

色差情報を用いたカラー動画の高速類似検索

Similarity Retrieval of Color Videos Using Color Differences

森近 優人† 宝珍 輝尚† 野宮 浩揮†
Yuto Morichika Teruhisa Hochin Hiroki Nomiya

1. はじめに

近年、コンピュータ技術の発達は目覚ましく、様々な分野でコンピュータが用いられるようになってきている。またコンピュータの普及によって、今までアナログであったモノの多くがデジタルに変換されつつある。それは例えば写真であったり、例えば音楽であったりする。その中の一つに映像データがある。

映像データは、カメラやビデオだけにとどまらず携帯電話などからでも作成することができ、その手軽さや利便さから様々な動画データが生み出され、個々人が持つ映像データの数は日々増え続けている。個々人が持つ映像データと類似した映像を検索したいという要求があるが、映像データは画像データとは違い時系列データであるため、一目見ただけではどのような動画であることを確認することはできず、人間の目視で探し出すには手間と時間がかかってしまう。

柏野らは、長時間の音や映像の中から既知の音や映像の位置を高速に検索することを目的として、スペクトル特徴のヒストグラムに基づいて検索を行う手法を提案している[1]。動画データの各フレームを、輝度を用いて圧縮しそれを用いて作成したヒストグラムで検索を行うことで、漏れなく高速に検索することができるように行っている。

武らは、長時間の放送映像アーカイブから CM を効率的かつ効果的に検出・同定することを目的として、教師なし、一般的、超高速な新たな手法を提案している[2]。そこでは、各動画データに対して輝度ベースの画像ハッシュ法を用いている。

しかし、これらのように動画データの検索に対して、輝度のみを扱おうと、形の区別をつけることができるが、色の区別はつけることができない。

そこで、本論文では、動画検索において類似した動画を検索することを目的として、フルカラー動画から色差を用いて特徴量として抽出し、そこから動画のヒストグラムを作成し、比較を行うことによって既存の手法では検索できないものが検索可能になる、新たな検索手法の提案を試みる。そして、動画の検索で用いる特徴量と色差との関係について考察する。

以降、2. では関連研究について述べる。3. では主に用いる特徴量である色差について述べる。4. では色差の利用とヒストグラムについて述べる。5. で本論文で用いる検索手法について述べ、6. で類似検索の評価実験を行う。7. で考察を行い、最後に8. でまとめる。

2. 関連研究

2.1 時系列アクティブ探索法

柏野らは、動画データに対して、各フレームの輝度から抽出される特徴量を用いて作成されたヒストグラムを用いることで、長時間の動画データの中から既知の動画が含まれているかどうか、また含まれていた場合にはその位置について高速に検索する手法を提案している[1]。動画データにはテレビ放送 6 時間分の映像と、15 秒の CM をいくつか用いている。

各フレームを 12 の領域（縦*3, 横*4）に分け、各領域の輝度値から各領域を二値化し、それを用いて 1 フレームの特徴量とする。この特徴量からヒストグラムを作成し、既知の動画データから作成されたヒストグラムと長時間の動画データから作成されたヒストグラムの類似度から、長時間の動画データの次に検索する位置を決定することによって検索時間の短縮を行っている。

この方法によって、検索時の所要時間は全探索と比較して 112 倍にまで向上している。

2.2 二段時間的共起ハッシュ

武子萌らは、膨大な映像アーカイブに対して、特定の CM を効率的かつ効果的に検出・同定する手法を提案している[2]。動画データには 10 時間と 1 か月間の放送映像を用いており、その動画内にはそれぞれ 66 個と 202 個の CM が含まれている。

各フレームを 16 の領域（縦*4, 横*4）に分け、各領域から輝度の相加平均と標準偏差を算出する。算出された 32 次元の特徴ベクトルを 2 値化し、それをもとにハッシュテーブルに格納する。同一のビンに連続で格納されたフレームの塊を映像断片とする。ここで離れた同一のビンに入れられた複数の映像断片は一致しているものとする。ビンが異なった映像断片同士の時間的な一貫性を検出することによって同一映像と雑音とを見分ける。

この方法による 1 か月間のアーカイブに対する処理時間は 42 分であり、検出精度（CM であるか否か）は 98.1%、位置特定精度（CM の出現時間）は 97.4%であった。

