画像ー音インデックスによる動画像検索

6Aa－3

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

動画像のディジタル化技術の普及にともない膨大な量のビデオデータベースの管理と、効率的なデータベースの検索技術が必要になってきている。ここでは、利用者の見たいシーンや、ハイライトシーンを検索するために、ビデオから画像と音を抽出し、これらを複合インデックスとして用いた動画像の検索方式について述べる。

2．ビデオシーン検索実験システム

図1に示すように、検索実験システムは、ビデオデータベースを持つＶＯＤサーバ、ビデオから抽出したインデックスファイルとイメージ検索エンジンを搭載[1]、類似画像検索を行う検索サーバ、そして検索要求を出すクライアントで構成される。これらは、ＡＴＭの高速ＬＡＮで接続されている[2,3]。

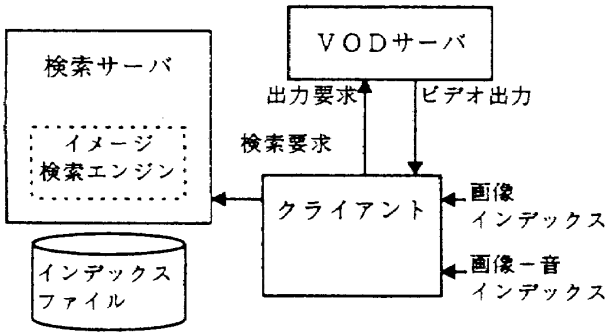

図1　ビデオシーン検索実験システムの概要

クライアントからの検索要求で検索サーバがビデオシーンの代表フレームを格納したインデックスファイルの中から類似画像検索し、その結果を候補画像として利用者に提示、その中から目的の画像を選び、そのインデックスに対応するビデオシーンをＶＯＤサーバから出力する[4]。 今回のビデオシーン検索実験では、ビデオシーンから画像と音をマニュアルで抽出、検索サーバ上にインデックスファイルを構成、これをクライアントから検索した。

3．マルチモーダルインデックス

我々は、画像、音、書誌的情報などで構成するインデックスをマルチモーダルインデックスと呼んでいる。 ここでは、ビデオから画像と音を抽出し、これらを一体化した画像ー音インデックスもマルチモーダルインデックスの1つとした。

（1）音の視覚化

実験システムでは鳥の鳴き声とサッカーゲームの音をカラーのサウンドスペクトログラムで表し音を視覚化している。 図2におおるりの鳴き声のサウンドスペクトログラムを示す。 音のエネルギーレベルを、色に対応させ、赤が高いエネルギーレベルを表し、青になるに従い低いレベルになっていく。 図2の例では3ＫＨｚ～5ＫＨｚの帯域で、高いエネルギーレベルの声（図中の黒表示）の特徴がよく現れている。

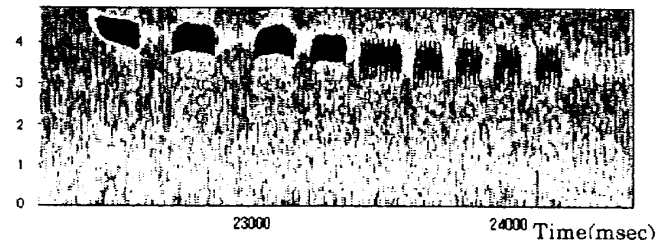

図2　サウンドスペクトログラムの例

（2）インデックシング

ビデオシーンから抽出したフレームを画像インデックスとし、そのフレームにに対応して抽出した音は、視覚化し音インデックスとした。

図3に示すように画像インデックスと、音インデックスとを一体化することで、2種類のビデオコンテンツによる、イメージ検索エンジンを使った類似画像検索ができ、データベース中のインデックスをブラウジングできる。 また、利用者にとっては視覚化した音の内容理解に多少の学習時間を必要とするが、画像との組み合わせにより感覚的に理解しやすくなっている。 ここでは、画像インデックス、音インデックスを5：1の面積比で一体化してビデオシーン検索に利用した。 面積比で画像と音の評価値に対する比重が変わるが、画像インデックスによる検索に比べ検索率が高く、インデックス全体が見やすいことで、この面積比のマルチモーダルインデックスを使った。

Video retrieval method using image-sound index / Yoshitsugu Hiwatari, Katsunobu Fushikida and †Hideyo Waki / Tokyo Waterfront Research Center, TAO Japan
†Tokyo Denki University

