

# 映像データの色彩情報を対象とした 時系列印象メタデータの生成方式

三浦 伸子<sup>†</sup> 倉林 修一<sup>††</sup> 清木 康<sup>†††</sup>

<sup>†</sup> 慶應義塾大学 総合政策学部

<sup>††</sup> 慶應義塾大学大学院 政策・メディア研究科

<sup>†††</sup> 慶應義塾大学 環境情報学部

E-mail: †{s02827nm,kurabaya,kiyoki}@sfc.keio.ac.jp

あらまし 映像データなどの時間軸に沿って内容および印象が推移するメディアを対象とした情報検索の実現には、映像の時間的推移に対応するメタデータ自動生成方式の実現が重要となる。本稿では、映像データを対象に、鑑賞者が受ける印象の大きな要素となる色彩情報を対象とし、映像中の印象の時間推移を反映した時系列印象メタデータの自動生成方式を提案する。本稿では、印象の時間的推移を、映像のストーリーと呼ぶ。本方式の特徴は、ユーザの視点に応じて、ストーリーの変化のポイントを動的に検出する点にある。ここで、ユーザの視点とは、ユーザが映像メディアから認識するストーリーの細かさ(以下、粒度とする)である。これにより、ユーザが発行する問い合わせが示すストーリーの粒度に適合した問い合わせ処理が可能となる。また、実際の映像を対象とした評価実験を行い、本方式の実現可能性および有効性について示す。

キーワード 映像検索, 色彩情報, 時系列印象メタデータ

## An Automatic Creation Method of Transitional Impression-Metadata for Color Information of Video Stream

Nobuko MIURA<sup>†</sup>, Shuichi KURABAYASHI<sup>††</sup>, and Yasushi KIYOKI<sup>†††</sup>

<sup>†</sup> Faculty of Policy Management, Keio University

<sup>††</sup> Graduate School of Media and Governance, Keio University

<sup>†††</sup> Faculty of Environmental Information, Keio University

E-mail: †{s02827nm,kurabaya,kiyoki}@sfc.keio.ac.jp

**Abstract** It is important to realize automatic metadata creation mechanisms for developing multimedia information systems. In particular, providing an automatic metadata creation mechanism for video data is significant. The video data have sequential structures and have different meanings according to various scenes. It is effective to deal with transition of meanings of sequential scenes for realizing story-based video search systems. In this paper, we present a novel metadata creation method dealing with transition of a story of video data. Our metadata creation method automatically generates transitional impression-metadata, which represents a transition of story, according to the color features of each scene. There are two main contributions. First, we develop an algorithm for detecting boundary of a story according to user's viewpoints for story recognition. We call the algorithm the Recursive Algorithm for Detecting Viewpoint Dependent Maximum Story-Granularity. Second, we implement a query processing mechanism for transitional impression metadata, and we clarify feasibility and effectiveness of our system by showing several experimental results.

**Key words** video retrieval, color information, Transitional Impression-Metadata

### 1. はじめに

放送番組の記録的・文化的価値に対する認識や知的活動の成

果としての利用の要望により、内容や印象の変化を含む連続メディアデータ(映像・音楽など)が豊富にデジタル情報として蓄積されている。特に映像データについては、(1)ブロードバン

ドの普及に伴ってネットワーク上の映像アーカイブが著しく増加している。(2) ソフトウェア及びハードウェアの進化により、高度な映像制作や映像編集が容易に行える。したがって、ネットワーク上に大量の映像データが置かれ、かつ、それらは多種多様化の傾向にある。

そのような大量の映像データの中から、利用者のニーズに合致した適切なデータを検索を実現するためには、自動的、かつ、客観的なメタデータ生成システムが重要となる。特に、映像データの印象は時間軸上で推移を伴うため、その時間的推移を反映したメタデータ生成(以下、時系列印象メタデータとする)方式の実現が肝要である[2][3]。そのような背景から、映像中の印象の時間的推移を反映したメタデータ抽出方式実現が有効である。

本稿の、映像データの中でも、特に、色彩情報が印象を決める主要な要素となっている映像を対象とする。映像全体の色彩情報の印象の時間的推移を色彩印象のストーリーと呼び、そのストーリーを反映させたメタデータ生成方式の実現を目指す。ストーリーの変化ポイント(以下、境界とする場合もある)を検出し、境界で区切られたまとまり(以下、ストーリー・グレイン)ごとの印象メタデータの順序集合を時系列印象メタデータとして採用する。

本研究の特徴は以下の3点にまとめることができる。

- ・映像中の色彩情報から印象を抽出する。
- ・印象の時間的推移(ストーリー)を扱う。
- ・印象でストーリーの境界を検知する。

本方式は、時系列印象メタデータを映像の物理的な変化ではなく、映像中の色彩情報から、心理的な特徴量を算出し、その特徴量の推移からストーリーを検出する点にある。ここで、特徴量とは、色彩情報から印象を形成するのに必要十分な単位時間分の映像データ(以下、アトミックシーンとする)が有する印象を表わした形容詞群を指す。本方式は、ストーリーが変化するポイントと物理的なシーンが変化するポイントは異なっていると仮定して、より鑑賞者が受ける印象を扱えるようなストーリーを重視したメタデータ生成方式を目指す。印象語によってストーリーを検出しているため、物理量が変化しても、印象は変化しないといったストーリー・グレインの算出も可能である。

