

映像データベースの内容記述へのコンテンツ情報の応用

4 Q-6

波多野 賢治[†] 亀井 俊之[†] 田島 敬史^{††} 田中 克己[†][†]神戸大学大学院自然科学研究科^{††}神戸大学工学部

1 はじめに

我々はこれまでに画像のコンテンツ情報のみを用いた自己組織化マップ(以下、SOM)[2]による自動分類を行ってきた[1]。従来までの結果を見てみると精度はそれほど高くないが、ある程度の自動分類が可能であることが判明している。一般にコンテンツ情報のみを用いると、動画像の内容の情報が欠落することから、コンテンツ情報のみを用いたクラスタリングには限界があることが判明している。よって更なるクラスタリング精度の向上のためには映像に対する記述情報が必要になってくるものと考えられる。

このような背景から、本研究ではSOMを用いた動画像データのコンテンツ情報をもとにしたクラスタリング結果を映像の内容記述を行なう際の補助的な手段として用いることにし、実際に動画像データのカットに対し記述を行い、その記述情報と動画像のコンテンツ情報を両方取り入れたハイブリッド型の特徴ベクトルを生成し、それをもとにSOMによる学習およびクラスタリングを行なった。

2 カットに対するハイブリッドシステムの適用

カットとは、コンテンツ間の変化が小さい連続したフレームの集合であるが、このカットに対するハイブリッド型特徴ベクトルの生成のために以下のような処理を行なう。

2.1 ハイブリッド型特徴ベクトルの生成

1. コンテンツ情報をもとにした、カットのクラスタリング

カットのコンテンツ情報特徴ベクトルの生成方法として従来より用いてきたフレーム特徴ベク

トルの重みつき平均値を用いる方式を採用した。
[1]

2. キーワード付与による内容記述

コンテンツ情報による分類結果を示すマップ中の同じセルに配置されたカットには同じ内容の記述を行ない、すべてのカットに対して人間が記述を行なうという手間を省くものである。

3. 記述情報のみを含む特徴ベクトルの生成

記述されたキーワードにより、まず、映像の内容の記述情報を含んだ特徴ベクトルを生成する。第*i*カットの記述情報特徴ベクトル $V_a(i)$ を式で表すと、次のようになる。

$$V_a(i) = w(k_1^i, k_2^i, \dots, k_n^i), k_j^i \in \{0, 1\}$$

ただし、全てのカットへの記述の際に用いたキーワード集合を $\{K_1, K_2, \dots, K_n\}$ 、 k_j^i は第*i*カットに対しキーワード K_j 与えられているとき1をとるとする。 w はコンテンツ情報に対する内容記述への重みを表す。ここで w はコンテンツのDC成分の大きさに合わせ100,200とした。

4. コンテンツ情報と記述情報を統合させる

コンテンツ特徴ベクトル、記述情報特徴ベクトルをそれぞれ

$$V_c(i) = (c_1^i, c_2^i, \dots, c_m^i)$$

$$V_a(i) = (wk_1^i, wk_2^i, \dots, wk_n^i)$$

とすると、ハイブリッド型の特徴ベクトル $V_h(i)$ は、次のようになる。

$$V_h(i) = (c_1^i, c_2^i, \dots, c_m^i, wk_1^i, wk_2^i, \dots, wk_n^i)$$

2.2 実行例

前述した方法により生成されたハイブリッド型特徴ベクトルを入力として、SOMを学習させ、その結果をVRMLによって出力する。実際に、15分のアニメーションをカット分割し、15×15のマップを用い

A Video Authoring Tool utilizing Video Feature Information
Kenji Hatano[†], Toshiyuki Kamei[†], Keishi Tajima^{††},
Katsumi Tanaka[†]
[†]Graduate School of Science and Technology, Kobe University
^{††}Faculty of Engineering, Kobe University

てマップ生成を行なった。この時、学習回数は12,000回、データとなったカットの総数は125個で、ユニット数225個のマップは約1分で生成された。分類結果を図1に示す。

