

映像データベースの動的クラスタリングと 素材検索機構について

波多野 賢治[†]

田中 克己[‡]

[†]神戸大学大学院自然科学研究科情報知能工学専攻

[‡]神戸大学大学院自然科学研究科知能科学専攻

{hatano,tanaka}@in4wolf.in.kobe-u.ac.jp

本稿では動画像データを、効果的にクラスタリングし類似検索を行なうシステムを構築する際に、従来のように映像の内容に対するキーワード付与や内容記述を行わず、画像を符合化することでその映像の特徴をとらえる方法について述べる。Kohonenの自己組織化マップの機能を用いて映像を動的・自己組織的にクラスタリングし、3次元表示や類似検索を行なえる試作システムとその実験結果について報告する。このシステムは、未整理の映像データ素材の再利用のための強力なツールとなり得るものと考えられる。

Dynamic Clustering and Movie Asset Retrieval for Video Databases

Kenji Hatano[†]

Katsumi Tanaka[‡]

[†]Division of Computer and Systems Engineering,
Graduate School of Science and Technology, Kobe University

[‡]Division of Intelligence Science,
Graduate School of Science and Technology, Kobe University

We propose a framework for effective clustering and similarity-based retrieval of image video data. Instead of giving keywords or authoring them, we use image coding in order to extract characteristics of image data. Coded image data are clustered by Kohonen's self-organizing map, and the result is visualized in a 3D form. By this, similarity-based retrieval is achieved. We implement a prototype system and report experimental results. We consider that our system effectively promote reuse of disordered image data assets.

1 まえがき

近年、文書、画像、音声などのマルチメディアデータを計算機上で扱うことに注目が集まってきており、それらのデータをデジタル化およびそのアーカイブ化の技術が進展してきている。それに伴いデータを複数の利用者に対して提供するだけでなく、ユーザの希望に応じた柔軟な処理や問い合わせなどの双方向通信を行なうなどのアプリケーションの開発が容易となっている。しかしその一方で未整理のデータが散在しその再利用のためにはそれらデータの組織化および検索技術の需要が高まってきている。

本研究はキーワード付与や内容記述を行なわない形で、World Wide Web(WWW)上で最近盛んになっているマルチメディアデータ、今回は特に映像のカットを素材として、効果的にクラスタリングし類似検索を行なうことによって、映像データベースの構築を手助けすることを目的としている。

我々は、従来より自己組織化マップ (Self Organizing Map, 以下 SOM) と呼ばれるニューラルネットワークを用いて、文書群の動的な分類、分類地図の段階的詳細化や曖昧検索が可能なブラウジング検索システムを開発してきている [1][2][3][4][5]。また、同様な研究も他で始められており、津田ら [6] はマップの階層化の方法とその初期値の設定法について提案している。

自己組織化マップの根本的な考え方 [7] は、与えられたデータ集合に対して各データの特徴ベクトル (パターン) を自己組織化マップに繰り返し学習させることにより、データ集合に隠されている類似構造を発見しそれを2次元自己組織化マップ上に反映し表示するというものである。本研究ではニュース映像を題材として、各カット素材 (asset) を再利用の単位と見なして自己組織化マップを用いたクラスタリングを試みた。

本稿では主に次の事項に関して述べる。

- 映像カットの特徴ベクトルの生成方法
- その特徴ベクトルを用いた3次元自己組織化マップの生成方法と検索機構
- ニュース映像を題材とした実験と特徴ベクトル生成方式の評価

2 基本的事項

2.1 自己組織化マップ

ニューラルネットワークの一種である SOM は、1990年に T.Kohonen によって提案された教師なし競合学習モデルである [7]。出力層の各ユニットが層の中で位置を持つという点が他の学習モデルと異なる点であ

る。このモデルの特徴はデータに隠されているトポロジカルな構造を学習アルゴリズムにより発見し、通常2次元空間で表示するというものである。

SOM法で用いられるネットワークは、ユニットを2次元上に配置したものである。入力データは通常高次元の特徴 (feature) ベクトル \mathbf{x} にパターン化され、ネットワーク中の各ユニット i は入力パターン \mathbf{x} と同次元のベクトル \mathbf{m}_i をもっている。学習はこれらのユニットを入力パターンに選択的に近付けることによって進行する。SOM法では入力パターンに一番近いパターンを持つ出力ユニット c およびその近傍 (図1) のユニットの集合 N_c のみを入力パターンに近付ける学習アルゴリズムを採用している。また、統計的に正確な学習効果を得るため、一定の学習回数 T をとる。