本稿では、ストーリー(印象の時間的推移)を扱う方式を提案している。鑑賞者によって印象の感じ方・受け取り方には差がある。したがって、時系列印象メタデータはユーザの視点に応じて動的に生成される必要がある。ユーザが映像を鑑賞する視点の違いに応じてストーリーの変化ポイントを検出することが重要となる。本稿では、ユーザの視点を、ストーリー・グレインを生成する際のアトミックシーンのグループのされ方の決定要因となる閾値とする。ユーザごとに閾値を設定することで、ユーザが映像メディアから認識するストーリーの細かさ(以下、粒度とする)を反映した時系列印象メタデータを自動生成できる。例えば、ユーザAがある映像データを2つの境界、3つのストーリー・グレインから構成されていると認識したとしても、ユーザBは、同じ映像をより細かく鑑賞していたため、その映像データを4つの境界、5つのストーリー・グレインから

構成されていたと認識する場合もある。このようなユーザの映画鑑賞粒度を用いて、ユーザごとの映像の印象の感じ方・受け取り方の違いに応じた時系列印象メタデータ生成を目指す。

したがって、本方式によって時系列印象メタデータを生成する過程において、次の2つの機能が重要である。

#### (1) アトミックシーンの特徴量生成機能

時系列印象メタデータを生成するプリプロセスとして、色彩情報による静止画像メタデータ生成方式[11][5]を用いて、あらかじめ映像を物理的なパラメータから印象語によるパラメータに変換する。色彩情報による静止画像メタデータ生成方式[11][5]の詳細については2.3.1節において述べる。これにより、アトミックシーン間の距離は、印象語と印象語の距離によって求めることができる。このアトミックシーン間の印象の距離を、ここでは、Semantic Distanceと呼ぶ。本方式の特徴は、映像中の物理量の変化によるシーン検出ではなく、印象語の距離によってストーリーの変化ポイントを求める点にある。ユーザの視点である閾値と同様に、アトミックシーン間の関連の強さを示す Semantic Distance は、ストーリー・グレイン生成に大きな影響を与える。

#### (2) ストーリーの変化の単位を動的に検出する機能

本方式は、ユーザのストーリー認識粒度に応じて、このストーリー・グレインの検出を動的に行う。アトミックシーンの特徴量は、映像の微細な印象変化を記述するものであるためそのままでは問い合わせ処理に適さない。そのため、映像中の印象の大局的な流れを反映した、より大きなグレインの時系列印象メタデータを生成する必要がある。

本方式では、隣接するアトミックシーン間の距離が小さいものからグレインとしてグループ化し、より大きなグレインごとにメタデータを算出する処理を再帰的に行うことで、印象変化を反映する必要最低限の数のグレイン数となるストーリー・グラニュアリティ検出を実現する。これを再帰的最大のストーリー・グラニュアリティ検出(Recursive Algorithm for Detecting Maximum Story-Granularity)と呼ぶ。本稿では、プリプロセスのアトミックシーンの特徴量生成方式、そして、メインプロセスの鑑賞粒度に応じた最大ストーリー・グラニュアリティ検出方式、時系列印象メタデータ生成方式、および、生成したメタデータに対する時系列問い合わせ処理方式について述べる。

本方式による問い合わせ処理では、ユーザは閾値と時系列問い合わせ語(Query by Keyword)を発行する。閾値に応じて生成された時系列印象メタデータと、発行された映像の時系列印象問い合わせ語列間において相関量計算を行い、相関が高いものから、順にランキングする。ユーザは、以前鑑賞した映像を参考に、検索時に、自身が映像を鑑賞する際の細かさを反映した閾値を選ぶ。本方式では、閾値は映像データに依存するために、以前鑑賞した映像の粒度を反映した閾値を参考に、利用者が検索時の閾値設定を行う。また、本メタデータ生成方式の実現により、パターン・マッチングで映像が有する属性(タイトル・制作者など)を頼りに映像を検索していたが、映像から受ける印象によって主観的に検索を行うことができる。

## 1.1 関連研究

映像のシーン検出に関しては、様々な研究が行われている。物理的なパラメータで映像をシーンの境目を決定するといった研究は [1] でおこなわれている。[1] では、MPEG-2 を対象に、そのシーン検出と索引付け自動生成方式を提案している。シーン検出では、ショットごとのオブジェクトのヒストグラムと背景のヒストグラムの変化を用いて、ショット中のオブジェクトの動きを判断する。オブジェクトの変化が少ないシーン群は、それぞれシーンとして扱われる。それに対して本提案方式は、物理的なシーンの切れ目ではなく、ストーリー・グレインを特定する方式である。色彩情報による静止画像メタデータ自動生成方式 [11] [5] を用いて、色彩情報を印象語に変換し、印象と印象の距離によってストーリー・グレインを特定する点で異なる。

