

映像特徴空間に基づく同系映像の分類

服部しのぶ[†] 亀山 渉^{††} 富永 英義^{†,††}

[†] 早稲田大学大学院 理工学研究科 電子・情報通信学専攻 〒169-8555 東京都新宿区大久保3-4-1

^{††} 早稲田大学大学院 国際情報通信研究科 〒169-0051 東京都新宿区西早稲田1-3-10

E-mail: †sally@tom.comm.waseda.ac.jp, ††{wataru,tominaga}@waseda.jp

あらまし 近年のマルチメディアコンテンツの大量化により、効率的なコンテンツの記述・管理・検索技術の必要性が高まりつつある。これを踏まえ、メタデータによるコンテンツ管理・流通に注目が集められている。しかし、現在、これらのメタデータは人手を介して付加されているため、人間の主観に依存した情報となり、必ずしも普遍的な情報とは限らない。そこで、本稿では、映像から自動抽出可能な特徴量(カメラパラメータ、色情報など)を用いた映像解析による同系映像の分類手法を提案する。これにより、従来の人間の主観による映像分類(ジャンルなど)とは異なった、映像特徴量という普遍的な情報を用いた映像の記述・管理・検索が可能となる。また、自動抽出可能な特徴量を用いることで、メタデータ付加作業における人的コストが削減され、効率的な映像アーカイブの管理が期待される。また、従来のジャンルによる映像検索ではなく、特徴量を基準とした新しい映像検索サービスへの応用なども可能となる。提案手法として、複数の特徴量を軸とする特徴空間を構築するという新しい概念を導入し、この特徴空間上で近距離に存在する映像を同系映像と定義する。

キーワード 映像アーカイブ, 映像分類, 特徴空間, 同系映像

Video Classification Based On Video Description Space

Shinobu HATTORI[†], Wataru KAMEYAMA^{††}, and Hideyoshi TOMINAGA^{†,††}

[†] Grad. School of Elec., Info. and Comm. Eng., WASEDA Univ., 3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169-8555 Japan

^{††} GITS, WASEDA University, 29-7 building 1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo 169-0051 Japan

E-mail: †sally@tom.comm.waseda.ac.jp, ††{wataru,tominaga}@waseda.jp

Abstract The advances in storage and communication technologies have made vast amounts of video contents available to consumer. This has developed techniques using metadata to describe, organize, and manage these contents. However, we still have limitations in creating metadata common to all users, automatically. Techniques for annotating and retrieving video content, such as content analysis and indexing therefore have become an active research topic. In this paper, we propose a video classification method using description space composed of low-level description, such as camera works and color information, automatically extracted from video. We determine the similarity of video based on description space. This method leads to an efficient automatic video archive management or a new application such as user preference usage for video retrieval.

Key words Video Archives, Video Classification, Description Space, Similar Video

1. はじめに

近年のマルチメディア技術の発展に伴った、映像・画像など、様々なコンテンツの大量化により、効率的なコンテンツの記述・管理・検索技術の必要性が高まりつつある。

現在、これらの有効的な技術として“メタデータ”が大きく期待されている。例えば、現在の検索に用いられているキーワードも、メタデータの1つとして考えられる。しかし、これらのメタデータは人手によって付加されているため、必ずしも普遍的な情報とは限らない。例えば、現在、一般的に広く用いられている検索ツールでは、検索したいコンテンツに対する語をユーザが入力する。その語と、登録されているキーワードが一致したコンテンツのみ提供される。しかし、これらのキーワードは、必ずしもユーザの意図を反映したものではないため、目的のコンテンツが取得できない場合が想定される。このように、人手によって付加された情報は、全てにおいて共通していないことから、普遍的な情報として扱うことは難しい。

このような現状を踏まえ、コンテンツから得られる特徴量を用いたコンテンツ解析やインデキシングを検索技術に応用する研究が盛んに行われている。

例えば、映像は、“映画の文法”[1]と呼ばれる、ある共通したカメラ操作や編集操作の規則に基づき、視聴者に作者の意図した印象が伝わるように作成されている。この“映画の文法”を用いて、映像の規則性を映像解析の特徴量として取り入れることにより、映像が与える印象を抽出する手法などが従来より提案されており、その有効性が報告されている[2]。