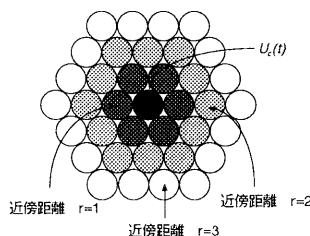


図1: SOMの出力層とユニットの近傍の取り方

この結果、入力ベクトル空間で近くにあるものは繰り返し学習の結果、各入力データはそのパターンと距離的に最も近いパターンを持つユニットに射影される。この結果入力ベクトル空間で近くにあるものはSOM上でも互いに近傍のユニットへと射影されるような写像が完成する。

2.2 DCT(Discrete Cosine Transform)

DCT(Discrete Cosine Transform、離散コサイン変換)とは、JPEG(Joint Photographic Expert Group)静止画像圧縮技術で用いられている画像の変換符合化方式である [8][9]。

1枚の静止画像を $N \times N$ 画素の正方形の領域 (ブロック) に分割し、各ブロックに対して変換処理を行うと、領域内の平均的な画像 (領域全体が一樣) に始まり、徐々に精細さを表現する画像へと段階的な画像に分解することができる。この分解操作を直交変換といい、精細さが高いことを別のいい方では、周波数が高いという。静止画像は、第1低周波項 (平均値画像) から順に、高周波項へと分解した画像の重ね合わせの表現になる。

DCTの利点は、変換前にランダムに分布していた画素値 (輝度など) が、変換後には低周波項に大きな

値が集中する性質があることである。したがって、高周波項を落とす操作(量子化)をすれば情報圧縮を行うことができる。

1枚の画像から分割された画素ブロックの大きさが $N \times N$ 画素のとき、画素信号を $f(x, y)$ 、変換によって得られる係数(DCT係数)を $F(u, v)$ とすると、 $F(u, v)$ は次のように求まる。

$$F(u, v) = \frac{2}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} C(u)C(v)f(x, y) \cdot \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}$$

ただし、

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u, v = 0 \\ 1 & \text{otherwise} \end{cases}$$

このようにして得られた DCT 係数のうち $F(0, 0)$ を DC(Direct Current、直流)係数といい、それ以外の DCT 係数を AC(Alternate Current、交流)係数と呼ぶ。DCT 係数は、画素数と同じ $N \times N$ 個求まり、低周波成分に集中する。DC 係数はブロック内の画素値の平均値を表し、AC 係数はその周波数の活性化を示す。

3 動画画像から 3次元自己組織化マップの生成

動画画像情報に対し自己組織化マップを利用した自動分類を行なうには、以下の課題を克服する必要がある。

- 入力データの単位決定
時系列データである動画画像データを再利用可能な素材として活用できる形に断片化する必要がある。本稿では、カットをこのための単位とする。従って、1つのニュース記事(article)は通常複数のカットに断片化される。
- カット(動画画像)からの特徴ベクトルの生成
静止画フレームの時系列データであるカットに対する特徴ベクトルをいかに生成するかという問題であり、本稿では各静止画フレームの DCT 表現を集約させる形で特徴ベクトルを生成する方式を提案する。
- カット(動画画像)の 3次元自己組織化マップの生成と表示
SOM を用いたカットの分類結果を基に素材検索やブラウジングを効果的に進めるための表示方式の開発が必要である。本稿では VRML(Virtual Reality Modeling Language)[10]を用いた 3次元ブラウザを開発した。