[13] では、動画像データの階層構造化を、シーン間の類似度をシーンに与えられたメタデータの相関量を求めることによって行っている。ここで、シーンのメタデータは、映像制作者のメモから得た出現オブジェクト名と、そのオブジェクトが映像中に現れる再生時間の出現割合によって定義する。まず、特徴が最も類似した 2 シーンを集約し、上位のノードを生成し、この類似度によるシーンの結合を繰り返すことによって、一つの映像データが階層（ツリー）構造として表現される。この階層構造は、映像の要約を提供するブラウジング機能において要約のポイントを提供するツールとして用いられる。各ノードは対応する映像データの内容を統合的に表現するビューを示す。これに対し、本論文で用いる階層構造も、シーン間をメタデータの類似性によって上位の節点を生成する点で [13] の方式と共通する。[13] の方式に対して、本方式は、2 シーン間の関連度を前のシーンの特徴が示す文脈に対する次のシーンの相関量を用いて測定するという点で異なる。また、本論文では、生成された階層構造を用いて検索の粒度を動的に設定し、文脈の時間的変化を扱う検索に適用するという点で異なる。また、本提案方式において、階層構造を生成する際、閾値を設定し、メタデータの精度に配慮した検索の粒度に限定している点も異なる。つまり、曖昧な時系列印象メタデータの生成を防ぐ範囲内において、ユーザの視点に応じた最大ストーリー・グラニューアリティを生成することが本研究の主要な特徴である。

[4] では、ショット間の関連を色情報の類似性によって求め、4 ショットを集約して上位の節点を生成することを繰り返し、最終的に 1 データが 1 頂点を持つ階層構造として表現される。上位の節点は、1 つのショットとして表現される。この階層構造によって様々な粒度からの動画像データブラウズやショットの検索が実現される。動画像データを色情報によって判別するといった点では、本提案方式と共通である。しかし、動画像データの内容や印象の変化を扱う検索は提案されていない。

[14] では、ショット間の色度数ヒストグラム差分、および、N-gram モデル、キャッシュモデルを用いて、動画像の意味的構造（ショットの内容に基づいた意味的にまとまった動画像区間のシーン）の発見に関する研究が行われている。N-gram とは、音声認識学の分野で盛んに用いられている言語モデルであり、ある単語をもとに次の単語を予想し、認識率の向上や計算時間

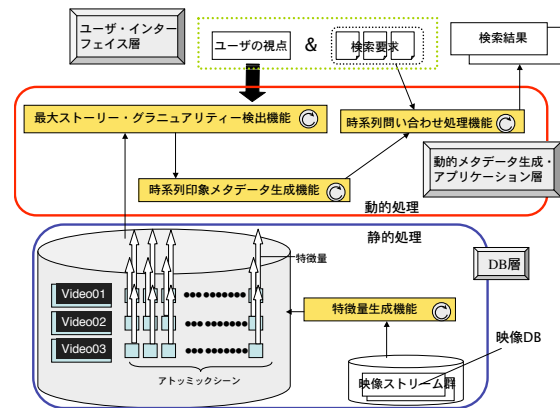


図1 システム概要

の削減を目指すものである。ヒストグラム差分をこの N-gram に関連づけることによって、動画像における内容や印象のまとまりを検出する。色彩情報によって、意味的なまとまりを検出する点で本研究と共通するが、本研究は、物理的なパラメータを印象語に関連づけることによって、ショット間で物理的なパラメータが変化し、かつ印象の変化がないストーリー・グレインの検出をおこなえる点で異なっている。

## 2. 提案方式

### 2.1 システムの概要

本システムは、図1に示すように、データベース層、動的メタデータ生成/問い合わせ処理層、そして、ユーザ・インターフェイス層の3層から成るデータベースシステムである。このレイヤ構造の特徴は、あらかじめ静的に生成可能なアトミックシーンの特徴量をデータベースとして共有し、ユーザごとの視点を反映させるメインプロセス機能群は、ユーザから検索要求がある度に、動的に処理する点にある。

(1) データベース層は映像ストリームを格納する映像データベースと、色彩情報によるアトミックシーンの特徴量生成機能により生成された特徴量を格納する印象データベースからなる。アトミックシーンの特徴量は、多数のユーザから共有可能であり、本システムでは、映像データが追加/更新された際に一度だけ生成・格納する。

(2) 動的メタデータ生成/問い合わせ処理層は、映像中のストーリーの変化ポイント検出を行う最大ストーリー・グラニューアリティ検出系、検出されたストーリーの粒度に応じてメタデータを生成する時系列印象メタデータ生成系、および、問い合わせとの相関量計量を行う時系列問い合わせ処理系から構成される。

(3) ユーザ・インターフェイス層では、ユーザの視点、および、時系列問い合わせ語を入力するのを支援するために、時系列問い合わせ処理に適したインターフェイスを提供する。これにより、ユーザは問い合わせ語の選定などに煩わされることなく、直感的に映像データを検索可能となる。

## 2.2 問い合わせ例

問い合わせ処理の際、ユーザは時系列印象問い合わせ語列とともに、閾値の設定を行う。本研究において、閾値は時系列印象メタデータを生成する際に、必要なストーリーの粒度を、個人の視点に応じて調節することを可能にするための重要なストーリー・グレイン決定要素となる。閾値の設定によって動的に生成される時系列印象メタデータ群に対して、時系列印象問い合わせ語（検索要求）を発行し、相関が高い時系列印象メタデータからランキングを表示する。ユーザは以下のような時系列問い合わせ語列および、閾値を発行する（表 1）。その後、データベースに格納されている映像データを対象に、最大ストーリー・グラニュアリティ検出系、および時系列印象メタデータ生成系が働き、ユーザが発行した閾値を反映した時系列印象メタデータ生成する（表 2）。

