類似理由の提示機能を具備した類似動画検索システムの構築

木村 彰吾 †， 林 貴宏 ††， 尾内 理紀夫 † ††

† 電気通信大学大学院 電気通信学研究科 情報工学専攻
†† 電気通信大学 電気通信学 部情報工学科

Content-Based Video Retriaval with reasons of similarities using images&sounds

ShogoKIMURA †， TakahiroHAYASHI ††， RikioONAI † ††

† Department of Computer Science, Graduate School of Electro-Communications, The University of Electro-Communications

†† Department of Computer Science, The University of Electro-Communications

動画を入力とする類似動画検索における問題点として、多数の類似尺度が存在することや、動画間の関連性が分かりづらいことが挙げられる。そこで、ユーザが検索目的に応じて類似尺度を選択でき、かつユーザへ類似理由を提示する機能を具備した類似動画検索システムの構築を目指す。本論文では、色の類似性、カメラワークの類似性、映像中の人物数の類似性、音の種類の類似性、の 4 種類の類似尺度によってデータベース中の類似動画を検索する機能、また類似理由をテキストで生成する機能を実装したシステムを試作した。また、これらの類似尺度と類似理由提示の有用性を被験者実験によって検証した。実験の結果、実装した 4 種類の類似尺度はそれぞれユーザの類似判断基準になり得ること、および類似理由を提示することでユーザの類似判断基準に影響を与えることを確認した。

1 はじめに

近年、ハードディスクの大容量化と動画投稿サイトの普及によって、非常に多くの動画に触れる機会を得た。それに伴い、動画の検索技術の発達が望まれている。現在はキーワードを入力とした、動画に付随するテキストを利用した検索が主流である。しかし、動画情報は文字情報に比べてはるかに多くの情報を含んでおり、多義性やあいまい性が高い。このため、キーワードのみでは目標とする動画を表現することが難しく、必ずしも満足な検索結果が得られない。また、人手でテキストを付加する手間も無視できない。それに対して、動画自体を入力として、その入力動画と類似する特徴を持った動画を検索する、Content-Based Video Retrieval(CBVR) と呼ばれる類似動画検索手法が存在する [1) 2)]。

しかし、CBVR にも問題点がある。動画は多義性のある非常に情報量の多いデータであり、類似する特徴といっても、色の類似性やカメラワークの類似性、音の類似性など、様々な類似尺度が存在する。そのため、どの類似尺度に着目するかによって、動画検索の結果は変動する。そこで、それぞれの類似尺度を個別に選択可能にすれば、ユーザの検索目的に応じた検索が可能となると思われる。

しかし、ユーザにとって、選択した類似尺度と自身の主観的な判断基準が必ずしも一致しているとは限らない。また、ユーザの動画間の類似判断基準には個人差があり、全てのユーザが満足する類似尺度を実装することは困難である。そのため、ユー

ザの判断基準に近い尺度での検索を行うことができない場合もある。このような場合、類似と判定された動画に関しては、どこが類似しているのかが分かりにくい。そこで、動画と共にその動画の類似理由を提示することが可能となれば、入力動画とシステムの提示する動画との関連性が分かりやすくなると思われる。

そこで本論文では、類似尺度をユーザが選択でき、類似動画を提示すると同時に類似理由もユーザに提示するシステムの構築を目指す。

以下、2 章〜5 章では試作した提案システムについて述べる。6 章では試作した提案システムに対する評価実験について述べる。7 章では既存研究とその問題点について述べる。

2　提案システム:似テ似テ動画 (仮)

システム構成を Fig.1 に示す。提案システムでは、まずユーザがシステムへの入力として動画を与え、用意された類似尺度の中から希望の類似尺度を選択する。システムは特徴量抽出部で入力動画の特徴量を抽出する。類似動画検索部において、入力動画とデータベース内の各動画に対し、選択された類似尺度に基づき、特徴量を比較し、類似度を計算する。また、類似理由生成部で類似理由を表す文の生成を行う。そしてユーザに対して類似度順に動画を提示すると共に、生成した類似理由文も合わせて提示する。なお、データベース内の動画は全てショット単位に分割されており、各ショットに対して特徴量の抽出を行っている。