3. 色空間 YUV と色差 U, V

色空間 YUV とは、色空間の 1 つであり、輝度 Y と色差 U, V を用いることで色を表現する。この色空間の利点は、人間の目は、明るさには敏感であるが、色の変化には鈍感であるという性質があることを利用して、明るさの情報は残したまま色の情報を削減することによって、人間の目からは画質が劣化することなく、データ量を削減することができるという点である。主にテレビの放送映像

や JPEG, MPEG など用いられている. Y のみを用いると, モノクロ画像 (映像) として表現することが可能である. 色差 U, V は輝度 Y から色空間 RGB の青成分 B と赤成分 R を引いた値に定数を掛けたものである. 色空間 RGB から色空間 YUV への変換の式を式(1)に示す.

$$\begin{aligned} Y &= 0.299 * R + 0.587 * G + 0.114 * B \\ U &= -0.1687 * R - 0.3313 * G + 0.5000B + 128 \\ V &= 0.5000 * R - 0.4187 * G - 0.0813B + 128 \end{aligned} \quad (1)$$

Y, U, V の取りうる範囲は式 (1) から分かる通り, $0 \leq Y, U, V \leq 255$ である. また, U, V が共に 128 の時がモノクロの色を表しており, Y が大きければ白に近づき, 小さければ黒に近づく.

色空間 YUV にはいくつかの種類があり, YUV444 や YUV422, YUV420, YUV411 などがある. これらは圧縮率が異なっており, 数字が小さい程色差 U, V が削減されている. 例えば, YUV420 では, 輝度 Y は全ピクセルに割り当てられているが, 色差 U と V は $2 * 2$ 四方の 4 ピクセル毎に 1 つの値が割り当てられている.

4. 色差の利用

本論文では, 動画の各フレームからの特徴量を取得する際に, 色差成分を用いる. ここでの色差とは, 色空間 YUV における色差 U と色差 V を指し示すものである. 使用する色空間 YUV の種類は YUV420 とする.

本論文では動画データのフレーム毎に色差成分から特徴量を取得し, その特徴量から動画データに対してのヒストグラムを作成する. また, 色差成分から特徴量を取得するためにもヒストグラムを使用する. まず, 動画データの 1 フレームに対して U, V の値をそのままビンとしてヒストグラムを生成し, ヒストグラムのビンをいくつかに分類することでヒストグラムのビン数を削減する. ビンの分類にはフェヒナーの法則[3]を適用し, 分類後のビン数を B とすると U, V の中央値を 128 とすることで $0 \sim 127$ と $128 \sim 255$ をそれぞれ $B/2$ 種に分類する. 色差を特徴量に変換する流れを図 1 に示す. 図 1 ではビン数が 8 の場合を示している. また, 式 (2) に元の U の値から分類されたビンに変換する式を示す. 式 (2) での U は YUV の U の値, H_u は変換後のヒストグラム内のビン, n は底である.

$$X_u = \begin{cases} 128 - U & (U < 128) \\ U - 128 & (U \geq 128) \end{cases}$$

$$H_u = \begin{cases} \text{floor}((B/2) + (B/2) \log_n X_u / \log_n 128) & U < 128 \\ \text{floor}((B/2) - (B/2) \log_n X_u / \log_n 128) & U \geq 128 \end{cases} \quad (2)$$

この方法によって得られたヒストグラムから 1 フレーム毎の特徴量を作成し, それを用いて動画データのヒストグラムを作成する. 作成されたヒストグラム間の比較には時系列アクティブ探索法[1]を用いる.