表 1 時系列問い合わせ例

User	検索要求	User's View (閾値)
User A	majestic-gentle-sad-appalling	0.7
User B	dark-intense-happy	0.4
⋮	⋮	⋮

表 2 時系列印象メタデータ例

時系列印象メタデータ
sad-hot-calm
dignified-calm-sad-appalling
⋮

## 2.3 メタデータ構造

本方式で扱う二つのデータ構造、1) アトミックシーン ( $i$ ) の特徴量  $\vec{v}_{[i]}$ 、および、2) ストーリー・グレインデータ  $S_i^l$  について述べる本方式は、映像データ中のアトミックシーンが有する特徴量から、ユーザの視点に応じて動的にストーリー・グレインデータを生成することで、映像データを対象した時系列印象検索を実現する。

### 2.3.1 アトミックシーンの特徴量

アトミックシーンの特徴量を生成する際、色彩情報による静止画像メタデータ生成方式 [11] [5] を用いる。色彩による静止画像メタデータ生成方式 [11] [5] では、マンセル表色系 [15] で表される色彩情報と人間の印象の関係が示されているカラーイメージスケール [10] を用いてマンセル表色系 [15] で表される色彩情報を自動抽出し、それらの要素群と印象語の相関量を計算するためにカラーイメージスケール [10] の示した相関関係を用いている。カラーイメージスケール [10] には、色彩と印象語の関表す統計データが示されている。本研究では、色彩による静止画像メタデータ生成方式 [11] [5] を用いて生成された静止画像メタデータをアトミックシーンの特徴量として採用する。[11] [5] で生成された静止画像メタデータは、182 次元の印象語ベクトルとして定義されている。したがって、アトミックシーンの特徴量も同様にベクトルとして表現されるため（特徴量:  $\vec{v}_{[i]}$ ）、ベクトル間の相関量計算を行うことで、アトミックシーン間の印象の距離 (Semantic Distance) を計量することが可能である。また、特徴量  $\vec{v}_{[i]}$  は、映像データベースから、あ

表 3 アトミックシーンの特徴量の例

key	value
authoritative	0.183324
majestic	0.180815
intense	0.158708
bold	0.158708
forceful	0.158708
⋮	⋮

らかじめ静的に生成される。特徴量は、全ての問い合わせにおいて、共有して使用される基本データ群である。表 3 に、アトミックシーンの特徴量の例を示す。

### 2.3.2 ストーリー・グレインデータ

ストーリー・グレインデータ  $S_i^l$  は、特徴量群を対象に、2.4.1 節で示す再帰的最大ストーリー・グラニュアリティ検出アルゴリズムを適用することで動的に生成される、ユーザの視点を反映したストーリー構造を示すツリー状のデータである。 $l$  はツリーの高さ (レベル) を示し、基底レベル ( $l = 0$ ) は、2.3.1 節で示した特徴量 ( $\vec{v}_{[i]}$ ) 群である。

$$S_{[i]}^l := \begin{cases} S_{[o\dots n]}^{l-1}, & l > 0 \\ \vec{v}_{[i]}, & l = 0 \end{cases} \quad (1)$$

ツリーのレベルが上がるほど、ストーリーの粒度は大きくなり、映像の大局的な印象の変化を示す時系列印象メタデータ生成に利用可能となる。一方、ツリーのレベルが低い場合は、映像の局所的な印象の変化を検索対象とする事ができるため、ユーザの幅広い視点に応じた問い合わせ処理が可能となる。このように、ストーリー・グレインデータ  $S_i^l$  は、時系列印象問い合わせをユーザが発行するたびに、それと同時に発行される閾値に応じて動的に生成されるため、データベース中には静的に格納されない。

### 2.4 メイン処理機能群

メイン処理機能群は、最大ストーリー・グラニュアリティ検出系と時系列印象メタデータ生成系、および、問い合わせ処理系から成り立つ。本節では、最大ストーリー・グラニュアリティ検出系と時系列印象メタデータ生成系について述べる。

#### 2.4.1 最大ストーリー・グラニュアリティ検出アルゴリズム

最大ストーリー・グラニュアリティ検出系は以下の 3 ステップからなる。

**Step-1.1 相関量計算:** 一定間隔の各アトミックシーンから抽出された特徴量  $S_{[i]}^{l=0}$  (ただし  $l = 0$  なので  $\vec{v}_{[i]}$  となる) を用いて、ストーリー・グレインを生成していく (図 2)。隣接するアトミックシーンの特徴量間の余弦 ( $c = \frac{S_{[i]}^l \cdot S_{[i+1]}^l}{|S_{[i]}^l| \cdot |S_{[i+1]}^l|}$ ) を計量する。