そこで本稿では、映像間の関連性を、各映像が有する特徴量(カメラパラメータ、色情報など)を用いて定量的に解析することにより、同系映像を分類する手法を提案する。

本提案手法の実現により、自動的に映像を整理し、大量化する映像アーカイブの効率的な管理が可能となる。これにより、ユーザの希望するコンテンツへのアクセスを容易とし、効率的な映像検索技術への応用が期待できる。また、映像から得られる特徴量を用いることにより、メタデータ付加作業における人的コストの削減も可能となる。

本稿では、2.において従来の映像解析技術についてまとめる。また、3.では、本提案手法について述べる。また、4.では、提案手法の評価実験について、5.では、まとめと今後の課題を示す。

2. 映像解析技術

現在、映像解析技術として、コンテンツ解析やインデキシングに関する様々な研究が盛んに行なわれている。これらの技術には、特徴抽出、構造解析、要約、記述処理など、主に4つの処理段階が存在し、様々な手法が検討されている。

また、映像解析技術には、手動・半自動的・全自動的ななど、様々な手法が存在する。しかし、手動による手法は、人手を介しているため、時間的コストにおける問題点が存在する。半自動的・全自動的による手法には、コンテンツの意味的情報を学習的に関連付けする手法や、統計学的モデルを用いた手法などが存在する。しかし、これらの手法は、独自の特徴量に依存する、特徴量を人手によって入力する、対象映像が限定されているなど、様々な問題点が存在する。

Aya[2]らは、映像の信号の特徴量から意味的特徴を定義すること目的とし、上記で述べた“映画の文法”の照明規則に着目し、映像の印象を抽出している。映像から、輝度を特徴量として抽出し、それらの変化や条件などの特徴が人の印象へ与える影響について検討している。

Cheng[3]らは、大量化する映像アーカイブの効率的な検索を目的とし、独自の特徴量である“映像署名(Video Signature)”を用いて、“IVS(Ideal Video Similarity)”と呼ばれる2つの映像間の類似性を計測している。“映像署名”とは、映像の特徴を基準とした映像の複数のフレームのセットのことである。ここで用いている特徴量は、映像の視覚的情報であるHSVである。

また、Smith[4]らは、マルチメディアの意味的インデキシングを用いた検索を目的とし、特徴量のモデルベクトル化による画像分類を行なっている。モデルベクトル化とは、画像の色・テキスト・動き情報を基に定義した“辞典用語”を用いて、学習的に作成されたベクトル空間に画像を位置付けすることである。

また、Chong-Wah[5]らは、類似映像ショットの分類による効率的な映像検索への応用を目的とし、映像のショットを段階的なクラスタリングを用いて分類する手法を提案している。特徴量としては、色情報と動き情報を用いている。まず、映像の色情報を基に類似ショットを分類し、その結果をさらに細かく動き情報を基に分類する段階的な分類処理を行なっている。

3. 提案手法

本稿では、映像が有する特徴量を利用した映像解析手法として、複数の特徴量を軸とした特徴空間に基づく同系映像の分類手法を提案する。

映像が有する特徴量を利用することは、全ての映像に共通した特徴量を用いることになるため、独自の特徴量への依存性をなくすことが可能となる。

また、映像の特徴は、映像に対するキーワードと同様に、複数の特徴量の組合せや、重なり合いによって定義・表現されると考えられる。

例えば、映画のジャンルについて考えた場合、1つの映画を1つのジャンルのみで表すことは難しい。ある映画は、ホラーでもあり、サスペンスでもあり、またドラマでもあると考えられる場合が想定される。このように、映像を“ジャンル”という1つの特徴量によって特徴付けることは困難であることが分かる。

また、同じ映画でも、個人により、ホラーやドラマなど印象が異なる。つまり、ジャンルなど、人手を介した情報は、必ずしも全員に共通する情報とは限らず、普遍的でない。よって、それを基準とした映像分類は、必ずしも正しい分類とは限らない。