3.1 カットの特徴ベクトルの生成

3.1.1 DCT 成分の抽出によるフレーム特徴ベクトル生成

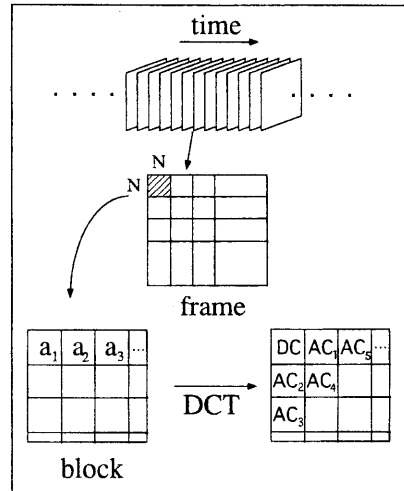


図 2: 静止画フレームの DCT 表現

動画を構成している各フレーム画像をいくつかのブロックに分割し、そのブロックを前述の 2次元 DCT により周波数成分に変換する。変換された成分は図 2 のように左上を直流成分とし、右下に向かって周波数が上がっていくように成分が並ぶ。このうちの第 1 成分(直流)DC、第 2、3 成分(水平、垂直周波数成分、交流)AC₁、AC₂の 3つの成分を用いる。これから 1つのフレームに対する DCT 成分を要素とするベクトル

$$(DC(1), AC_1(1), AC_2(1), DC(2), AC_1(2), AC_2(2), \dots, DC(n), AC_1(n), AC_2(n))$$

を構成し、これをフレーム特徴ベクトルと呼ぶ。ここで $DC(k), AC_1(k), AC_2(k) (k = 1, 2, \dots, n)$ は分割されたブロックにおける DC 成分と AC 成分である。

3.1.2 カット検出

カット検出には、有木ら[11]のカットの変わり目のフレーム(カット点フレーム)を検出する手法を用いている。3.1.1 節に述べた方式で生成したフレーム特徴ベクトルは多次元空間の 1点として表される。同じカット内の隣接したフレームのフレーム特徴ベクトルは、この多次元空間において近いところに存在し、クラスタを形成すると考えられる。ところがカットが変わると、多次元空間中の点はこのクラスタから逸脱しはじめるので、この逸脱を持ってカット点検出を行なうことができる。

3.1.3 カット特徴ベクトル生成

カットの特徴ベクトルを生成するには、ベクトルの次元と要素を何にするかが問題となる。文書の場合は含まれる単語の総数から次元数を決定し、要素はその単語の文書中の出現頻度に比例した値が用いられたが、動画像には、単語のような意味をなすキーワードが存在しない。そこで本研究では、カットの特徴ベクトルの生成方法として、カット検出で分割された各カットに含まれる全てのフレームの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。この方法では、DCTを行なうと低い周波数にその画像のパワーが集中するという特徴を利用することにより、情報が膨大な動画像の特徴をできるだけ低次元のベクトルに反映させることができる。以下にカットの特徴ベクトル生成のアルゴリズムを示す。

1. 動画像を構成する各フレームを $n = L \times M$ 個のブロックに分割し、各ブロックに対し3つのDCT成分 (DC, AC_1, AC_2) を抽出することにより、フレーム全体で $3 \times L \times M$ 個のDCT成分を得、これからフレーム特徴ベクトルを生成する。なお、フレーム特徴ベクトルを生成する際に3つのDCT成分に重み w_1, w_2, w_3 をつけ、抽出するDCT成分を $(DC \times w_1, AC_1 \times w_2, AC_2 \times w_3)$ にすることで重みをつけている。
2. カット検出により抽出された1つのカットに含まれる全てのフレームの特徴ベクトルの重みつき平均を求め、これをカット特徴ベクトルとする。このときのフレームの範囲は、カット点フレームから次のカット検出時のアウトフレームが生じる直前のフレームまでとする(図3参照)。これによりカットに含まれる全てのフレームの特徴を集約できる。

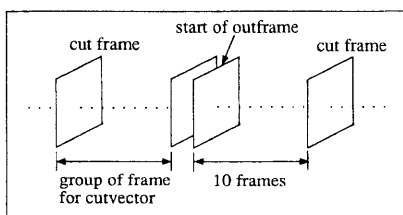


図3: カットとカット特徴ベクトルの計算

4 システムの実現

4.1 システムの全体構成

このシステムは図4に示すように、次の5つの部分からなる。

● カット検出部分

この部分は龍谷大学の有木研究室で開発されたカット検出・シーン切り出しソフト [11] を用いた。このソフトはIndy上に実装されているCosmo Compressorと呼ばれる画像圧縮・伸張ハードウェア上で実現される。