3　特徴量抽出部

入力動画中の全ショット、およびデータベース中の動画群の全ショットに対して、類似動画検索のための特徴量を抽出する。動画をショットに分割するために、動画中の各フレームで RGB ヒストグラムを算出し、時間的に隣接するフレーム間で RGB ヒストグラムが大幅に変化している点を、

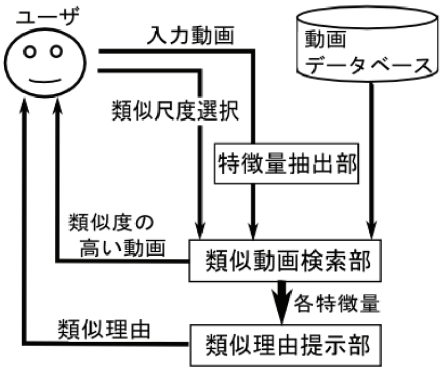

Fig. 1　システム構成

ショットの切り替わる点と判定する処理を用いる [3]。

3.1　色に関する特徴量

動画中に占める割合が多い色ほど、その色はその動画の特徴を現す色であると考えられる。本論文では、各色の含有率 (どの色がどの程度の割合で存在するのか) が近い動画ほど動画から受ける印象も類似すると仮定する。そこで、色情報に着目し、色の含有率を計算する。

具体的な手法としては、まずショットの先頭フレームを抽出する。そしてそのフレームを 16 × 12 のブロックに分割し、各ブロック内で最も出現頻度が高い色をそれぞれ決定する。ただし、色は 11 の基本色 (白、黒、灰色、茶色、赤、オレンジ、黄色、青、緑、ピンク、紫) のいずれかとする。この 11 色は画像処理においても重要な色であるとされる [4] [5] [6]。基本色への分類は、HSV の各値 ($0 \leq H < 360$, $0 \leq S \leq 255$, $0 \leq V \leq 255$) から、図 2 のように行う。

3.2　カメラワークに関する特徴量

各ショットのカメラワークを、フィックス、ズーム、パン (右)、パン (左)、チルト

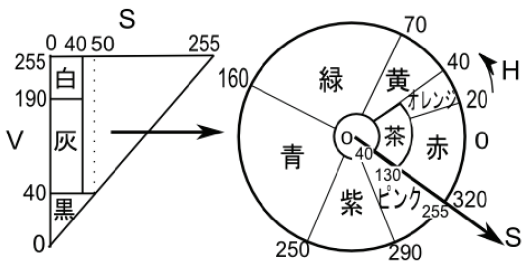

Fig. 2　HSV と基本 11 色の対応関係

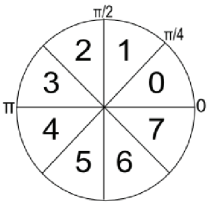

Fig. 3　θ の分類

(上)、チルト (下)、ハンドカメラの 7 種類に分類する。フィックスとはカメラが固定された状態で撮影された映像、パンはカメラが左右に動きながら撮影された映像、チルトはカメラが上下に動きながら撮影された映像である。ハンドカメラはカメラワークの種類を表す用語ではないが、本論文では短時間で様々な方向にカメラが動く、手ぶれのあるような映像を表すこととする。