**Step-1.2 グレイン形成:**  $c > t$  (閾値) の場合、 $S_{[i]}^l, S_{[i+1]}^l$  は、双方のベクトルの平均値を求め ( $\frac{S_{[i]}^l + S_{[i+1]}^l}{2}$ )、 $l + 1$  階層において、ストーリー・グレイン ( $S_{[i]}^{l+1}$ ) を生成する。

$c < t$  の場合、 $S_{[i]}^l, S_{[i+1]}^l$  は、 $S_{[i]}^l$  自体は、 $l + 1$  階層にそのまま持ち越され、単一の特徴量としてストーリー・グレイン

$(S_{[i]}^{l+1})$  を生成する．その際， $S_{[i+1]}^l$  は  $S_{[i+2]}^l$  との間で，余弦を計量する．

**Step-1.3 再帰的呼び出し:** Step-1.2 で生成されたストーリー・グレイン  $S_{[i]}^l$  は，階層  $l$  が上がるたびに，ストーリー・グレイン生成関数  $f_{\text{divide}}()$  が再帰的に呼び出され，徐々に大きなグレインへと変化する．大きなストーリー・グレインが形成されるにつれて，ストーリー・グレイン間の Semantic Distance はしだいに大きくなっていく．グレイン間の合成が一度も起きていない場合，その当該階層を最大ストーリー・グラニューアリティとする．

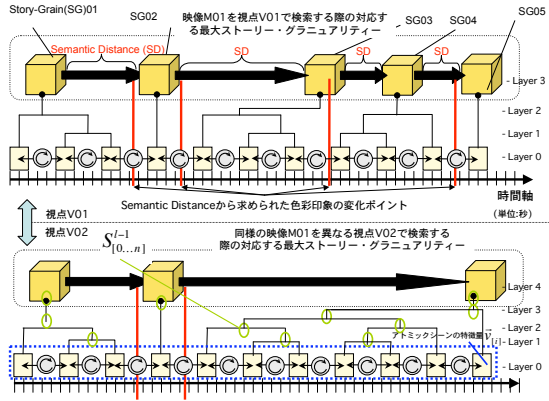


図 2 再帰的的最大ストーリー・グラニューアリティ検出アルゴリズム

### 2.4.2 時系列印象メタデータ生成方式

**Step-2.1 トラバース:** Step-1.3 で検出された最大ストーリー・グラニューアリティに応じて時系列印象メタデータを生成する．まず，再帰的アルゴリズムで生成されたグラニューアリティの構造をトラバースする関数  $f_{\text{traverse}}()$  を適用し，最大ストーリー・グラニューアリティを構成するアトミックシーンの特徴量の集合  $S_{[n]}^l$  をストーリー・グレインごとに収集する．

**Step-2.2 時系列印象メタデータ生成:** 時系列印象メタデータ生成関数  $f_{\text{generate}}()$  を用いて，最大ストーリー・グラニューアリティが有する各ストーリー・グレインに対応する印象メタデータを生成すると共に，その印象メタデータ群の順序集合を生成する (図 4)．まず，Step-2.1 で収集した，各ストーリー・グレインを構成するアトミックシーンの特徴量の平均値を算出する．平均値が有する印象語群中で，値が高い上位 5 印象語を，各ストーリー・グレインの印象メタデータとして採用する．次に順序集合生成オペレータ ( $\cup$ ) を用いて，印象メタデータの順序集合を生成する．これが時系列印象メタデータとなる (式 (2))．

$$f_{\text{generate}}(S_{[o..n]}^l, t) := \bigcup_{l=1}^{p_l > 0} \left\{ \frac{\sum_{j=0}^{j > m-1} |V|}{m} \leftarrow f_{\text{traverse}}(S_{[i]}^l) \right\} \quad (2)$$

```
function:  $f_{\text{divide}}(S_{[o..n]}^l, t)$ 
for all  $S_{[i]}^l$  such that  $0 \leq i \leq n$  do
  if  $S_{[i]}^l * S_{[i+1]}^l > t$  then
     $s \leftarrow S_{[i]}^l + S_{[i+1]}^l$  //(Step1.1)
     $p_l \leftarrow p_l + 1$  //(Step1.1)
  else
     $s \leftarrow S_{[i]}^l$ 
  end if
  addElement( $S^{l+1}, s$ ) //(Step1.2)
end for
if  $p_l > 0$  then
   $f_{\text{divide}}(S_{[0, \dots, (n-p_l)]}^{l+1})$  //(Step1.3)
else
  return  $S$ 
end if
```

図 3 再帰的ストーリー・グラニューアリティ検出アルゴリズム

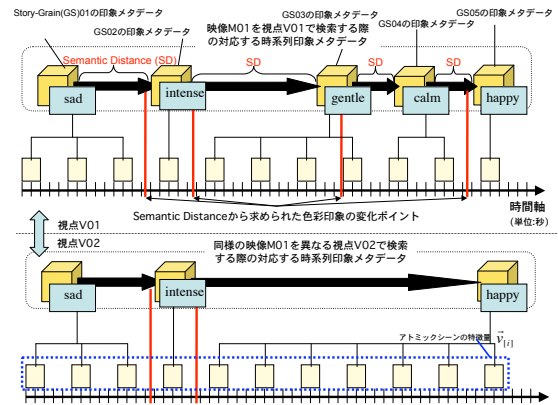


図 4 時系列メタデータ生成方式

## 3. 問い合わせ処理

本方式における問い合わせ処理は，時系列印象問い合わせ語列と，第 2.4 のメタデータ生成方式により生成された時系列印象メタデータ群との相関計量により実現される．ただし，双方ともに時系列的なメタデータベクトル列であるため，以下に示す 2 ステップにより問い合わせ処理を行う．