そこで本稿では、従来の人手による情報を基準とするのではなく、映像から得られる普遍的な特徴量を基準とする映像分類手法を提案する。

また、“色情報のみ”や“動き情報のみ”など、1つの特徴量のみで解析するのではなく、複数の特徴量を同時に用いて解析を提案する。

複数の特徴量から作成される特徴空間を基に映像間の関連性を解析し、分類を行なう。本稿では、この特徴空間において、関連性の高い、近距離に存在する映像を、“同系映像”と定義する。

映像間の関連性を、映像特徴量を用いて抽出することにより、例えば、新しい映像検索システムへの応用が可能となる。現在のアーカイブにおいて同系と判定されている映像の基準は、人間のジャンルなどの主観評価が大半である。

しかし、映像特徴量を基準とすることにより、人間の感覚とは異なった映像の関連性が抽出される可能性が考えられる。

これにより、従来の検索システムにおいては、普段接点のない分野における映像の発見へと導くと考えられ、新しい映像検索の応用技術となる。

3.1 特徴量

本稿で用いる映像特徴量について、以下にまとめる。

3.1.1 カット点情報

以下に示すカット点情報から得られる特徴量を用いることとする。カット点検出手法には、MPEG 映像において、映像のカット点前後のマクロブロックタイプの特徴に着目したカット点検出手法を用いる。ここで、マクロブロックタイプの特徴とは、P Picture ではイントラ符号量が増大し、B Picture では双方向予測が減少、又は順方向予測が逆方向予測が増大することである。

- Cut Extraction Rate (C_R)

Cut Extraction Rate とは、カット点の検出率であり、以下の通りに定義する。

$$C_R = \frac{C_n}{V_T} * 100 \quad [\%] \quad (1)$$

ここで、 C_n は合計カット数、 V_T は映像の長さを示す。

- Average Cut Time (C_{AT})

Average Cut Time とは、平均カット時間であり、以下の通りに定義する。

$$C_{AT} = \frac{1}{C_n} \sum_{t=1}^{C_n} C_t \quad [\%] \quad (2)$$

ここで、 C_T は、各カット長のことを示す。

3.1.2 カメラパラメータ

以下に示したカメラパラメータから得られる特徴量 [6] を用いることとする。抽出されるカメラパラメータ (C_χ) は、以下に示した Fix, Pan, Zoom, Shake の 4 種類である。Shake とは映像がぶれていると判定された場合のカメラパラメータである。

$$\chi \in (\text{Fix}, \text{Pan}, \text{Zoom}, \text{Shake}) \quad (3)$$

カメラパラメータ抽出手法には、映像から動きベクトルを抽出し、その方向と量をもとに 8 方向にクラスタリングし、そのヒストグラムの時間的変化によりカメラパラメータを判定する手法 [7] を用いる。以下に、それぞれの特徴量についてまとめる。

- Camerawork Extraction Rate (CR_χ)

Camerawork Extraction Rate とは、カメラワーク検出率のことであり、以下の通りに定義する。

$$CR_\chi = \frac{N_\chi}{N_T} * 100 \quad [\%] \quad (4)$$

ここで、 N_χ は各カメラワークの検出数、 N_T は合計カメラワーク検出数を示す。

- Camerawork Unit (CU_χ)

Camerawork Unit とは、カメラワーク区間のことである。本稿では、1GOP 単位 (15 フレーム毎) にカメラワークを検出するため、連続して検出される同カメラワーク区間を結合し、1つのカメラワーク区間として定義し、その区間時間を $t(CU_\chi)$ と定義する。

- Camerawork Transition (CT_{χ_1, χ_2})

Camerawork Transition とは、カメラワークの状態遷移のことである。カメラワークの状態遷移とは、カメラワーク間の移り変わりを定義したものであり、例えば、Fix から Pan へ切り替わった場合、これを Fix-Pan Camerawork Transition ($CT_{\text{Fix}, \text{Pan}}$) と定義する。 CT_{χ_1, χ_2} は以下の通りに定義する。

$$CT_{\chi_1, \chi_2} = \frac{N_{\chi_1, \chi_2}}{N_{CT}} * 100 \quad (\chi_1 \neq \chi_2) \quad [\%] \quad (5)$$