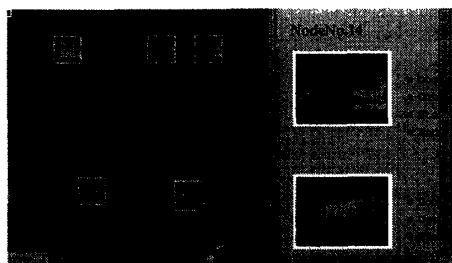


図1: ハイブリッド型特徴ベクトルを用いたビデオカットに対するクラスタリング結果 ©東映株式会社

2.3 評価

実際、行なった実験がどれほどの結果なのかを判断するために「精度(適合率)」と「再現率」を測定した。図2では上から素材データを変えた場合、アニメーション映像のコンテンツ情報のみによる分類、アニメーション映像のコンテンツと記述情報による分類でコンテンツを評価基準に取り入れた場合、同じ実験に対し、コンテンツと記述情報を評価基準に取り入れた場合、重みが100,200の時のコンテンツ、記述情報を用いた分類の精度と適合率を、それぞれ測定範囲が中心のセルから近傍距離が1,2の場合について示す。

評価基準	特徴ベクトル	コンテンツ情報(実写映像)		コンテンツ情報(アニメ映像)	
		近傍距離	適合率	再現率	適合率
コンテンツのみ	1	53.71	43.52	73.35	45.08
	2	52.17	54.73	33.75	61.93

評価基準	特徴ベクトル	コンテンツ情報		コンテンツ情報、記述情報	
		近傍距離	適合率	再現率	適合率
コンテンツのみ	1	73.75	45.08	64.59	47.67
	2	33.75	61.93	37.95	60.69

評価基準	特徴ベクトル	コンテンツ情報		コンテンツ情報、記述情報	
		近傍距離	適合率	再現率	適合率
コンテンツと記述情報	1	61.22	41.74	65.03	47.67
	2	39.98	52.36	37.95	60.69

評価基準	特徴ベクトル	コンテンツ、記述(重み=100)		コンテンツ、記述(重み=200)	
		近傍距離	適合率	再現率	適合率
コンテンツと記述情報	1	61.22	41.74	67.82	48.88
	2	39.98	52.36	40.47	65.70

図2: 適合率および再現率

像であるため、映像に用いられている色、構図などがほぼ決まっていることから、記述の前段階でも普通のテレビ映像を用いた場合に比べかなり高いクラスタリング精度が得られているまた、映像のコンテンツのみを類似度の評価基準とした場合にハイブリッド型特徴ベクトルによる分類結果の適合率が下がったが、これは記述情報が学習時に反映された影響であると考えられる。一方、コンテンツ、記述情報とともに類似度の評価基準に取り入れた場合、両情報を取り入れた分類の方が全体的に適合率、再現率は多少ではあるが良い結果となった。

3 おわりに

本研究では、SOMを用いて動画データに対する内容記述をカットに対して行ない、その記述情報および動画データのコンテンツ情報を共に用いたカットのクラスタリングを行うことにより精度の向上が認められた。

また、コンテンツ情報のみを用いたカットのクラスタリング結果を動画データに対する内容記述に利用することで記述作業を軽減化させ、ある程度意味内容による分類ができるようになった。今後の課題として

- コンテンツ情報と、内容記述情報の定式化された重みの付け方
- シーンの自動分割の方法の提案
- 複数の記述者がいる場合の記述情報の統合化

などを挙げておく。

謝辞

本研究において、貴重な映像資料の学術利用を許可して下さった、東映株式会社に感謝致します。また、本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」および文部省科学研究費重点領域研究(課題番号08244103)による。ここに記して誠意を表します。

参考文献

[1] K.Hatano, Q.Qing and K.Tanaka. *A SOM-Based Information Organizer for Text and Video Data*. Proceedings of the Fifth International Conference on Database Systems for Advanced Applications (DASFAA'97), pages 205-214, Apr. 1997.

[2] T.Kohonen. *The Self-Organizing Map*, Proc. of the IEEE, Vol.78, No.9, pp.1464-1480, 1990