

感性に基づく動画検索について

井田俊博 寶珍輝尚[†] 都司達夫[‡] 樋口健[‡]

福井大学大学院 工学研究科 情報工学専攻

[†]大阪府立大学 総合科学部 数理・情報科学科

[‡]福井大学 工学部 情報・メディア工学科

〒910-8507 福井市文京3丁目9-1

ida@pear.fuis.fukui-u.ac.jp

本論文では、感性の主因子を用いた動画検索システムに生じる新たな動画データに対応できない問題を解消する方法について述べる。感性の主因子は Semantic Differential 法と因子分析によって求められる。感性の主因子を用いた動画検索システムは因子得点を使用しているため、想定された動画データにしか適用できないという問題が生じる。この問題を解消するために、つまり、想定されていない動画データにも対応した感性にも基づく動画検索システムを実装するために、動画データの特徴量と感性の主因子との関係を導き出す必要がある。そのために、本論文ではどのような特徴量を使用するかについて考察する。

On the Video Retrieval Based on Kansei

Toshihiro IDA Teruhisa HOCHIN[†] Tatsuo TSUJI[‡] Ken HIGUCHI[‡]

Graduate School of Fukui University

[†]Osaka Prefecture University

[‡]Deps. Of Information Science, Faculty of Eng., Fukui University

3-9-1, Bunkyo, Fukui-shi, Fukui 901-8507 Japan

ida@pear.fuis.fukui-u.ac.jp

This paper describes how to solve the problem which cannot treat new video data in the Video Retrieval System Based on Kansei by using the human sensitivity factors. These factors are obtained by using the Semantic Differential Method and the Factor Analysis. The Video Retrieval System can only treat the video data examined in the Semantic Differential Method because this system uses factor scores of these video data. In order to solve this problem, the correlation between the feature values of video data and the main factors of sensitivity must be obtained. This paper studies what kinds of feature values of video data are used in the Video Retrieval System.

1. はじめに

近年、スケッチ画像をキーとする類似画検索、内容を表現する一種のネットワークを用いる内容検索、特徴量空間における所望の範囲を指定する検索、感性語を用いた検索などのマルチメディアデータに対する検索の研究が盛んに行なわれている。しかし、現在の

マルチメディア検索は、一般ユーザーが簡単に扱える適切な検索方法がないため、操作が複雑になり、初心者が容易に利用できるまでには至っていない。

そこで本研究では、マルチメディア検索方法の中で利用者が理解しやすく利用しやすい『感性語を用いた検索』を試みる。しかし、マルチメディアデータを扱

う場合、そのメディアの種別や対象に依存した特徴量を用いており、多種多様なデータを扱う場合好ましくない。

このような問題を解決する手法として『人間の感性の主因子』を利用する方法が提案されている。この手法では、対象となるデータから『人間の感性の主因子』の因子得点を抽出し検索に利用している。ここで、人間の感性の主因子は、SD法と呼ばれる心理学で使用される方法でマルチメディアデータの印象を定量化し、これを因子分析して得られるものである。現在までの研究で、画像データ、自然音データ、音楽データ、動画データを対象とした感性に基づくマルチメディアデータ検索システムが実現されている。また、画像データ、自然音データ、ならびに、音楽データに対しては、これらの特徴量からそれぞれの因子得点を推定する方法が検討されている。従って、画像データ、自然音データ、ならびに、音楽データに対しては、任意のデータを検索の対象とすることができる。

しかし、これまでの感性に基づく動画検索では因子分析で得られた因子得点を使用しており、想定したモデル(動画データ)にしか適用できない。つまり、新たな動画データが加えられたとき、それに対する因子得点を使用することはできない。これを解決するためには、(1)新たなデータを加えて再度評価してもらい因子分析を行う、(2)現在の因子得点とデータの特徴量の関係から推定するという二つの方法が考えられる。再度評価してもらう方法は非常に手間がかかり、非常実的である。

そこで、本研究では特徴量から因子得点を推定し、新たな動画データにも対応した感性に基づく動画検索システムを実現することを目的とする。特に本論文では、MPEG形式の動画を対象とし、動画データの特徴量としてどのような特徴量を使用するかについて考察する。