ここで、 N_{χ_1, χ_2} は C_{χ_1} と C_{χ_2} 間のカメラワークの移り変わった回数、 N_{CT} は CT の合計数を示す。

3.1.3 色情報

映像から得られる色情報には、様々な値が存在する。本稿では、カラーイメージとして公認されている色相・明度・彩度 (HSV) の値を用いることとする。以下に示した特徴量を HSV から得られる特徴量として用いる。

- 各フレーム毎の平均 HSV 値とその分散値
- 各フレーム毎の Dominant HSV 値とその分散値

ここで、Dominant HSV とは、MPEG-7 [8] でも規定されている "Dominant Color" のことであり、各フレームにおける主要な色のことを示す。

3.2 特徴解析

本稿で用いる特徴解析手法について以下にまとめる。本稿では、複数の特徴量を用いた特徴空間の作成を目的としている。しかし、映像が有する全ての特徴量を軸として、特徴空間を作成した場合、次元数が非常に多くなり複雑な特徴空間になることが予想される。従って、解析が困難になると考えられる。

そこで本稿では、特徴解析手法として、複数からなる特徴空間の次元数を重要な特徴量に重みを与えることにより削減する手法を提案する。次元数を削減し、かつ重要な特徴量に重みを与えた分類を行なうことにより、最適な分類を、より簡単な計算を用いて得ることが期待できる。

多変量から構成される空間軸の削減手法として、主成分分析が存在する。主成分分析とは、多種類の変数の情報を、互いに無相関な少数個の総合特性値に要約する分析手法のことである。以下に、主成分分析手法について簡単にまとめる。

主成分分析

主成分分析の目的は、複数の要因が絡んでいるデータから、総合的な評価式を作成すること、つまり、次元を減らすことにより、もとのデータの持つ意味を理解しやすくする手法である。総合的な評価式とは、用いたデータの特徴付ける特徴量のことを意味する。この評価式で、最もデータを差別化する評価式を表す座標軸を "第 1 主成分" と呼ぶ。n 個のデータから、最大 n 個の主成分が出力される可能性がある。

変量 (x_1, x_2, \dots, x_p) に対して、任意の係数 (a_1, \dots, a_p) を用いて、以下に示す線形結合 (1 次式による合成変用) が得られる。

$$z = a_1 x_1 + \dots + a_p x_p = a^t x \quad (6)$$

$$a = (a_1, \dots, a_p)^t, x = (x_1, \dots, x_p)^t \quad (7)$$

ここで、 z は合成変量、任意の係数 (a_1, \dots, a_p) を固有ベクトルと呼ぶ。これは、各変数の主成分に対する影響度を表している。

また、合成変量 z の分散を最大にする、最大固有値に対応する固有ベクトル $a_1 = (a_{11}, \dots, a_{p1})^t$ の要素を係数として表した式 (8) が "第 1 主成分" となる。

$$z_1 = a_1^t x_1 = (a_{11} x_1 + \dots + a_{p1} x_p) \quad (8)$$

本稿で用いる変数を a_i と定義した場合、変数 i は 3.1 章で示した特徴量のことであり、以下の通りになる。

$$i \in (CR_\chi, CU_\chi, t(CU_\chi), CT_{\chi_1, \chi_2}, \text{etc.}) \quad (9)$$

また、主成分の分散を λ_i 、もとの変量の分散を s_{ii} とした場合、求められた主成分の分散と、もとの変量の分散の間には、以下のような関係が存在する。

$$\lambda_1 + \lambda_2 + \dots + \lambda_p = s_{11} + s_{22} + \dots + s_{pp} \quad (10)$$

ここで、 $\lambda_k / \sum_{j=1}^p \lambda_j$ を、第 k 主成分の寄与率、 $\sum_{j=1}^k \lambda_j / \sum_{j=1}^p \lambda_j$ を、第 1-第 k 主成分までの累積寄与率と呼ぶ。

これらは、各主成分が、どの程度特徴を表しているかを示す尺度である。寄与率は、採用する主成分数を決める基準となり、寄与率の最も高いものから順に第1主成分として抽出される。