### (1) 構造的マッチング

時系列問い合わせ語列と関連性を計量可能な時系列印象メタデータ群を検索するために，与えられた時系列印象問い合わせ語列の語数と等しい数の印象語から構成される時系列印象メタデータ群を検索する．本方式では，時系列印象問い合わせ語列とともに，ユーザの視点として閾値が与えられ，それに応じて時系列印象メタデータを動的に生成するので，構造的マッチングにより相関量計算対象外となった時系列印象メタデータ群は，ユーザの求める構造と異なるストーリーを有することとなる．つまり，ユーザの視点である閾値によってストーリー・グレインが設定されているため，問い合わせ語列の数と異なる構造を有する時系列印象メタデータ群は，ユーザの視点を反映していないと見なし，検索対象集合から外す．



## (2) 相関量計量

次に、時系列印象問い合わせ語列 ( $S$ ) と、時系列印象メタデータ ( $S'$ ) 間の相関量を求める。(式 (3)) すでに構造的マッチングによって、時系列印象メタデータ群は  $S$  の印象語数と等しいため、 $S$  と  $S'$  のそれぞれが有する各印象語の余弦を合計した値を算出する。算出した相関量のスコア順に検索結果として出力する。

$$f_{\text{eval}}(S, S') := \sum_{n>0}^{i=0} (S[i] \cdot S'[i]) \quad (3)$$

## 4. 実験

### 4.1 実験概要

ここでは、提案方式による時系列印象メタデータ生成実験について述べる。映像データの色彩情報を対象とした時系列印象メタデータ生成方式の実現において、以下の二つの実験を行う。

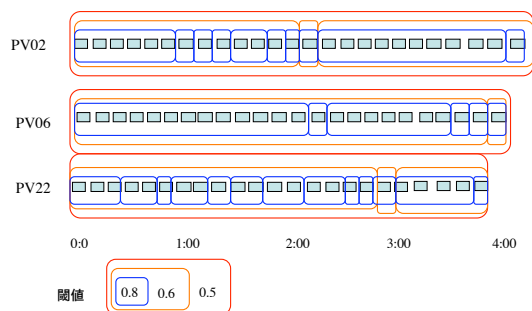


図 5 閾値ごとのストーリー・グレイン結果サンプル

### 4.2 実験 1: 閾値に応じたストーリー・グレイン生成の評価

ユーザの視点となる閾値は、ストーリー・グレインの生成のされ方にどのように影響するのかを検証する。

対象データ: 本実験で用いる対象データとして、色彩情報が印象を与える主要な要素となっている映像集の中から、色彩情報が豊富なプロモーションビデオクリップ (PV) 40 本を選択した。

閾値: 閾値を  $t = [0.5, 0.6, 0.8]$  に応じて、生成される最大ストーリー・グラニュアリティにおけるストーリーグレインの数の変化を確認する。

### 4.3 実験 2: 印象語を用いたストーリー・グレイン検出

印象語によってストーリーを扱うことのメリットを提示するために、物理的パラメータを用いた場合のシーン検出と、印象語を用いた場合のストーリー・グレイン生成の違いを明らかにする。

対象データ: 実験 1 で使用したプロモーションビデオクリップ (PV) 40 本の時系列印象メタデータを生成する事で、映像の物理的な変化に依らない時系列印象の変化を有するデータを検証する。

閾値: 閾値を与えずアトミックシーンの特徴量の比較を行う。

表 4 時系列印象メタデータ (閾値:0.5)

SGID	PVID	impression
1	PV05	light(0.2339)-neat(0.1973)-clean(0.1941) clear(0.1888)-cleanandfresh(0.1861)
2	PV05	free(0.4886)-sweetsour(0.487)-citrus(0.328) healthy(0.325)-fresh(0.3245)
3	PV05	light(0.233)-clean(0.194)-clear(0.1885) cleanandfresh(0.1861)-pure(0.1831)
1	PV06	warm(0.2693)-dry(0.2554)-cool(0.226) pristine(0.184)-mirthful(0.1709)
2	PV06	fresh(0.284)-free(0.106)-yourthful(0.147) citrus(0.0894)-cute(0.0819)
3	PV06	warm(0.271)-restful(0.1699)-tranquil(0.1691) plain(0.1547)-simpleandappealing(0.1345)
⋮	⋮	⋮

表 5 時系列印象メタデータ (閾値:0.8)