以降、2. では、動画の感性の主因子とこれを用いた感性語による動画検索システムについて述べる。次に、3. で、MPEGについて概説する。そして、4. で、使用する特徴量について考察する。最後に、5. でまとめる。

2. 動画の感性の主因子

2.1 感性の主因子

心理学の研究では、印象語が表す感性的な性質はいくつかの基本的な因子(感性の主因子)によって表現

されることが明らかとなっている。これは Semantic Differential(SD)法という手法を用いて解析した結果得られるものである。SD法とは心理学者 C. E. Osgood が考案した手法で、複数の反対の意味を持つ印象語の対を尺度とし、その間をいくつかの段階に分けてある対象物を被験者に評価させるものである。図1は「父」についてどのような印象を受けるかを評価した例である。この例では、7段階で評価している。例えば、happy-sad では、いくぶん happy であり、hard-soft では、かなり hard であるといった具合である。

```

happy  : __ : __ : X : __ : __ : __ : __ : sad
hard   : X : __ : __ : __ : __ : __ : __ : soft
slow   : __ : __ : __ : X : __ : __ : __ : fast
  
```

図1 : SD法の例

これらの印象語対がそれぞれ単次元の特性を抽出すると仮定し、因子分析を適用すると、次元の少ない、より簡潔なものとしてすることができる。因子分析では、 n 個の観測対象に対する p 変量のデータの行列を X とすると、 X をした式のように表現する。

$$X = FA' + E$$

ここで、 F は $n \times m$ の行列、 A は $p \times m$ の行列 A の転置行列、 E は $n \times p$ の行列であり、 m をできるだけ小さくとり、かつ、 E を十分小さくなるように分解し、 F と A を求める。 F は因子得点行列、 A は因子負荷量行列、 E は残差行列と呼ばれる。変数の数 p よりもかなり小さい m を用いることで潜在的な因子を求める。

ここで、 E を十分小さくするにはいくつかの方法がある。 E の分散・共分散行列を $U = E'E$ とすると、 U の各要素の2乗和を最小にする方法は主成分分析法と呼ばれ、 U の非対角要素の2乗和を最小にする方法は Minres 法と呼ばれる。

また、因子負荷行列 A は一意に決定できるものではなく、自由度がある。通常は、ある変量は絶対値が大きく、他の変量は絶対値が小さくなるような回転を施し、説明をしやすくするのが一般的である。よく利用される方法にバリマックス法がある。

また、因子得点行列 F も一意に決定できるものではない。よく知られている方法には以下の方法がある。

$$F_1 = XR^{-1}A \quad (1)$$

$$F_2 = XWA(A'WA)^{-1} \quad (2)$$

ここで、 W は重み行列である。この因子分析を SD 法によって求めた得点に適用すると、ものや概念に内在する因子が得られる。これを感性の主因子と呼ぶ。

2.2 動画の感性の主因子

動画に対して、SD 法により評価し因子分析した。被験者が SD 法により、2, 1, 0, -1, -2 の 5 段階で評価し、このデータを各動画、各印象語対ごとに平均を求め、Minres 法により因子分析を行って得ている。ここで、使用した形容詞対の数は 16 (図 2) である。また、因子は、1 より大きい固有地を持つ因子を採用することにして得ている。さらに、因子負荷行列は、バリマックス回転を行って得ている。

対象の動画は、Moonpocket から販売されている「ビデオ素材時点」から 4 2 個、おのおの、約 10 秒前後である。被験者は男子大学生 10 名である。

さらに、感性の主因子の意味を説明する変数を、因子負荷量行列の因子負荷量の大きい印象語対がその因子を表しているという従来の考え方、ならびに、他の因子と比較して特に重みの重い印象語対を選択するようにして決定している。この方法により得られた印象語対を表 1 に示す。

2.3 動画検索システム

本研究の検索システムでは、因子分析の結果得られた因子負荷量、指定された感性語の重み、ならびに因子分析から推定された因子得点によって検索を行う。ここで、あらかじめ因子分析により因子負荷量を抽出し、全ての動画データにおいて因子得点を推定しておく。検索は以下のように行う。