主成分数決定法

主成分分析を適用する場合、主成分をいくつまで採用するかが問題となる。もとの変量の持っている情報の大部分が、主成分で説明されている必要があるが、現象を単純化して理解するという観点からは、できるだけ少数個であることが望ましい。主成分数の決定は、一般的にスクリープロット(岩くず図)と呼ばれる固有値をグラフで示したものをを用いて、以下の2つの基準によって行なわれる。スクリープロットは、主成分が、どの程度全体を説明しているかを示した折れ線グラフである。

- スクリー基準

固有値の減少が、他に比べて大きくなった時点

- カイザー基準

各主成分の固有値 $\lambda_i > 1.0$ となった時点

本稿では、この2つの基準を用いて、主成分数の決定を行ない、実験的に最適な主成分数の決定を行なうこととする。

3.3 映像分類

似ている物同士をまとめる手法(分類手法)として、一般的にクラスター分析が存在する。クラスター分析には、主に階層的クラスター分析・非階層的クラスター分析の2種類が存在する。本稿では、後者の非階層的クラスター分析手法であるK-means法を採用することとする。以下に、本稿で用いるK-means法によるクラスター分析についてまとめる。

K-means法によるクラスター分析

K-Means法はPartitioning Method(分割手法)と呼ばれるクラスター分析の一種で、データを与えられたk個のクラスターに分割する。この方法では、全てのデータに対して、k個のクラスターの中心値とデータとの距離を計算し、最も近いクラスターにデータを配分する。次に、クラスターに配分されたデータから、そのクラスターの中心値を計算する。クラスターの中心値は、そのクラスターに属する全ての点の平均値とする。これを、全てのデータに対して、データとデータが属するクラスターの中心値との距離の合計が最小になるまで繰り返す。なお、本稿では距離として、平方ユークリッド距離を用いている。

n個のデータそれぞれについて、p種類の特性が測られた場合のi番目のデータの特徴を表すベクトルを $(x_{i1}, x_{i2}, \dots, x_{ip})$ とした場合の平方ユークリッド距離は式(11)のように定義される。

$$d_{ij}^2 = \sum_{k=1}^p (X_{ik} - X_{jk})^2 \quad (ij = 1, 2, 3, \dots) \quad (11)$$

クラスター数決定法

非階層的クラスター分析は、集団をある群に分割する手法であるため、分析者が、いくつのクラスターに分類するか、クラスター数を決定する必要がある。そこで、本稿では、以下に定義されるクラスター数自動決定法[5]を用いて最適クラスター数 \hat{k} を決定することとする。

まず、クラスター数評価値 $p(k)$ は以下のように定義される。

$$p(k) = \frac{1}{k} \sum_{i=1}^k \max_{x_i \leq k} \left\{ \frac{\eta_i + \eta_j}{\xi_{ij}} \right\} \quad (12)$$

ここで、kはクラスター数、 η_i, η_j は各クラスター内の平均距離、 ξ_{ij} はクラスターiとjのクラスター間距離を示す。

表1 Experimental Contents Condition

MPEG1 Sequence Factors	Parameters
Frame Rate	30[frame/s]
I/P Frame Distance	3[frames]
Frame Size	352[pixel]*240[line]
Bit Rate	1.8[Mbps]
Encoder	Canopus MPEG Station
Ave Sequence Time	2.00[min]
Total Number of Sequence	50

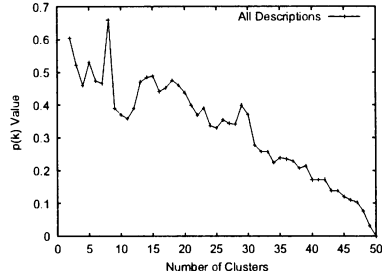


図1 p(k) Value (All Descriptions)

また、最適クラスター数 \hat{k} は以下のように定義される。

$$\hat{k} = \min_{1 \leq k \leq N} p(k) \quad (13)$$

つまり、K-means法を用いたクラスター分析を $k = \{1, 2, 3, \dots, N\}$ まで行ない、最小のクラスター数決定値 $p(k)$ を与えるクラスター数 k が最適クラスター数 \hat{k} となる。