カメラワークの推定には、時刻 t でのショット中のフレーム F_t と時間的に隣接するフレーム F_{t-1} においてオプティカルフローを計算する。本システムでは勾配法のひとつである Lucas-Kanade 法 [7] を利用した。オプティカルフローは、フレーム画像中の横 40 点、縦 30 点の計 1200 点の格子点中で、画面端に近い点を除いた 936 点 (縦 36 ×横 26 点に対し算出する。画面端に近い点を利用しないのは、画面外に移動する点が存在すると、その点に対するオプティカルフローを正しく算出することができないためである。点 i(i=0,1, ⋯ ,935) の動きベクトル (v_{x_i}, v_{y_i}) に対し、動きベクトルの方向 Θ_i を、次式、

$$\Theta_i = \tan^{-1} \frac{v_{y_i}}{v_{x_i}} \qquad (1)$$

により求める。さらに Θ_i の値に応じて各点を Fig.3 に示す 8 種類の方向に分類し、向きヒストグラム p_j(j=0,1, ⋯ ,7) を作成する。この向きヒストグラムにおける各要素の出現回数の分散 δ_p を、

$$\delta_p = \frac{1}{936} \sum_{j=0}^{7} (p_j - \frac{936}{8})^2 \qquad (2)$$

と定義する。

さらに、あるフレームでの動きベクトルの平均ベクトルの大きさ Q を次式、

$$Q = \frac{1}{936} \sqrt{(\sum_{i=0}^{935} v_{x_i})^2 + (\sum_{i=0}^{935} v_{y_i})^2} \qquad (3)$$

により定義する。また、動きベクトルの大きさの平均 W を次式、

$$W = \frac{\sum_{i=0}^{935} \sqrt{v_{x_i}^2 + v_{y_i}^2}}{936} \qquad (4)$$

により定義する。δ_p、W、Q を用いて、以下の条件によりカメラワークを分類する。

- $W > 0.75$, $Q < 0.1$, $\delta_p < 0.25$
  ⇒ ズーム
- $W > 0.75$, $Q > 0.75$, $\delta_p > 0.6$
  ⇒ パン・チルト
  方向は argmax(p_j) に応じて決定
- 上のいずれの条件も満たさない場合
  ⇒ フィックス

対象とするショット内の全フレーム $F_t(t = 1, \cdots, L)$(L はフレーム長) に対して分類を行い、各カメラワークの出現回数からカメラワークヒストグラム $w_n(n = 0, \cdots, 5)$ を算出する。なお、フィックスのとき $n = 0$、ズームのとき $n = 1$、パン (右) のとき $n = 2$、パン (左) のとき $n = 3$、チルト (上) のとき $n = 4$、チルト (下) のとき $n = 5$ とする。カメラワークヒストグラムの分散 δ_w を、

$$\delta_w = \frac{1}{L}\sum_{n=0}^{5}(w_n - \frac{L}{6})^2 \qquad (5)$$

により定義し、δ_w と w_n を用いて、ショット全体のカメラワークの種類 C を以下の条件により推定する。

- $\delta_w \leq 0.4$

  $\Rightarrow$ C = "ハンドカメラ"

- $\delta > 0.4$

  $\Rightarrow$ C = argmax(w_n) に対応するカメラワーク

3.3 映像中の人物数に関する特徴量

各ショットに映る人物の人数と、それらの人物が映像内で占める面積の割合を計算する。人数は映像中のフレームに対して顔検出を行い、"検出した顔の数＝ショット中の登場人物数"であると推定する。顔検出は Viola らの手法 [8] によって行う。顔の検出精度はフレーム内のノイズや顔の角度、髪型等に強く依存しており、大幅に低下する場合もある。そこで、各ショットの全フレームに対して顔認識を行い、全フレーム数の 1/3 以上のフレームで検出できたものを結果することで、この問題に対応する。そして検出した顔に対して、それらが映像中で占める面積の割合を、検出した顔の外接円の面積により近似する。検出した顔の数を f、各顔の占める割合を $t_i(i = 1, 2, \cdots, f)$ とする。

3.4 音の種類に関する特徴量

各ショット中の音を、"音楽"、"音声"、"無音" のいずれかに分類する。さらにショット内での音の平均パワーを計算する。

音楽と音声の分類は、高柳らのソナグラムの画像特徴を用いた分類手法 [9] を用いる。ソナグラムは横軸に時間、縦軸に周波数を 255 段階で分割したものとして、ある時刻のある周波数のパワーを色濃度で表した画像であり、その色濃度は $p(x, y)(0 \leq x \leq L, 0 \leq y \leq 255)$($L$ はショット長に応じたピクセル数) により定義される。p(x,y) から短時間区間を抽出し、その区間において "音楽"、"音声"、"無音" の分類を行う。ショット内で区間を動かしながら分類を繰り返す。全区間で"無音"と判定されれば"無音"、"音声"と判定された回数が"音楽"よりも多ければ"音声"、"音楽"が多ければ"音楽"であると判定し、この判定結果により音の種類 s を定義する。

