

ビデオデータベース検索用のインデックス構造

4U-6

樋渡 良継 伏木田 勝信 脇 英世[†]

通信・放送機構 東京臨海部リサーチセンタ †東京電機大学

1. はじめに

デジタル映像ライブラリが普及し、インターネットを通して、簡単に大規模な映像データベースにアクセスできるようになった。利用者が大規模な映像データベースをブラウジングし、効率的に検索できるインデックス構造の提供が求められている。ここでは画像検索エンジンによるクラスタリング方式で構築したインデックス構造、視覚化音インデックス[1,2,3]にこのインデックス構造を適用し行った検索実験と、その結果について報告する。

2. ビデオインデックスのクラスタリング

(1) インデックスのクラスタリング

図1に示すように、インデックスのクラスタリングは、インデックスデータベースの全インデックスに対し、画像検索エンジンを使って類似度を計算し、この類似度をもとにインデックスをクラスに分類する。

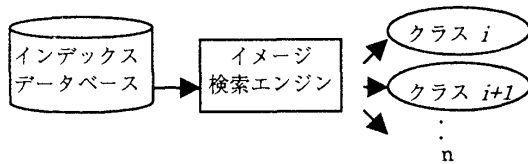


図1 インデックスのクラスタリング

クラスタリング方法として、平田氏は検索の高速化のため、画像のマッチングプロセスに複数のクラスに重複して画像を分類する、オーバーラッピングクラスタリング方式を採用し、その成果を報告している[4]。しかし、この方式によるクラスタリングでは、データベースの種類によって、クラス数を特定し分類することは難しく、全データを網羅して分類しようとするクラス数が増加する。我々は、比較的短時間で簡単にクラスタリングが可能なシリアル分割と、k近傍決定則(k-NN)によるクラスタリングを行い、オーバーラッピングクラスタリングとの検索効率についての比較評価を行った。

●シリアル分割クラスタリング

データベースの中から任意に選択した入力キーを

An index structure for video database / Yoshitsugu Hiwatari, Katsunobu Fushikida and [†]Hideyo Waki / Tokyo Waterfront Research Center, TAO Japan [†] Tokyo Denki University

使って、画像検索エンジンによる類似度計算を行い、類似度の高い順に n 個のブロックに分割、それぞれのブロックをクラスとする。

●k-NNクラスタリング

データベースの中から任意に入力キーを選択し、画像検索エンジンの類似度計算で、類似度の高いk個のインデックスを1つのクラスにする。次に類似度計算でリジェクトされたインデックスの中から、新たに入力キーを選択し、先のk個のインデックスを除いたデータベースを対象に類似度計算を行い、次のクラスを作る。これをn個まで繰り返す。

(2) インデックス構造

シリアル分割クラスタリングでは、分割したブロックの先頭のインデックスを各クラスの代表インデックスとした。また、k-NNクラスタリングでは、入力キーとして使った全てのインデックスを各クラスの代表インデックスとした。この代表インデックスと、それに対応するクラスを階層構造に配置してインデックス構造を構築している。図2にk-NNクラスタリングで構築したインデックス構造の例をあげる。

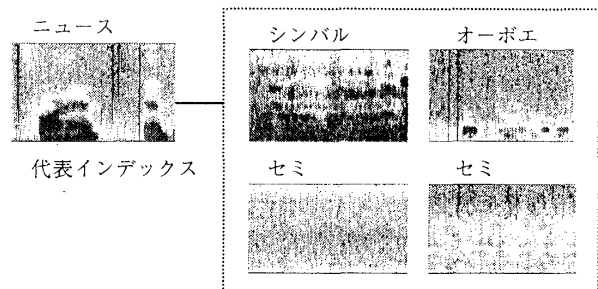


図2 インデックス構造の例

(3) 画像検索エンジンによる類似度計算

インデックスを分類する類似度の計算には、画像検索エンジンのマッチングアルゴリズム[4,5,6]を使った。画像検索エンジンは画像を 24×24 ピクセルの大きさに正規化、概略画像を作成し、以下のマッチング処理を行い類似度を計算する。

- ・概略画像を領域分割
 - ・キー画像と検索画像間で領域の対応付け
 - ・対応領域間で形、色の距離を基に類似度計算
- 画像の類似度 I は形類似度 S と色類似度 C の線形