SGID	title	impression
1	PV07	solemn(0.247)-exact(0.135)-provincial(0.1121) formal(0.1076)-bitter(0.0897)
2	PV07	light(0.2038)-chic(0.1455)-sober(0.1389) dry(0.1377)-metallic(0.1337)
3	PV07	light(0.225)-refresing(0.1239)-simple(0.1239) festive(0.1239)-pure(0.124)
4	PV07	warm(0.2693)-dry(0.2554)-cool(0.226) pristine(0.184)-mirthful(0.1709)
5	PV07	chic(0.1892)-sober(0.1681)-solemn(0.1647) subtleandmysterious(0.1511)-exact(0.1221)
6	PV07	warm(0.1861)-subtleandmysterious(0.1834)-exact(0.1222) alluring(0.1144)-bitter(0.0882)
⋮	⋮	⋮

### 4.4 実験結果

#### 4.4.1 実験 1: 閾値に応じたストーリー・グレイン生成の結果

対象サンプルデータ 40 本を対象に、 $t = [0.5, 0.6, 0.8]$  ごとにストーリー・グレインを生成し、そのストーリー・グレインの個数をグラフにまとめた (図 6)。閾値が低い場合は、ストーリー・グレイン生成個数が極めて少ない。それに対して閾値を高くするほど、ストーリー・グレイン生成個数は増加していく。閾値はユーザがその映像を鑑賞する際の粒度を反映した値である。つまり、閾値を低く設定すると、映像のストーリーの大局的な流れに着目したストーリー・グレイン生成がおこなえ、また、閾値を高く設定すると、映像の微細な変化を考慮したストーリー・グレイン生成が行える。本研究は、ユーザが映像データを鑑賞する際の粒度をストーリー・グレインの生成個数によって扱うことが可能であることを検証できた。本実験結果は、本方式を実際の映像データに適用させる場合、閾値ごとに異なる個数のシーン群が検出でき、それによってユーザごとの鑑賞粒度に応じた検索が可能となることを示している。加えて、閾値を低く設定した場合の印象の変化ポイントは、閾値を高く設定した場合の印象の変化ポイントを覆ってストーリー・グレインを形成することはない (図 5)。つまり、閾値を低く設定した際に生成されるストーリー・グレイン間の Semantic Distance は、極めて離れていることがわかる。さらに、 $t = [0.5, 0.8]$  での、実際に生成された最大ストーリー・グラニュアリティにおける、時系列印象メタデータを表に示す。表の、S GID はストーリー・グレインごとの ID を、PVID は映像番号を、impression は各 S G ごとの印象メタデータを表す (表 4)(表 5)。

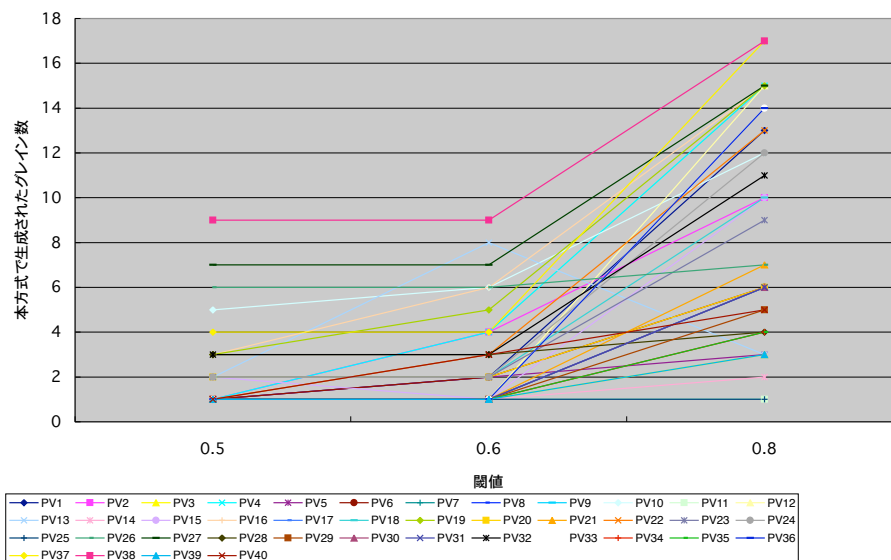


図 6 閾値ごとのストーリー・グレイン数集計結果

#### 4.4.2 実験 2: 印象語を用いたストーリー・グレイン検出の結果

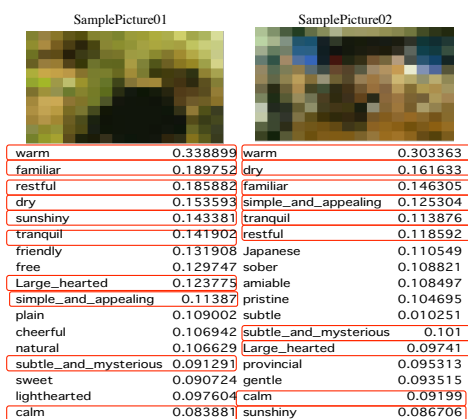


図 7 印象語を用いたストーリー・グレイン検出具体例

隣接するアトミックシーンに対して特徴量をそれぞれ生成し、ベクトル内の上位 17 個の印象語群を図に表した (図 7<sup>注1)</sup>)。双方のベクトル内における上位印象語において、多少の変動は見られるが、SamplePicture01 のベクトルの上位にある印象語 10 個中の全ての単語が、SamplePicture02 のベクトルの上位に出現した。この 2 つのアトミックシーンは、グレイン生成の段階においても全ての閾値 ( $t = [0.5, 0.6, 0.8]$ ) において、同じストーリー・グレインに含まれており、双方は似ている印象