また、音の平均パワー E は、全ての時刻における全周波数のパワーの総和を時刻で割ったものであり、ソナグラムにおける色濃度 p(x,y) を用いて、次式、

$$E = \sum_{y=0}^{255}\sum_{x=0}^{L} p(x, y)/L \qquad (6)$$

により定義する。

4 類似動画検索部

入力動画から切り出したショットの特徴量と、データベース中の動画群の各ショットにおけるそれぞれの特徴量とを、ユーザが選択した類似尺度に基づいて比較し、類似度を計算する。そして類似度の高い順に、そのショットを含む動画をユーザに提示する。なお、入力動画から切り出すショットは、重要なシーンにおいては音のパワーが強くなると仮定して [12]、音の平均パワーが最も大きいショットとした。

4.1 色に基づく類似度

動画間の類似度を、基本 11 色各色の出現頻度の類似性に基づいて計算する。

3.1 で示した 16 × 12 の各ブロックにおいて最も出現頻度の高い色に対し、各色の出現回数をカウントした色ヒストグラム $c_i(i = 0, \cdots, 10)$ を作成し、色ベクトル $v = (c_0, c_1, \cdots, c_{10})$ を定義する。そして、入力動画 V_0 の色ベクトル v_0 と、データベース中のショット $V_k(1 \leq k \leq N)$(N はデータベース中の総ショット数) の色ベクトル v_k との類似度 S_k を、

$$S_k = \frac{v_0 \cdot v_k}{|v_0||v_k|} \tag{7}$$

により定義する。しかし、フレーム全体の色ヒストグラムだけでは、各色の出現位置の類似性を考慮することができない。そこで、フレームを縦・横それぞれ 2 分割した計 4 つの領域においてもそれぞれ色ヒストグラムを作成し、同様に類似度を計算することで、”画面上部の色が似ている”といった位置も考慮した類似度計算が可能となると考える。それぞれの結果を $s_{1,k}$、$s_{2,k}$、$s_{3,k}$、$s_{4,k}$ とする。

以上、S_k と $s_{1,k}$〜$s_{4,k}$ を用いて、入力動画 V_0 とデータベース中のショット V_k の、色に基づく類似度 S_{color} を、

$$S_{color,k} = 4S_k + s_{1,k} + s_{2,k} + s_{3,k} + s_{4,k} \tag{8}$$

により定義する。

4.2 カメラワークに基づく類似度

入力動画中のカメラワークと、データベース中の各ショットにおけるカメラワークとを比較することで類似度を計算する。

入力動画 V_0 とデータベース中のショット V_k のカメラワークによる類似度 S_{camera} を、3.2 において判定したカメラワークの種類 C_0、C_k、および動きベクトルの大きさの平均 W_0、W_k を用いて、

$$S_{camera} = \begin{cases} \frac{1}{|W_0 - Q_k|} & (C_0 = C_k) \\ 0 & (C_0 \neq C_k) \end{cases} \tag{9}$$

により定義する。カメラワークの種類が一致する場合には、W が表すカメラの動きの激しさが同程度のものほど、S_{camera} は高くなるように定義されている。

4.3 映像中の人物数に基づく類似度

検出した顔の個数によって、映像中の人物数に関する類似度を計算する。

入力動画 V_0 とデータベース中のショット V_k の人物数による類似度 S_{person} を、3.3 における検出した顔数 f_0、f_k、およびそれらの顔がそれぞれ映像中に占める割合 $t_{0,i}(i = 1, \cdots, f_0)$、$t_{k,i}(i = 1, \cdots, f_k)$ を用いて、

$$S_{person} = \begin{cases} \frac{1}{|T_0 - T_k|} & (f_0 = f_k) \\ 0 & (f_0 \neq f_k) \end{cases} \tag{10}$$