5. 検索手法

本論文の以降では, 簡略化のために, 検索対象となる

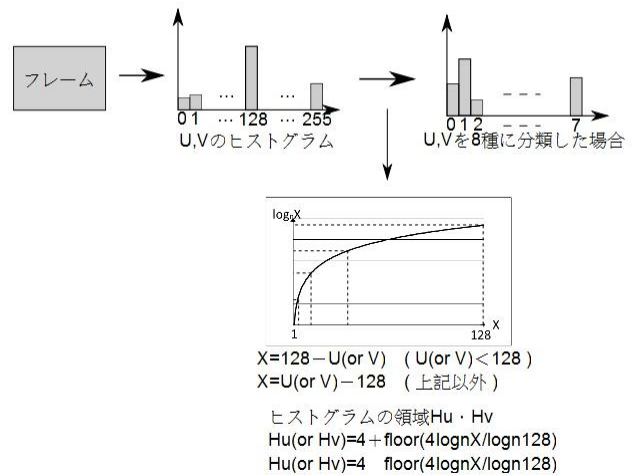
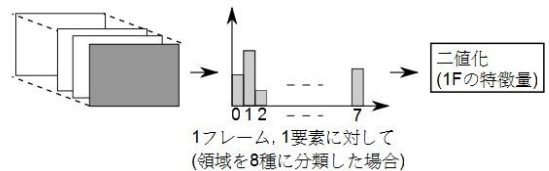
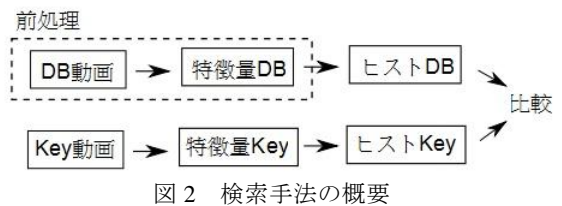


図 1 フレームから特徴量抽出の概要

大量の動画データ群を **DB 動画** (DataBase 動画), 類似したものを求めたい動画データを **Key 動画** と表すこととする. また, 要素とは Y, U, V を表し, 値とはヒストグラムのビンに入る量 (ヒストグラムの縦軸) を表す.

図 2 に検索手法の概要を示し, 図 3 に要素ごとの特徴量抽出の概要を示す. 図 3 ではビン数を 8 と仮定したものである.



検索の手順を以下に示す.

- ・ (前処理)
- a) DB 動画の各動画からの色差ヒストグラムの取得
各動画の要素 U, V を用いて 4. で示した色差ヒストグラムを作成し, 各ビンに対して一定以上の値 (閾値) を取るものを検出する. そのビンに対応する値を 1 とし, 他のビンの値を 0 とする (2 値化).
- b) a) で 2 値化されたデータの変換
要素 U, V に対してそれぞれ 2 値化されたデータが存在するが, ここでは 2 つの値をまとめて 1 つの特徴量に変換する.
- c) 特徴量をファイルに保存
b) で得られた特徴量を 1 動画毎に 1 ファイルを用いて作成する. DB 動画の特徴量のファイルに変換したものの集合を **特徴量 DB** とする.
- ・ (検索)

- d) Key 動画からの色差ヒストグラムの取得
a), b), c)と同様にして, Key 動画から特徴量を抽出したファイルを作成する. このファイルを**特徴量 Key**とする.
- e) 時系列アクティブ探索法 (1)
特徴量 Key からヒストグラムを作成する (**ヒスト Key**). 同様にして, 特徴量 DB の始点から特徴量 Key と同じ長さのヒストグラムを作成する (**ヒスト DB**).
(ここでの始点は最初の検索の場合は特徴量 DB の最初の値である.)
- f) 時系列アクティブ探索法 (2)
ヒスト Key とヒスト DB の類似度に基づいて検索を行う. 類似度とはヒスト Key とヒスト DB との重なり具合を指す.
- g) 時系列アクティブ探索法 (3)
f)で得た類似度によって特徴量 DB の始点の位置を検索漏れが出ない最大の幅で移動させ, e)の操作へ戻る. この手順を特徴量 DB の最後まで行う.

6. 評価

6.1 実験方法

本研究は, フリーのビデオ素材集である[4]~[11]内の全動画 460 個 (81 分 3 秒) に対して類似検索を行った. 全ての動画データは 30fps である.

動画の大きさはそれぞれ, 320*240, もしくは 640*480, 一部が 640*486 であるが, 特徴量抽出の際には 80*60 に縮小することとした. これは, 精密性を求める検索が目的ではなく, データのサイズを縮小することも目的とするからである.

探索するべき動画データ (Key 動画) と, それに類似した動画データ (以下, **正解データ**) はそれぞれ[4]~[11]内から選出する. 正解データの選出は筆者の一人が目視で行った. ここでの類似とは, 画面全体の雰囲気 (色合い) が似ているものを指す.