1. ユーザーは www 上より検索に用いる印象語を 1 つ以上選択する (図 2)。但し、相反する形容詞 (例: 「速い」と「遅い」) は同時に選択できない。
2. 選択した形容詞の重みを決定する。
3. 選択された形容詞の因子負荷量・重みと動画データが持つ因子得点の積をとり、検索対象となっている全てのデータごとに和を求め評価値とする。ここで、重みをかけることにより重み付けを行い、微妙なニュアンスのデータの検索に対応している。
4. 評価値の大きいものから 10 個を検索結果として出力する (図 3)。

表 1 : 感性の主因子

主因子	形容詞対
明快性	美しい - 醜い 明るい - 暗い 愉快的な - 不愉快的な
力量性	動的な - 性的な 大胆な - 繊細な 迫力のある - 迫力のない
活動性	規則的な - 不規則的な 複雑な - 単純な
軽重性	軽快な - 重厚な
堅鋭性	直線的 - 曲線的



図 2 : 検索実行画面

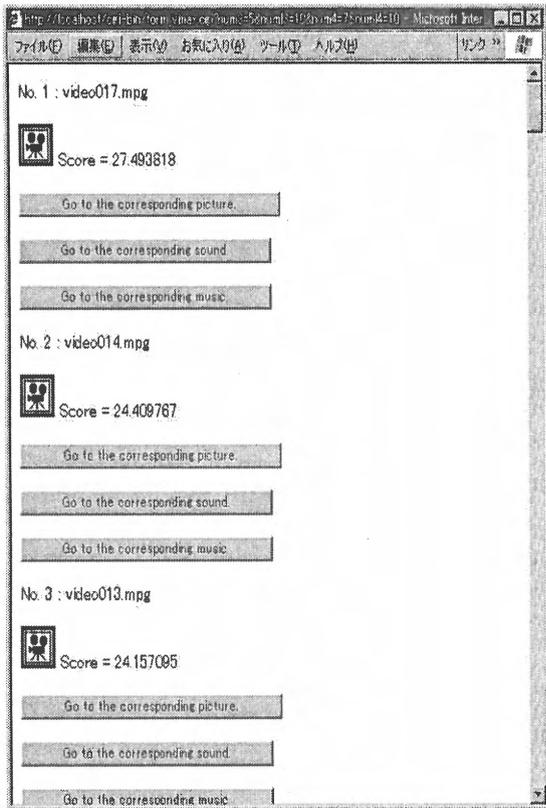


図3：検索結果画面

3. MPEG(Moving Picture Experts Group)

3.1 動画データ

動画データは画像データの並びであり、画像サイズと画像レートで表される。画像データは画素値の2次元配列である。画素値は通常8bitのデータであり、0から255の値をとる(255が最も明るい画素を表し、0が黒を表す)。

画像が白黒の場合はこのような輝度画像のみで十分であるが、カラー画像はRGBの3つの値または3つの画像で表される。これを線形変換して、輝度(Y)と2つの色差(Cb,Cr)で表すと、色差画像の空間分解能を減らすことができる。

3.2 データ構造

画像データは図4に示すように、シーケンス、グループ・オブ・ピクチャ(GOP)、ピクチャ、スライス、マクロブロック(MB)、ブロックの6層の階層構造から成っている。

GOP内の画面(Picture)はタイプをもつ。主要なピクチャタイプは、つぎの3種である。まず、Iピクチャはフレーム内(Intra frame)符号化画像である。この画像では、すべてのマクロブロックがIntra符号化される。つぎに、Pピクチャはフレーム間予測(Inter

frame prediction)符号化画像である。この画像は、マクロブロックごとにIntra符号化とInter符号化が選択できる画面タイプである。Inter符号化とは、入力画像と、以前の再生画像との差を符号化することである。この差(ずれ)は動きベクトルと呼ばれる。最後にBピクチャは双方向予測符号化画像である。この画像は、過去のIまたはPピクチャを予測に使うだけでなく未来のI,Pピクチャをも予測に使うことができる画像である。