しかし、 $p(k)$ 値は、クラスター数が増加するにつれ、減少することが予想される。その場合、最適クラスター数の決定を、単純に $p(k)$ の最小値とすることは適切でないことが分かる。

そこで、本稿では、局所的な最小値を用いることとする。つまり、 $p(k)$ 値の増加量が、最大となったクラスター数までをクラスターリング範囲(N)とし、 $1 \leq k \leq N$ までの最小 $p(k)$ 値のクラスター数を、最適クラスター数 \hat{k} と定義する。

4. 実験

本提案手法における映像の分類条件として、以下に示した2つの分類条件が考えられる。

- 全特徴量(39種類)を用いた映像の分類
- 主成分分析による主成分を用いた映像の分類

そこで、本稿では、これらの条件における分類実験を行なった。また、各実験条件において、クラスター分析における最適クラスター数の評価実験や主成分分析における最適主成分数の評価実験を行なった。

実験映像は、映画の予告編の映像を対象とした。映像の符合化の条件は、表1に示した通りである。合計50本の映像を対象とした。また、抽出した特徴量は、合計39種類(カット情報:2種類、カメラパラメータ:25種類、色情報:12種類)である。

4.1 全特徴量を用いた映像分類実験

以下に、3.1に示した特徴量を、映像から抽出し、それらを用いた分類実験を行なった結果を表2に示す。次に、分類を行なう際の最適クラスター数の評価実験について以下に示す。

最適クラスター数の評価実験

映像を分類する際に必要となるクラスター数の評価実験を行なった結果を図1に示す。

表 2 Cluster Result (All Descriptions)

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Video 1	Video 41	Video 24	Video 4
Video 45	Video 20	Video 10	Video 23
	Video 7	Video 5	Video 44
	Video 13	Video 21	Video 36
	Video 26	Video 18	Video 37
	Video 16	Video 9	Video 48
	Video 47	Video 43	Video 30
	Video 27	Video 29	Video 2
	Video 22	Video 33	Video 25
	Video 42	Video 14	
	Video 50	Video 17	
	Video 46	Video 32	
	Video 12	Video 3	
	Video 49	Video 34	
	Video 6	Video 39	
	Video 40	Video 38	
	Video 19	Video 35	
	Video 31	Video 11	
	Video 8		
	Video 28		
	Video 15		

表 3 Cluster Result (Screen Criterion)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Video 1	Video 30	Video 16	Video 23	Video 12	Video 47
Video 45	Video 11	Video 22	Video 2	Video 14	
Video 8	Video 38	Video 13	Video 4	Video 34	
	Video 36	Video 26	Video 48	Video 6	
	Video 9	Video 20		Video 40	
	Video 35	Video 17		Video 42	
	Video 21	Video 18		Video 32	
	Video 44	Video 15		Video 41	
	Video 25	Video 28		Video 29	
	Video 37			Video 39	
				Video 49	
				Video 10	
				Video 19	
				Video 7	
				Video 24	
				Video 33	
				Video 43	
				Video 27	
				Video 50	
				Video 46	
				Video 31	
				Video 3	

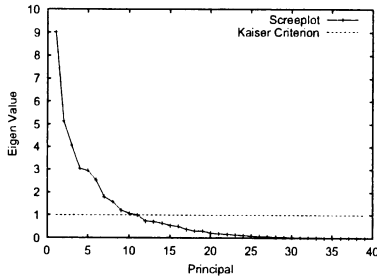


図 2 Scree Plot

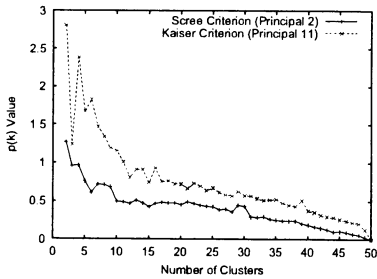


図 3 p(k) Value (Scree & Kaiser Based)

表 4 Cluster Result (Kaiser Criterion)