と定義する。ただし、

$$T_0 = \sum_{i=1}^{f_0} t_{0,i} \quad , \quad T_k = \sum_{i=1}^{f_k} t_{k,i} \tag{11}$$

である。

4.4 音の種類に基づく類似度

類似度を音の種類とパワーによって決定する。

入力動画 V_0 とデータベース中のショット V_k の人物数による類似度 S_{sound} を、3.4 において判定した音の種類 s_0、s_k、および音の平均パワー E_0、E_k を用いて以下のように計算する。

if $s_0 \neq$ ”無音”

$$S_{sound} = \begin{cases} \frac{1}{|E_0 - E_k|} & (s_0 = s_k) \\ 0 & (s_0 \neq s_k) \end{cases} \tag{12}$$

if $s_0 =$ ”無音”

$$S_{sound} = \begin{cases} 1 & (s_0 = s_k) \\ 0 & (s_0 \neq s_k) \end{cases} \tag{13}$$

5　類似理由生成部

入力動画との類似度が高いショットを含む動画を提示すると同時に、それが類似であると判定した理由をテキスト形式で生成し、ユーザに提示する。また、色に基づいた類似理由や人物数に関する類似理由を提示する場合にはフレーム画像の表示、カメラワークに基づいた類似理由を提示する場合には動画の再生、の処理を行い、類似理由をより分かりやすくする。

類似理由の提示は、ユーザが選択した類似尺度に基づいて行う。

5.1　色に関する類似理由の生成例

色に関する類似理由生成の際には、以下の点に着目する。

- 色ベクトルの分散 δ
- 画面全体で最も出現回数の多い色 m_0
- 画面全体で二番目に出現回数の多い色 m_1
- 画面上部で最も出現回数の多い色 m_2
- 画面下部で最も出現回数の多い色 m_3
- 画面左部で最も出現回数の多い色 m_4
- 画面右部で最も出現回数の多い色 m_5

色ベクトルのコサイン測度の値が大きい場合、これらも類似する可能性が高い。なお、$m_i \in$ { 白、黒、灰色、茶色、赤、オレンジ、黄色、青、緑、ピンク、紫 }(i=0,1, … ,5) とする。

入力動画 V_0 の各値 δ_0、$m_{0,0}$〜$m_{5,0}$、提示する動画 V_k の δ_k、$m_{0,k}$〜$m_{5,k}$ を用いて、Fig.4 のように類似理由を生成する。なお、下線部には、m_i に対応した色名や、d_i に対応した位置 (d_2 = "上"、d_3 = "下"、d_4 = "左"、d_5 = "右") が入る。また、各条件を複数満たす場合には、それぞれのテキストを連結したものを提示する。

($\delta_0 < 0.4$, $\delta_k < 0.4$ のとき)

「どちらの動画もカラフルです。色々な色が含まれてますね。」

($m_{0,0} = m_{0,k}$, $\delta_0 > 0.7$, $\delta_k > 0.7$ のとき)

「$\underline{m_{0,k}}$ いです。一言で言うと $\underline{m_{0,k}}$ いんです。」

($m_{0,0} = m_{0,k}$, $m_{1\,0} \neq m_{1,k}$, $\delta_0 \leq 0.7$, $\delta_k \leq 0.7$ のとき)

「全体的に $\underline{m_{0,k}}$ っぽいです。」

($m_{j\,0} = m_{j,k} (2 \leq j \leq 5)$ のとき)

「$\underline{d_j}$ の方が $\underline{m_{0,k}}$ っぽいところとか。」

Fig. 4　色に関する類似理由例

5.2　カメラワークに関する類似理由の生成例

入力動画とカメラワークの種類が一致していると判定したショットには類似理由を付加する。類似理由は、4.2 で用いたカメラワークの種類 C_0、C_k に応じて、Fig.5 のように生成する。d はパンやチルトの際のカメラの移動方向であり、$d \in$ { 上, 下, 左, 右 } とする。