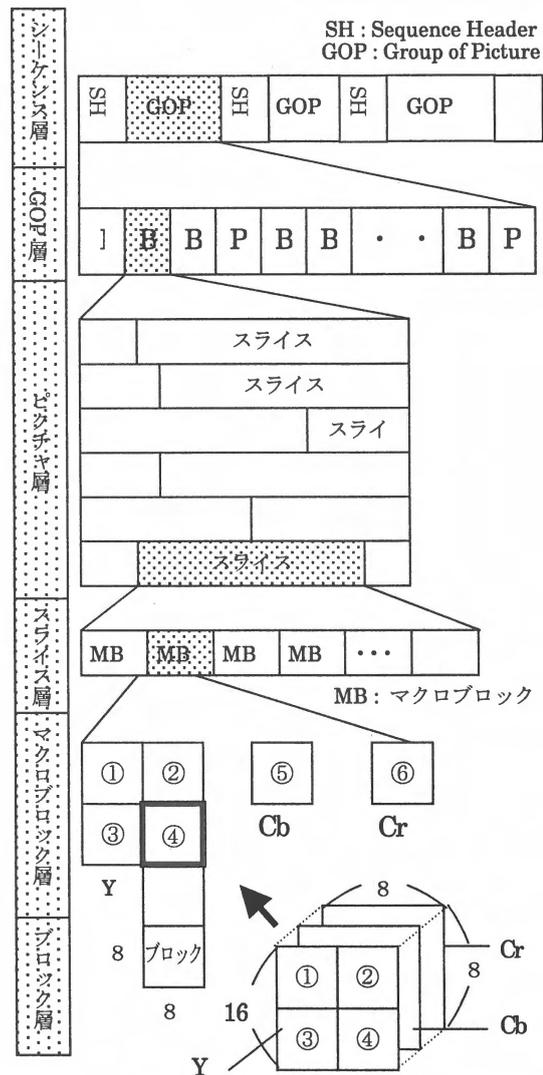


図4 MPEG1の階層構造

3.3 DCT (離散コサイン変換) 符号化の流れ

入力画像は、まず8×8画素のブロックに分割される。このブロック単位にDCT演算を行い、得られたDCT係数をDC成分、AC成分で独立して量子化する。量子化に用いる量子化テーブルには規定値は存在

せず、個別に設定される。一般には、図に示した量子化テーブル値が多く用いられる。量子化した DCT 係数のうち、DC 係数は、直前のブロックの DC 係数を予測値とした差分値を符号化する。残りの AC 成分は、ブロック内でジグザグ・スキャン (図) によって並び替えた後、符号化する。

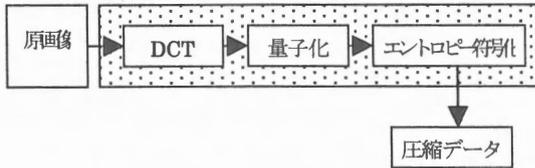


図5 : DCT 符号化方式の符号器

表2 : 量子化テーブル

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	34	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

表3 : ジグザグ・スキャン順表

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

4. 特徴量について

4.1 動画ファイルについて

検索対象に用意した動画像データのファイルタイプは QuickTime 形式であったため、WWW 上で容易に閲覧できるように、Macintosh 上のフリーの MPEG 変換ツールと、Panasonic の MPEG 変換プ

ラグインをインストールした Adobe Premiere で MPEG 1 ファイルに変換した。

動画の例を図6と図7に示す。図6は木星の動画で、円弧を描いて木星が近づいてくる動画である。図7は風景面の動画で、風で枝が微妙に動く動画である。



図6 : 動画1 (木星接近)



図7 : 動画2 (風にそよぐ樹木)