を受ける関係を持つことが証明できた。このように、本方式によって、物理的なパラメータのみを用いるだけでは検出しにくいシーン間の関連の強さの検出が可能となることを検証できた。本実験結果は、本方式を実際の映像データに適応させた場合、類似する印象を有するシーン群を検出することができ、それによって生成された時系列印象メタデータに対する連続的な印象語を用いた検索が可能となることを示している。

## 5. まとめ

本稿では、映像中の色彩情報を対象に、時系列印象メタデータの生成方式を提案した。本提案方式により、映像データを対象に、鑑賞者が受ける印象の大きな要素となる色彩情報を用いた映像中の印象の時間的推移を反映した時系列印象メタデータの生成が可能となった。本提案方式によって、映像の属性に頼らない、映像のストーリーを考慮した主観的な検索が行える。

本研究の特徴は、映像中の色彩情報から取得される印象の変化に着目し、また、鑑賞者ごとに異なる印象の変化ポイントも考慮した上で、ユーザごとに動的に時系列印象メタデータを生成している点にある。

今後の課題としては、映像から受ける印象をよりの確に反映した時系列印象メタデータ生成方式提案に向けて、被験者を募り、時系列印象メタデータ精度実験を行う必要がある。そして、本研究の有効性を検証するために、実際の映像データを対象とした解析的実験を今後行う必要がある。

謝辞 本研究には、様々な方の貴重な指導とご意見をいただきました。中でも、本研究に貴重なご助言をいただきました吉田尚史氏 (慶應義塾大学大学院 政策・メディア研究科)、および河本穰氏 (慶應義塾大学大学院 政策・メディア研究科) に感謝

(注 1): 著作権保護のために、モザイク処理を施しているが、本方式は色彩情報を用いるため、ほぼ同様の結果が得られる。

いたします。

## 文 献

- [1] Haoran Yi, Deepu Rajan and Liang-Tien Chia: “A motion based scene tree for browsing and retrieval of compressed videos”, *MMDB '04: Proceedings of the 2nd ACM international workshop on Multimedia databases*, pp.10–18 2004.
- [2] 伊地智 麻子, 清木 康: “音楽データの印象の時間的推移を扱う印象メタデータ自動生成方式”, Vol.44 No.SIG 18(TOD 20) 情報処理学会論文誌:データベース Dec. 2003.
- [3] Ijichi, A. and Kiyoki, Y.: “A Kansei Metadata Generation Method for Music Data Dealing with Dramatic Interpretation”, *Information Modeling and Knowledge Bases*, Vol.XVI, IOS Press, (accepted, 13 pages).
- [4] J-Y Chen, C. Taskiran, A. Albiol, E. J. Delp and C. A. Bouman: “ViBE: A Compressed Video Database Structured for Active Browsing and Search”, submitted to *IEEE Transactions on Multimedia*.
- [5] 北川 高嗣, 中西 崇文, 清木 康: “静止画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用” 情報処理学会論文誌:データベース Vol.43 No.SIG12 - 004
- [6] 清木 康, 金子 昌史, 北川 高嗣: “意味の数学モデルによる画像データベース検索方式とその学習機構”, Vol.J79-D, No.4, pp.509-519 (1996) 電子情報通信学会論文誌
- [7] Kiyoki, Y., Kitagawa, T. and Hayama, T.: A Metadata System for Semantic Image Search by Mathematical Model of Meaning, *Multimedia Data Management Using Metadata to Integrate and Apply Digital Media*, Chapter7, McGrawHill(1998).
- [8] Kurabayashi, S. and Kiyoki, Y.: “An Active Multidatabase System Architecture with an Aspect-Oriented Media Modelling Framework”, *Information Modelling and Knowledge Bases*, Vol.XVI, IOS Press, (accepted, 20 pages).
- [9] Kurabayashi, S. and Kiyoki, Y.: “Aspect-ARM: An Aspect-Oriented Active Rule System for Heterogeneous Multimedia Information”, *The Fifth International Conference on Web Information Systems Engineering (WISE2004)*, Lecture Notes in Computer Science, Springer-Verlag, No. 3306, pp. 659–667. (November, 2004).
- [10] 小林 重順: “カラーイメージスケール”, 講談社 (1990).
- [11] 小谷拓矢, 清木康, 北川高嗣: “色彩情報による静止画像メタデータ生成方式と意味的画像検索への適用”, 第9回データ工学ワークショップ (DEWS'98) 論文集, 電子情報通信学会, 1998 .
- [12] Yumiko Sato, Yasushi Kiyoki “A semantic associative search method for media data with a story”, *IASTED International Conference on Applied Informatics*(February,2000)
- [13] 柴田 正啓: “映像の内容記述モデルとその映像構造化への応用”, 電子情報通信学会論文誌, Vol.J78-D-II, No.5, pp.754-764 (1995).
- [14] 谷澤 和昭: “動画画像の特徴量を用いた意味的構造の自動検出”, データベースシステム, 120-11(2000,1,24)
- [15] 千々岩英彰: “色彩学概説”, 東京大学出版会, 2001.