5.3　人物数に関する類似理由の生成例

入力動画と登場する人物数が一致していると判定したショットに対して、類似理由を付加する。類似理由は、4.3 における検出顔数 f_0、f_k を用いて Fig.6 のように生成する。下線部には、顔数 n に応じた整数を記入する。

5.4　音の種類に関する類似理由の生成例

入力動画と音の種類が一致していると判定したショットには、4.4 における音の種類 s_0、s_k を用いて、Fig.7 のように類似理

($C_0 = C_k =$ ”ズーム” のとき)

「どちらもズームのある動画です。ズームということは、何かに着目している動画です。」

($C_0 = C_k =$ ”パン” or ”チルト” 、方向が d のとき)

「カメラを d に動かしながら撮っているところとかが似ています。」

($C_0 = C_k =$ ”ハンドカメラ” のとき)

「映像がぶれているところとか似ています。ハンドカメラとかで撮ってるのでないかと … 。」

Fig. 5　カメラワークに関する類似理由例

由を付加する。

6　評価実験

試作したシステムに対して、実装した類似尺度と類似理由提示の有用性を調査した。

6.1　実験方法

12 人の大学生・院生を対象として評価実験を行った。最初に入力動画、続いてその動画を入力した場合の、各類似度に基づいてシステムが提示したショットを 4 本、そしていずれの類似度も低い動画 1 本の計 5 本を、どのような類似尺度で検索した結果であるかを伝えずに視聴してもらい、Fig.8 のアンケートに回答してもらった。各被験者に対して、それぞれ 3 本の入力動画を用いてこの実験を行った。また、アンケートには自由記述欄を設け、被験者の意見を募った。

質問 1 は、試作した提案システムにおいて用いた 4 種類の類似尺度が、ユーザ自身の持つ類似尺度とどの程度一致しているのかの評価を目的とした質問である。質問 2

($f_0 = 0$, $f_k = 0$ のとき)

「どちらの動画も、人が映っていない動画、もしくは人がメインではない動画です。」

($f_0 = f_k = n$ かつ $n > 0$ のとき)

「どちらも、人が n 人くらい映っている動画です。」

Fig. 6　人物数に関する類似理由例

($s_0 = s_k =$ ”音楽” のとき)

「音楽が流れている動画です。つまり、どちらにも歌とか BGM があります。」

($s_0 = s_k =$ ”音声” のとき)

「人の声がする動画というところが似ています。」

($s_0 = s_k =$ ”無音” のとき)

「どちらも無音の動画です。無音だと雰囲気も似てくると思いませんか？」

Fig. 7　音の種類に関する類似理由例

は、システムが提示する類似理由の精度の評価と、各類似尺度の必要性の評価を主な目的とした質問である。質問 3 は、システムが類似理由を提示することで、ユーザにどの程度の心理的影響を与えるかの評価を目的とした質問である。

6.2　結果と考察

各類似尺度で提示した動画に対して、質問 1 において「思う」と判断した割合を、Table 1 に示す。また、質問 2 において「納得できる」と回答した類似尺度別の割合を Table 2 に、質問 3 において「なった」と

質問 1:提示された動画は、入力動画と類似していると思いますか?

1. 思う
(類似と判断した理由:)

2. 思わない (質問 2 へ)

質問 2:システムが提示した類似理由に納得することはできましたか?

1. 納得できる (質問 3 へ)

2. 納得できない

3. 類似理由が理解できない

質問 3:類似理由を読んだことで、提示動画と入力動画は類似していると思えるようになりましたか?