図３　マルチモーダルインデックスの例

４．マルチモーダルインデックスによる検索実験

（１）鳥のビデオシーンの検索

鳥のビデオシーン検索実験では、サッカー、鳥類のビデオから抽出した７６種のフルカラーのインデックスを持つデータベースを用いた。　実験はキーとなる鳥のインデックスを決め、これをもとに検索を行い、複数の同一種の鳥のインデックスのリコールランクの平均を求めた。

①イメージインデックスによる検索実験

表１の Image は画像インデックスによる検索の結果である。　鳥や動物のビデオでは、望遠撮影が頻繁で、背景の変化があるため、画像の特徴だけで特定のシーン検索を行うことは難しく、#1,#3,#5,#7 のリコールランクが低い結果になった。

②マルチモーダルインデックスによる検索

この検索では、音インデックス部に、キーに用いたシーンの鳴き声と異なるシーンの鳴き声を用いて、同一種インデックスを構成した。

表１の Multi modal がその実験結果で、画像インデックスによる検索に比べて全体の平均リコールランクは良く、特に#2,#3,#8 のリコールランクの改善が著しい。　これは、画像部分の望遠効果によるフレーム間の差異を音部分で補完し、検索精度を向上できたものと考えられる。

表１　平均リコールランクの比較

Key pic.	Image	Multi modal	同一種インデックス数
#1	35	22	3
#2	12	1	2
#3	20	5	5
#4	16	10	5
#5	38	28	8
#6	17	16	8
#7	32	13	5
#8	24	2	3
Average	23	11	－

数値はリコールランクの平均、データベースサイズ＝７６

（２）サッカーのハイライトシーンの検索

次に、サッカーゲーム３試合のビデオから、フィールド上のパスシーンと、シュートシーンの４４種の代表フレームを、画像と音を対応させて抽出し、マルチモーダルインデックスを構成した。　これを使って、ゲームのハイライトであるシュートシーンを検索する実験を行った。　データベースは４４種のうち、シュートシーンのインデックスは１４種、パスシーンなどは３０種である。

図４の左に示すように、シュートの場面では、サウンドスペクトログラムは高いエネルギーレベルを示し、音インデックス部は赤い色（黒の表示）がほぼ全域を占める。　これに対し、図４右のパスの場面では、音インデックス部は黄色と緑色が多い（濃淡の薄い表示）分布となる。　この特徴をとらえて、芝の緑と赤をペイントし、スケッチ入力で検索を行った。

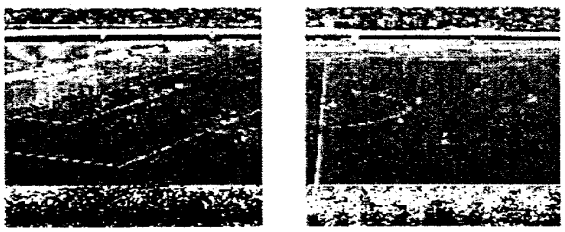

図４　サッカーのマルチモーダルインデックスの例

表２に、この検索で上位に現れたシーンの内容を示す。　リコールランク１０位以内に全てシュートシーンのインデックスがランクされた。これは、シュートシーンのインデックス１４種の７１％にあたり、また２０位以内には９３％がランクされた。

表２　上位リコールランクのシーン

リコールランク	1	2	3	4	5	6	7	8	9	10
シーン	S	S	S	BS	BS	G	G	S	S	G

S; シュート，　BS; シュート直前, G; ゴール直後
データベースサイズ=44

５．まとめ

マルチモーダルインデックスを用いて、ある特定のシーンを検索する実験を行い、検索精度に於いてその有効性を確認した。　ここでは、マニュアルで画像フレームと音を抽出しインデックシングを行ったが、今後は、インデックスの自動作成を行い、効率的な検索についての研究開発に取り組む計画である。

参考文献

[1]平田、原，　概略画像を用いた画像検索，電子情報通信学会,1992　[2]伏木田、樋渡，ｲﾝﾄﾗﾈｯﾄを利用した対話型分散画像検索システム，電子情報通信学会総合大会'97　[3]樋渡、伏木田，ｲﾝﾄﾗﾈｯﾄを利用した類似画像検索システムのアーキテクチャ，情報処理学会第５５回全国大会　[4]K. Fushikida, Y. Hiwatari, H. Waki, Content-based Image Query Method using Parallel Retrieval Scheme, ICCIMA '98 Feb. 1998