● カット特徴ベクトル生成部分

この部分はカット検出プログラムに組み込まれており、カット検出と同時にベクトルが生成される。

● 自己組織化マップ (SOM パッケージ) による学習部分

得られたカット特徴ベクトル群を入力データとして自己組織化マップに学習させる。

ここでは、フィンランドのT.Kohonenが率いるプログラム開発チームにより開発されたSOMパッケージ [12] を用いた。

● VRML による3次元マップの生成

3次元マップとユーザーインターフェース部分はVRMLを用いて実現されている。

● 類似カット一覧表示部

検索のためHTMLを用いてドキュメントリスト一覧表示用のリストファイルと、ユニット情報表示用のカットファイルを作成した。

4.2 実行例

実際にニュース映像(10分×3、56392フレーム)を用いて、以下の処理を行なった。

1. カットのマップ生成

3.1節の方式により約200のカットが検出され、それぞれ144次元のベクトルに変換された。これに 15×15 のマップを用いてマップ生成を行なった。自己組織化マップの学習回数は11,000回とした。マップ生成には10分を要した。

(a) 3次元マップ全体像

図5はシーンを入力としたマップの実行結果がViewer上に現れる。図において円筒の高さはそれらに写像されたカットの数を表しており、またそれぞれに貼りつけられたカットは円筒に分類されたカットの中でその円筒の持つベクトルの値に最も近いカットが表示されている。

(b) 分類結果

図5上で任意の円筒(ノード)を選択すると、図6のようにそのユニットに写像されたシーンに含まれる各カットのカット点におけるイメージを含んだテキストがWWWブラウザ上に表示される。

(c) リスト一覧表示

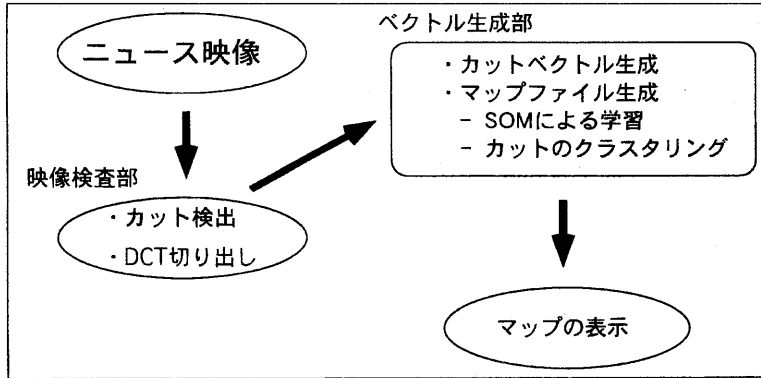


図 4: システム構成



図 5: 3次元自己組織化マップ全体図

任意の数字(ノード)を選択すると、マップ上の全てのユニットに写像されたカット点におけるイメージデータがWWWブラウザ上に現れる。

本システムでは、ベクトル化されたカットをSOM法で学習させた後、その学習結果を3次元マップに表示させ、さらに3次元マップ上で行なえる点検索機能や一覧リスト表示機能をもたせた。

これらの諸機能により、データの分布状況が容易に把握でき、また曖昧な検索にも対応できることから、情報閲覧・検索システムとしての有効性が確かめられたといえる。

5 システムの評価

カットの分布状況を一別すると、マップ全体にデータが分散したという感を受ける。マップ上で集中して

いる部分を見てみると、雰囲気や明暗、背景、場面状況などといった視覚的に似通ったカットが近い所に配置されており、中でも動作やカメラワークが比較的静かなカットはある程度同じところにまとまりやすいといえる。

しかし、従来の方法だとカットの明暗でマップ上に分散しているのははっきり分かるため、カットベクトルの要素として用いたDCT成分のうち画像の濃淡の変化に左右されるという直流成分の特徴が強くていゝものと推察される。このようなことを避けるためには、ベクトル生成時に直流成分を省くか重みづけを行なうといった作業を行なった訳である。

そしてそれぞれについて、「精度(適合率)」と「再現率」を測定した。ここでいう「精度」と「再現率」は、検索されなかった適合情報をA、検索された適合

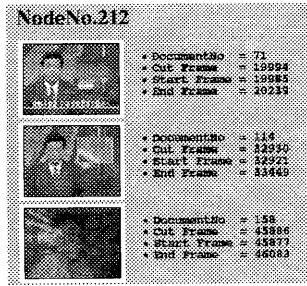


図 6: ユニット検索

情報を B 、検索された不適合情報を C とした場合、適合率は $B/(B+C)$ 、再現率は $B/(A+B)$ で表される。今回の実験ではあるユニットと近傍距離 1 のユニットを検索範囲とした。

w_1	w_2	w_3	精度 (%)	再現率 (%)
1	1	1	44.45	36.08
1/DC	1	1	46.99	50.50
1/DC	10	20	-	-
1/DC	20	10	-	-
1	10	20	53.71	43.52
1	20	10	52.66	47.93
1	10	100	40.22	51.53