1. なった

2. ならない

3. わからない

Fig. 8 評価実験アンケート

回答した割合を Table 3 に示す。

Table 1 から、被験者が類似していると判定した割合は、システムの類似尺度によって差があることが分かる。特にカメラワークに着目した場合に著しく低くなっている。これにより、動画が類似しているかどうかを判定する時に、カメラワークに着目する人はほとんどいないことが分かる。

しかし、Table 2 において、カメラワークに着目したことを記した類似理由に対しては、「納得した」という回答が 92%と、比較的高い値となった。また Table 3 において、類似していると思えるようになった、という回答が 3 件あり、Table 1 の 1 件と比べて類似していると判断した回答が増加している。これは類似理由を提示したことによって、ユーザの類似判断基準に何らかの心理的影響を与えたということを示している。特に、自由記述欄において類似理由に対して「そういう視点もあるのか」や「”なるほど”と思った」との記述があったことから、類似理由の提示によってユーザの気づかなかった類似尺度に気づかせるという効果を確認することができた。他の尺度に関しても、質問 3 に対して「なった」との回答が多くはないが存在し、同様の効果があったと考えられる。

色に着目した場合に関しては、Table 1 では高かったのに対して、Table 2 では一番低くなるという結果になった。色は、今回実装した 4 種類の類似尺度の中では、最も被験者間で差が大きくなった類似尺度であった。本論文では基本色の中で多く出現する色に着目したが、その基本色の分類は主観に依存するので、システムの行う分類は被験者の感性と異なっていた可能性もある。また、被験者によっては動画中の一部の領域内の色に着目したり、本論文では考慮していない明るさの違いを重く捉えたりと、一概に色と言っても着眼点は様々であることも判明した。改善のためには、個人差を吸収するための手法の検討が必要である。

音の種類に着目した場合には、質問 2 で「納得できない」と回答した被験者の中には、「歌と BGM は違う」等の自由記述があった。試作した提案システムでは、歌も BGM も音楽に分類され、区別していないためである。人数に関しても、顔の向きなどによって顔検出の精度が低下したことで、「実際の動画中の人数と、類似理由で書かれている人数が異なる」という声が複数挙がった。これらに関しても、より良い手法の検討や、精度の向上が必要である。

また、自由記述欄では他にも、「音は確かに似ているが、画像が似ていない」という記述があった。本実験では、それぞれ 1 つの類似尺度のみで検索を行っているが、複数の類似尺度を選択することが必要となると思われる。またその場合には、それぞれの類似尺度をどれだけ重視するかを考慮する必要がある。

Table 1　質問 1 の結果

色	カメラワーク	音の種類	人数
22.2 % (8/36)	2.8 % (1/36)	22.2 % (8/36)	13.9 % (5/36)

Table 2　質問 2 の結果

色	カメラワーク	音の種類	人数
68.2 % (15/22)	92.3 % (24/26)	70.8 % (17/24)	84.0 % (21/28)

Table 3　質問 3 の結果

色	カメラワーク	音の種類	人数
13.3 % (2/15)	12.5 % (3/24)	17.6 % (3/17)	4.8 % (1/21)

一方で、類似理由に納得した場合でも、「やはり自分の判断基準にはない」といった意見もあった。そのようなユーザの類似判定基準に影響を与えることは、現状のシステムでは難しいと思われる。しかし、中には「おもしろい」とシステムを評価する声が複数あった。本システムの改良により、エンターテインメント性の向上も期待できる。

7　関連研究

関連研究としては、本論文と同種の類似尺度を利用しているものとして、色の類似する動画を検索する手法[10]や、カメラワークの類似性に基づいて類似シーンを検索する手法[11]などがある。[10]では、ショット単位、及び関連性の強いショットの集合単位で HSV ヒストグラムを作成し、色の類似するシーンを検索する。[11]では、スポーツ映像において、類似シーンであればカメラワークも類似するという特徴を利用している。カメラワークから特徴ベクトル列を生成し、DP マッチングによって類似シーンを検索することで、野球のホームランシーンのみを抽出する、といったことが可能である。しかしこの 2 つの研究では、対象とする動画を特定の分野に限定する場合には高い精度での類似シーン検索が可能であるが、限定しない場合には精度は大幅に低下すると思われる。また、[10]、[11]は類似尺度を 1 種類しか用いておらず、本システムのように対象とする動画に応じて複数の類似尺度をユーザが切り替えて用いるという仕組みはない。

