各線分の左端と右端はそれぞれ、その動画オブジェクトの f_H と f_T を表している。

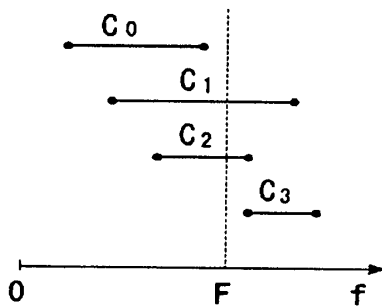


図2: 動画オブジェクトの表示範囲(1)

この例では、 C_0, C_1, C_2, C_3 という4つの動画オブジェクトが写っており、そのうち、フレーム F に写っている動画オブジェクトは C_1 と C_2 である。一般的には、次の条件式

$$f_H \leq F \leq f_T$$

を満足するような動画オブジェクトを抽出すればよい。しかし、動画オブジェクトの総数が非常に大きい場合には、各動画オブジェクトに対して、上の条件式を評価していたのでは処理の効率が非常に悪くなる。

そこで、図3に示すように、 f_H を横軸に、 f_T を縦軸にとり、動画オブジェクトが写っているフレーム範囲を2次元の点で表現する。フレーム F に写っている動画オブジェクトは、図中の網かけの領域に含まれている点を検索することで抽出できる。

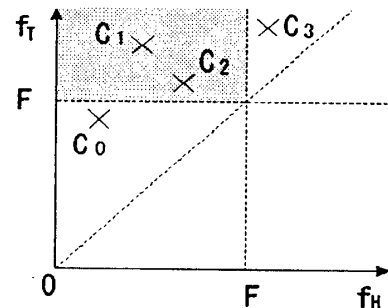


図3: 動画オブジェクトの表示範囲(2)

この2次元の点として表現された、動画オブジェクトが写っているフレーム範囲をk-d木[3]と呼ばれる木構造で管理する。k-d木は、空間を点データ数が半分ずつになるように順次領域分割していき、最終的に1点となるまで分割していき、その構造を蓄積しておくことができる木構造であり、空間の範囲検索に適している。

4 まとめ

動画像中の動画オブジェクトの動きを3次の自然スプライン関数で表現し、また、動画オブジェクトが写っているフレーム範囲をk-d木に蓄積して管理することで、動画オブジェクトを検索する手法を提案することができた。

参考文献

- [1] 川崎, 平田, 高野, 原: “ハイパーメディア「雅」における分散動画実装方式”, 情報処理学会第49回全国大会 1W-4 (1994).
- [2] 桜井, 石井, 吉村, 高山: “スプライン関数入門”, 東京電機大学出版局 (1981).
- [3] 坂内, 大沢: “画像データベース”, 昭晃堂 (1987).