和で求めている。

$$I = s \cdot S + c \cdot C \quad (1)$$

ここで s, c は重み係数である。

3. 検索実験

検索実験では、音楽、ニュース、スポーツ、昆虫と鳥などのビデオから抽出して作った視覚化音声インデックス[7]を 3,000 以上持つデータベースを使った。このデータベースを 60 個のクラスに分類しインデックス構造を作成した。さらに、検索の目的インデックスを設定、この目的インデックスが含まれるクラスの代表インデックスを検索することで検索効率を求めた。実験に先立って、目的インデックス 100 個と、入力キー 200 個をデータベースから選択した。目的インデックス 1 個に対し、そのインデックスと同じビデオシーンから抽出した 2 個の入力キーで検索する。

この検索実験では、3 種類の分類法で作ったインデックス構造の検索率を比較するために、検索精度 R_a を次のように定義した。

$$R_a = N/Tt \times 100 \quad (2)$$

ここで N は規定ランク内に入った目的クラスの数、 Tt は検索回数である。実験では規定ランクを 10 とし、11 以上をリジェクトとした。

(1) 実験で使用したインデックス構造

次にあげる 3 種類のクラスタリングを行い、インデックス構造を構築し実験を行った。いずれのクラスタリングでも、最初の入力キーはデータベースから任意に選択したインデックスを使っている。

① オーバーラッピングクラスタリング

k -NN 法と同じ手順でクラスタリングをする。インデックスのオーバーラップを許すためデータベースからはクラスに分類されたインデックスを除かない。ここでは 60 個のクラスを作るため、オーバーラッピングで 40 クラスを作成、これらのクラスに属していないインデックスをシリアル分割し、20 クラスを作成した。

② シリアル分割クラスタリング

任意に選択した入力キーで、類似度計算を行い、類似度の高い順から 50 個のインデックス毎に分割し、60 クラスを作成した。

③ k -NN クラスタリング

このクラスタリングでは、初めに 200 個のインデックスをもつ 15 クラスを作成し、それぞれから 50 個のインデックスを持つクラスを 4 個作成、全体で 60 クラスを作成した。

(2) 実験結果

表 1 に 3 種類のインデックス構造を使った検索実験での検索精度の比較を表す。実験結果から、オ

ーバーラッピングクラスタリングの検索精度は、65%、リジェクト数が 70 であった。しかしこの結果は、60 クラスに限定するためにシリアル分割クラスタリングと併用したので、それが検索精度に影響していると考えられる。また、この方式のインデックス重複の検索への効果は、検索精度で 3% であった。表の中で、シリアル分割クラスタリングの検索精度が最も低いが、1 度の類似度計算で、代表インデックスとクラスの類似度の関係が最も弱い場合、この結果になったと考えられる。これは類似度計算を数回実行することで検索精度を向上できると考えている。この実験では、 k -NN クラスタリングの検索精度が最も高く、このインデックス構造によってビデオデータベースを効率的に検索できることが確認できた。しかし、 k -NN 法によるクラスタリングはデータベースが大規模になるに従い、処理時間が長くなることが知られており[8]、これをシリアル分割で補完するなど、クラスタリング方式の併用を検討する必要があると考えている。

表 1 検索精度の比較

インデックス構造	検索精度(%)	リジェクト
オーバーラッピング	65	70
シリアル分割	36	128
k -NN	94	12

4. まとめ

視覚化音声インデックスデータベースを、画像検索エンジンを使いクラスタリングし、3 種類のインデックス構造を構築、それぞれの検索精度を求め比較した。3 種類のインデックス構造を使った検索精度の比較では、 k -NN クラスタリングのインデックス構造による検索精度が最も高く、このクラスタリングによる音インデックス構造が、ビデオ検索に有効であることが確認できた。今後は、大規模ビデオデータベースへの適用について検討を進める予定である。

参考文献

- [1] 樋渡、伏木田、脇、画像-音インデックスによる動画像検索、情報処理学会 56 回全国大会 1998. [2] K. Fushikida, Y. Hiwatari, H. Waki, "A Content-Based Video Retrieval Method Using a Visualized Sound Pattern" Proc., IFIP VDB4, pp.208-213, 1998. [3] K. Fushikida, Y. Hiwatari and H. Waki, Visualized sound retrieval and categorization, advances in computer Cybernetics. Vol. V, ISAVIIA-98, 68-73, 1998. [4] K. Hirata, Y. Hara, N. Shibata, F. Hirabayashi, Media-based Navigation for Hypermedia system, *ACM Hypertext*: 157-173, 1993. [5] 平田、原、概略画像を用いた画像検索、電子情報通信学会、1992. [6] K. Hirata, Y. Hara et al. Content-oriented integration in hypermedia system, *ACM Hypertext'96 March*, 1996. [7] 伏木田、樋渡、最適矩形波近似による代表ビデオフレーム選択、情報処理学会第 57 回全国大会 [8] 鳥脇純一郎、認識工学、テレビジョン学会