4.2 使用する特徴量

どのような特徴量を使用する必要があるのかを検討するにあたり、表1に見られる5つの感性の主因子を考慮する。まず、明快性では、美しい-醜い、明るい-暗い、愉快的-不愉快なという形容詞対が見られるため、画像の色に関する特徴量が必要と考えられる。そこで、明度、輝度を使用することとした。次に、力量性、活動性、軽重性、堅鋭性では、動画本来の動きに関する特徴がみられる。そこで、動きベクトルから特徴が得られると考えた。また、動画データには始まりと終わりで雰囲気が変わる場合もあるため、全体、ならびに、全体を三つに分けた開始部分、中央部分、終了部分の合計4パターンに対して特徴量を得ることとする。ここで、明度、輝度に関する特徴量はIピ

クチャから獲得し、動きベクトルに関する特徴量はPピクチャから獲得する。

また、特徴量として各特徴量の平均と標準偏差を求めることにする。ここで、動きベクトルは全てのマクロブロックに対して存在するとは限らないので、存在するもののみで平均した場合と全マクロブロックで平均したものを求めることとした。さらに、前後画像の差も求めることとする。

使用する特徴量は、明度、前後画像の明度の差、輝度、前後画面の輝度の差、全マクロブロックを対象とした画像内動きベクトルの大きさ、その前後画像の差、Inter符号化マクロブロックのみを対象とした画像内動きベクトルの大きさ、ならびに、その前後画像の差のおおの平均と標準偏差である。これを、全体、ならびに、全体を三つに分けた開始部分、中央部分、終了部分の合計4パターンに対して得るので、合計64の特徴量を得ることになる。

動画1の特徴量を表4、6、8、10に示す。また、動画2の特徴量を表5、7、9、11に示す。

表4：動画1の全体の特徴量

特徴量	平均	標準偏差
明度	27.37	21.20
前後画像の明度の差	9.73	14.25
輝度	52.54	35.88
前後画面の輝度の差	17.19	25.11
画像内動きベクトル (全MBを対象)	9.74	8.73
前後画像の差	3.30	3.84
画像内動きベクトル (Inter符号化MBを対象)	22.22	9.65
前後画像の差	6.18	5.65

表5：動画2の全体の特徴量

特徴量	平均	標準偏差
明度	33.73	8.47
前後画像の明度の差	2.53	7.52
輝度	76.74	19.29
前後画面の輝度の差	5.97	17.06
画像内動きベクトル (全MBを対象)	1.04	0.76
前後画像の差	0.13	0.18
画像内動きベクトル (Inter符号化MBを対照)	1.64	0.70
前後画像の差	0.14	17.06

表6：動画1の開始部分の特徴量

特徴量	平均	標準偏差
明度	11.56	2.82
前後画像の明度の差	1.87	0.85
輝度	26.16	4.91
前後画面の輝度の差	3.20	1.48
画像内動きベクトル (全MBを対象)	3.00	2.73
前後画像の差	-0.34	13.94
画像内動きベクトル (Inter符号化MBを対象)	19.11	9.16
前後画像の差	-0.86	24.57

表7：動画2の開始部分の特徴量

特徴量	平均	標準偏差
明度	34.40	0.12
前後画像の明度の差	-0.03	0.20
輝度	78.36	0.46
前後画面の輝度の差	0.18	0.58
画像内動きベクトル (全MBを対象)	0.34	0.22
前後画像の差	-0.50	3.76
画像内動きベクトル (Inter符号化MBを対照)	1.14	0.25
前後画像の差	-1.12	8.57

表8：動画1の中央部分の特徴量

特徴量	平均	標準偏差
明度	24.44	8.80
前後画像の明度の差	8.04	5.60
輝度	48.54	15.26
前後画面の輝度の差	12.20	8.95
画像内動きベクトル (全MBを対象)	7.37	4.01
前後画像の差	0.00	0.00
画像内動きベクトル (Inter符号化MBを対象)	18.13	8.54
前後画像の差	0.00	0.00

表9：動画2の中央部分の特徴量

特徴量	平均	標準偏差
明度	33.53	0.62
前後画像の明度の差	-0.12	0.70
輝度	76.20	1.76
前後画面の輝度の差	-0.50	1.66
画像内動きベクトル (全MBを対象)	1.35	0.71
前後画像の差	0.00	0.12
画像内動きベクトル (Inter符号化MBを対照)	1.83	0.72
前後画像の差	0.02	0.25