Cluster 1	Cluster 2	Cluster 3
Video 37	Video 1	Video 4
	Video 2	Video 24
	Video 3	Video 48
	Video 5	
	Video 6	
	Video 7	
	Video 8	
	Video 9	
	Video 10	
	Video 11	
	Video 12	
	Video 13	
	Video 14	
	Video 15	
	Video 16	
	Video 17	
	Video 18	
	Video 19	
	Video 20	
	Video 21	
	Video 22	
	Video 23	
	Video 25	
	Video 26	
	Video 27	
	Video 28	
	Video 29	
	Video 30	
	Video 31	
	Video 32	
	Video 33	
	Video 34	
	Video 35	
	Video 36	
	Video 38	
	Video 39	
	Video 40	
	Video 41	
	Video 42	
	Video 43	
	Video 44	
	Video 45	
	Video 46	
	Video 47	
	Video 49	
	Video 50	

図 1 は、各クラス数におけるクラス数評価値 $p(k)$ を示したものである。図 1 より、 $p(k)$ 値の最大増加は、クラス数が 8 個になる時点であることが分かる。よって、本分類におけるクラスタリング範囲は $N = 8$ となる。

次に、クラス数 8 個までの最小 $p(k)$ 値を求める。グラフより、最小 $p(k)$ 値は、クラス数が 4 個の時となることが分かる。よって、本分類における最適クラス数として、4 個が得られる。

4.2 主成分を用いた映像分類実験

次に、本提案手法である、主成分分析を用いた特徴軸の削減による分類実験についてまとめる。まず、主成分分析によって得られる主成分で、いくつまでの主成分を採用するのかを決定するための最適主成分数の評価実験を行った。その結果をもとに、4.1 と同様に、映像を分類する最適クラス数を決定し、分類を行った。

最適主成分数の評価実験

図 2 に、主成分分析によって得られた主成分とその固有値の関係を表したスクリープロットを示す。図 2 より、スクリー基準では、最も固有値の減少が起る時点は第 2 主成分である。

よって、最適主成分数は第 2 主成分となる。また、カイザー基準の場合、固有値が 1 以下になる時点は、第 11 主成分となる。よって、最適主成分数は第 11 主成分となる。

最適クラス数の評価実験

上記の 2 つの最適主成分数を基準として、最適クラス数の評価実験の結果を図 3 に示す。図 3 より、スクリー基準では、 $p(k)$ 値の最大増加は、クラス数が 7 個になる時点となる。よって、クラスタリング範囲は $N = 7$ となり、最適クラス数は 6 個になる。また、カイザー基準の場合、 $p(k)$ 値の最大増加は、クラス数が 4 個になる時点となる。よって、クラスタリング範囲は $N = 4$ となり、最適クラス数は 3 個となる。これらの結果を用いて、映像の分類を行った結果を表 3 と表 4 に示す。

表5 Cluster Result Genre (Scree Criterion)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Horror, SF, Suspense	Action	SF, Family	Action, SF	Action	SF, Cult
Classic, Cult, SF	Drama	Suspence, Thrillar	SF, Action	Horror	
Action, Comedy	Drama, Romance	Horror	Comedy	Drama, Thrillar	
	Action, Drama	Drama	Drama, Horror, Mystery	Romance, Comedy	
	Romance, Comedy	Drama		Drama	
	Comedy	SF, Family		Drama	
	Drama	SF, Family		Comedy	
	Comedy, Romance	Action		Drama	
	Action	Action		Drama, Romance	
	Drama			Comedy, Drama	
				Drama	
				Action	
				Comedy	
				Classic, Romance	
				Action	
				Classic, Romance	
				Drama, Romance	
				Comedy, Romance	
				Drama	
				Drama, Romance	
				Action	
				Action	
				SF, Action	

4.3 評価

本稿における実験における評価をまとめる。まず、最適主成分数の評価実験における、2つの基準(スクリー基準・カイザー基準)について考察する。

スクリー基準を用いた場合とカイザー基準を用いた場合のクラスタ数評価値 $p(k)$ を比較する。図3より、スクリー基準を用いた場合の $p(k)$ 値の方が、全体的に小さいことが確認できる。つまり、本稿におけるクラスター分析において、スクリー基準に基づいて主成分を決定した場合の方が、良いクラスタ数評価値を得られていることから、分類結果においても良い結果が期待できることが分かった。