これらを見ると精度、再現率ともに 40~60%前後と検索結果としてはそう良いものではない。しかし、似ているカットの分類方法や、分類した結果似ているだろうと思われ処理されたカットが全体のだいたい半数(その他は関係のないカットとして処理)であることを考えれば、この結果は十分に納得のいく結果だろうと思われる。また、以上の結果からカットベクトルの生成には $w_1 < w_2 < w_3$ という関係が満たされていればよいことが経験的に判明した。ただ、どれくらいの大関係が必要かなどの詳細はもう少し実験を行なう必要があると考えられる

以上より視覚的には一応の成果が見られるが、画像を符合化するという点でカットの内容に対する意味的なものをとらえられたかどうかは疑問であるといえる。ただ、今回の実験で用いたニュース映像ではカットのジャンルが限られているために正確な判断はできない。今後はスポーツ映像などを用いているようなカットに対して実験を行なっていく予定である。

6 あとがき

本研究では、自己組織化マップをデータベースのブラウジングツールおよび問い合わせインタフェースとみなして動画データを自動分類し、3次元自己組織

化マップの形で表示するという機能を実現した。

試作したシステムの利点として次のものが挙げられる。

- マップを 3 次元で表現することによる見通しの良さ
- 曖昧な検索への柔軟な対応
- VRML 利用による portability の確保

また、検討すべき課題として以下のものがあげられる。

- カットベクトルの要素に対するパラメータ設定法の検討
- 自己組織化マップの学習に関する各種パラメータの設定方法の検討
- より見やすいユーザインタフェースの開発
- カット分類の成果の article 分類への適用

謝辞 なお本研究は一部文部省科学研究費重点領域研究(課題番号 08244103)による。

参考文献

- [1] 仁木 和久、田中 克己、「ニューラルネットワーク技術の情報検索への応用」、人工知能学会誌, Vol.10 No.1, pp.1-7, January 1995
- [2] Qing,Q., Shi,X., and Tanaka,K., *Document Browsing and Retrieving based on 3D Self-Organizing Map*, Proc. of Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with CIKM'95, Baltimore, USA, Nov.28-Dec.2 1995
- [3] Qing,Q., *Study on Hypertext Database Systems and Incremental Data Organization Mechanisms*, Doctoral Dissertation, Graduate School of Science and Technology, Kobe University, July, 1995
- [4] Qing,Q., Hatano,K., Sasao,T., Tanaka,K., *A VRML-Based Information Organizer by 3D Self Organizing Map*, 電子情報通信学会第 7 回データ工学ワークショップ (DEWS '96), pp.85-90, March 1996
- [5] 笹尾 年男, 「動画画像情報の自己組織化に関する研究」, 神戸大学工学部計測工学科卒業論文, March 1996
- [6] 津田 宏治, 黒田 崇, 美濃 導彦, 上林 彌彦, 「文書検索のための自己組織化セマンティックマップの階層化」, 情報処理学会第 52 回全国大会, No.4, pp.217-218, March??, 1996
- [7] Kohonen,T., *The Self-Organizing Map*, Proc. of the IEEE, Vol.78, No.9, pp.1464-1480, 1990
- [8] 加藤 茂夫, 「画像データ圧縮の基礎知識」, インターフェース, No.175, pp.132-159, December 1991
- [9] 藤原 洋, 「最新 MPEG 教科書」, (株)アスキー, 1994
- [10] <http://vrml.wired.com/>
- [11] 岩成 英一, 有木 康雄, 「DCT 成分を用いた動画シーンのクラスタリングとカット検出」, 電子情報通信学会, パターン認識と理解研究会, PRU93-119, pp.23-30, 1994
- [12] SOM Programming Team of the Helsinki University, *Self-Organizing Map Program Package Version 3.1*, April 1995