複数の類似尺度を用いているものとしては、色・テクスチャ・モーションから類似動画を検索する手法[1]がある。色ヒストグラム、テクスチャ、モーションヒストグラムの 3 種類の特徴量から類似度を計算することで、スポーツ映像や映画を対象にした検索で比較的良い結果が得られている。しかし、類似理由の提示機能がない点で本論文と異なる。

8　おわりに

本論文では、類似尺度をユーザが選択でき、その類似尺度に基づいて類似と判定した動画とその類似理由を提示する類似動画検索システムを試作し、評価した。評価の結果、実装した色の類似性、カメラワークの類似性、映像中の人物の類似性、音の種類の類似性に基づいた 4 種類の類似尺度は、いずれもユーザの類似判断基準となる可能性があることを確認した。さらに類似理由を提示することで、ユーザの類似の判定基準に影響を与えることができることを確認した。

また今後は、評価実験の結果をふまえ、4 種類の類似尺度に関する特徴量の改善や、類似尺度を同時に複数利用して検索を行うことが必要である。さらに、今後はシステ

ムで用いる類似尺度を増やすことを検討する。

参考文献

1) Yining Deng; Manjunath, B.S. *Content-based search of video using color, texture, and motion*, Image Processing, 1997. Proceedings., International Conference on Volume 2, pp.534-537, 1997.

2) Su, C.-W. Liao, H.-Y.M. Tyan,H.-R. Lin,C.-W. Chen,D.-Y. Fan, K.-C. *Motion Flow-Based Video Retrieval*, IEEE Transaction on Multimedia, Volume 9, Issue 6, pp.1193 - 1201, 2007.

3) 長坂晃朗、田中譲. カラービデオ映像における自動索引付け法と物体探索法, 情報処理学会論文誌, Vol.33, No.4, pp. 543-550, 1992.

4) E. L. van den Broek, P. M. F. Kisters, L. G. Vuurpijl. *The utilization of human color categorization for content-based image retrieval*, In B. E. Rogowitz and T. N. Pappas, editors, Proceedings of Human Vision and Electronic Imaging IX, pp. 351-362, volume 5292, 2004.

5) CHANG. Y, SAITO. S, NAKAJIMA. M. *A framework for transfer colors based on the basic color categories*, In Computer Graphics International, pp176-183, 2003.

6) Millet. C, Grefenstette. G, Bloch. I, Moellic. P.A, Hede. P, *Automatically populating an image ontology and semantic color filtering*, International Workshop, pp.34-39, 2006.

7) Lucas. B, Kanade. T, *An Iterative Image Registration Technique with an Application to Stereo Vision*, In. DARPA Image Understanding Workshop, pp.121-130, 1981.

8) Viola. P, Jones. M, *Rapid object detection using boosted cascade of simple features*, IEEE Conference on ComputerVisionandPatternRecognition,2001 pp.511-518 vol.1, 2001.

9) 高柳直，林貴宏，尾内理紀夫，ソナグラムの画像特徴に着目した音声・音楽・ノイズ区間識別手法の提案，電子情報通信学会技術研究報告, pp.17-22, 2007.

10) T. Lin, H.J. Zhang. *Automatic Video Scene Extraction by Shot Grouping*, Proceedings of the International Conference on Pattern Recognition, p.4039, September 03-08, 2000

11) 片岡 良治，遠藤 斉. MPEG 符号化情報に基づく類似シーン検出方式, 情報処理学会論文誌，Vol.41，No.SIG. 3(TOD6), pp.38-45, 2000

12) 青柳 滋己, 佐藤 孝治, 高田 敏弘, 菅原 俊治, 尾内 理紀夫. 映像短縮再生システムの教育映像への適用評価，情報処理学会論文誌，Vol.46，No.5，pp.1297–1305, 2005.