表10：動画1の終了部分の特徴量

特徴量	平均	標準偏差
明度	40.63	28.78
前後画像の明度の差	-15.76	63.33
輝度	72.41	49.97
前後画面の輝度の差	-26.26	61.70
画像内動きベクトル (全MBを対象)	18.96	8.46
前後画像の差	0.00	3.81
画像内動きベクトル (Inter符号化MBを対象)	29.62	6.43
前後画像の差	0.00	2.66

表11：動画2の終了部分の特徴量

特徴量	平均	標準偏差
明度	26.55	13.31
前後画像の明度の差	-6.74	39.02
輝度	60.41	30.26
前後画面の輝度の差	-15.20	35.07
画像内動きベクトル (全MBを対象)	1.44	0.66
前後画像の差	0.02	0.51
画像内動きベクトル (Inter符号化MBを対照)	1.95	0.70
前後画像の差	0.01	0.44

4.3 考察

今回の試みで使用した特徴量には、違いがはっきり

とみられ、適当であると考えられる。しかし、似通ったもの、単純なものが多く、多種の特徴量をもっと増やしていく必要があると考えられる。これにより、より複雑な動画データに対しても十分対応できると考えられる。また、データセットに関しても、より複雑な(途中で内容が多様に移り変わる)動画データの使用も必要と考えられる。

5. おわりに

感性に基づく動画検索システムにおいて、新たな動画データにも対応可能とするための動画データの特徴量について検討した。この結果、明度、輝度、ならびに、動きベクトルを利用することとした。

しかし、今回は使用する特徴量の提案に留まり、特徴量の評価が行われていない。また、特徴量からの因子得点の推定も行っていない。今後は、推定した特徴量の評価を行い、本来の目的である新たな動画にも対応する感性に基づく動画検索システムの実現を行う予定である。

参考文献

- [1] 寶珍 輝尚、高田 伸介、都司 達夫：感性に基づくマルチメディアデータの相互アクセスについて、情報処理学会研究報告 DBS124-3 FI62-3、2001、44、17-24 (2001)。
- [2] T. Hochin、T. Tsuji：Mutual Multimedia Access using Kansei Factors、*Kansei Engineering international*、Vol. 2、No. 4、pp. 9-18 (2001)。
- [3] 田中 豊、脇本 和昌：“多変量統計解析法”、現代数学社出版 (1984)。
- [4] 井口 征士：“感性情報処理”、オーム出版 (1994)。
- [5] 超 智弘、黒田 英夫：“図解でわかる画像圧縮技術”、日本実業出版 (2000)。
- [6] 藤原 洋：“最新MPEG教科書”、アスキー出版 (1994)。
- [7] TECHI (Vol. 4)；“画像&音声圧縮技術のすべて”、CQ出版 (2000)。
- [8] スティーブ・リマー：“Windowsマルチメディアプログラミング”、アスキー出版 (1995)。

付録

使用した動画データの一覧を示す。

- (1) 光玉

- (2) 光
- (3) 光の波紋
- (4) オーロラ
- (5) 業火
- (6) 暗雲
- (7) 地球 (回転)
- (8) 木星 (回転)
- (9) 木星 (接近)
- (10) 小惑星軍
- (11) 銀河 (離反)
- (12) フィルム
- (13) 壁
- (14) フュージョン
- (15) ノイズ
- (16) 水
- (17) 三角群
- (18) 梅
- (19) ラベンダー
- (20) 花畑
- (21) 菜の花
- (22) 風にそよぐ樹木
- (23) 竹林
- (24) 吹雪
- (25) 吹雪 (夜)
- (26) 夕日
- (27) 木々にさしこむ光
- (28) ツリー
- (29) 車
- (30) 東京タワー
- (31) ダンス
- (32) 晴天
- (33) 不気味な雲
- (34) 夕日2
- (35) 山脈
- (36) 雲に突進
- (37) 浅瀬
- (38) 波
- (39) せせらぎ
- (40) 滝
- (41) 泡
- (42) 水