また、表3と表4を比較した場合、スクリー基準を用いた分類結果である表3の方が、分類結果に偏りが少なく、分散された比較的良好な分類結果が得られていることが分かる。

次に、それぞれの条件における分類結果について考える。表2-表4より分かるように、これらの分類結果には大きな違いが存在することが分かる。しかし、これらの分類は、分類基準が異なるため、単純に比較することは困難である。

そこで、1つの比較基準として、現在の映画の分類として、最も一般的に用いられているジャンルにおいて比較を行なう。本稿における評価実験より、スクリー基準による主成分数を用いた分類結果の各映像をジャンルで表したものを表5に示す。

これらの表を比較すると、様々なジャンルが各クラスタ内に混在していることから、従来のジャンルを用いた分類とは異なった分類結果が得られていることが確認できる。

例えば、表5のクラスタ1の結果を例として考察を行なう。ここにある3つの映像は、本提案手法によって、同系映像と判定された映像である。従来の分類基準であるジャンルという観点で考えると、これらの内、2つの映像は、“SF”というキーワードで同系と判定されているが、それ以外の“Horror”や“Classic”や“Action”などのキーワードでは一致しない。よって、検索に応用した場合を考えると、ユーザの選択したキーワードが“SF”でない場合、これらの映像は同時にユーザに提供されない。

しかし、本提案手法による分類結果を用いた場合、これらの3つの映像は、特徴量の観点からは同系映像と分類される映像であり、ユーザに同系映像として提供可能となる。これにより、ユーザは従来の映像検索では発見することができなかった、新しいコンテンツを見つけることが可能となる。また、これらの映像を実際に比較した結果、ある程度類似した印象を与える映像の分類結果であることが確認できた。

映像のジャンルのみでなく、様々な観点から提案手法の評価が可能であることが分かる。提案手法を用いた分類結果の評価手法については、今後も検討を続ける予定である。

5. まとめと今後の課題

近年のマルチメディア技術の発展に伴った映像コンテンツの大量化により、効率的なコンテンツの管理・検索技術の必要性を踏まえ、本稿では、効率的な映像管理・検索のための有効な映像解析を目的とし、映像から自動抽出可能な特徴量を軸とする特徴空間による同系映像の分類手法を提案した。

提案手法の評価実験として、映像から得られる特徴量を全て用いた映像の分類、主成分分析によって得られる主成分を用いた映像の分類を行なった。また、それぞれの分類条件における最適クラスタ数や最適主成分数の評価実験を行なった。その結果、ある程度映像の印象に基づく分類結果が得られていることが確認できた。

今後の課題としては、まず、これらの提案手法による分類結果の評価方法について検討する必要がある。また、現在分類に用いている特徴量には、類似するものが存在する。従って、特徴量の類似度に基づき、分類に用いている特徴量数を削減することで映像分類における計算量の削減が期待できる。

文献

- [1] Daniel Arijonf. "Grammar of the Film Language. ", London: Folcal Press. 1976
- [2] Aya Aner-Wolf, "Determining A Scene's Atmosphere by File Grammar Rules. " Proc. ICME, vol.2, pp.365-368, 2003
- [3] Cheung S-C C et al. "Efficient Video Similarity Measurement with Video Signature. " Trans CSVT. Vol13, No.1, pp.59-74. 2003
- [4] John R Smith et al, "Multimedia Semantic Indexing Using Model Vectors. " Proc. ICME, vol.2, pp.445-448. 2003
- [5] Chong-Wah Ngo et al, "On Clustering and Retrieval of Video Shots Through Temporal Slices Analysis. " Trans Mlt. Nol.4. No.4, pp446-458. 2002
- [6] Shinobu Hattori et al, "Content-based Video Classification for High-level Description Extraction. " Proc. SCI. vol.2, pp.461-464. 2003
- [7] Kentaro Dobashi: "Camera Working Parameter Extraction For Constructing Video Considering Camera Shake." ICIP, Vol.3, pp.382-385, 2001
- [8] ISO/IEC 15938, MPEG7: "Multimedia Content Description Interface," <http://mpeg.telecomitalia.com/>