Key 動画には, [4]~[11]から 1 個ずつランダムに取得した (表 1). Key 動画のサムネイル一覧を図 4 に示す. DB 動画は, 素材集毎のいくつかの動画データを結合したものとする. この操作から DB 動画は合計で 9 個になった (表 2).

定数については, 4. の $B=8, n=8$, 5. の a)での閾値を 150 (総ピクセル $(80*60/4) / \text{ビン数 } B (8)$) とする. 検索時の閾値は, 0.6 とする. これは Key 動画と DB 動画の特徴量が Key 動画の 0.6 倍以上一致している部分を検索対象とすることを意味する.

処理時間については, CPU : 3.20Ghz, メモリ 2.00GB の PC で測定を行った. 測定には `gettimeofday` 関数を用いた.

6.2 実験結果

DB 動画から特徴量を抽出するのに要した時間を表 3 に示し, それぞれの Key 動画に対する検索時間と検索結果を表 4 に示す. 表 4 の結果から得られた DB 動画のフレームのサムネイル一覧を図 5 に示す. なお, 表 4 で類似度が 1.0 の動画は Key 動画と同一の動画データであったので図 5 では省略している.

表 1 Key 動画

ファイル名	再生時間	サイズ[pixel]
k01.mov[4]	0m 10s	320*240
k02.mov[5]	0m 12s	320*240
k03.mov[6]	0m 06s	320*240
k04.mov[7]	0m 15s	640*480
k05.mov[8]	0m 15s	640*480
k06.mov[9]	0m 15s	640*480
k07.mov[10]	0m 15s	640*480
k08.mov[11]	0m 15s	640*480

表 2 DB 動画

ファイル名	再生時間	サイズ[pixel]
d01.mov[4][5][6]	26m 34s	320*240
d02.mov[7]	10m 45s	640*480
d03.mov[8]	3m 59s	640*480
d04.mov[8]	6m 17s	640*486
d05.mov[9]	10m 52s	640*480
d06.mov[10]	11m 08s	640*480
d07.mov[11]	10m 29s	640*480

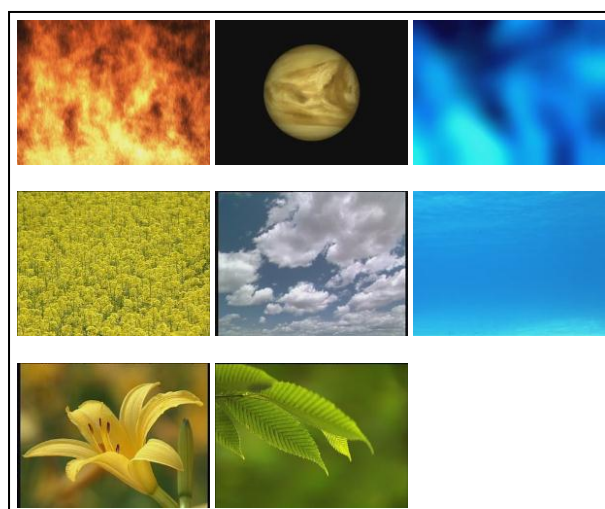


図 4 Key 動画のサムネイル
(左上から右の順で k01.mov, k02.mov, k03.mov, k04.mov, k05.mov, k06.mov, k07.mov, k08.mov)

表 3 DB 動画変換時間

ファイル名	変換時間[sec]
d01.mov	124.44
d02.mov	145.11
d03.mov	66.97
d04.mov	100.64
d05.mov	113.40
d06.mov	157.54
d07.mov	294.90

表 4 Key 動画の検索結果

ファイル名	検索時間[s]	類似動画の最初のフレーム [f](類似度)
k01.mov	0.95	d01.mov : 7520. (1.00000)
k02.mov	0.79	d01.mov : 22026. (1.00000) d01.mov : 33258. (0.96667)
k03.mov	0.89	d01.mov : 44086. (1.00000)
k04.mov	0.48	d02.mov : 4200. (1.00000)
k05.mov	0.56	d03.mov : 0. (1.00000) d04.mov : 7393. (0.64000)
k06.mov	0.84	d05.mov : 582. (0.70667) d05.mov : 7710. (1.00000)
k07.mov	0.58	d02.mov : 10200. (0.71778) d06.mov : 9180. (1.00000)
k08.mov	0.44	d07.mov : 2280. (1.00000)

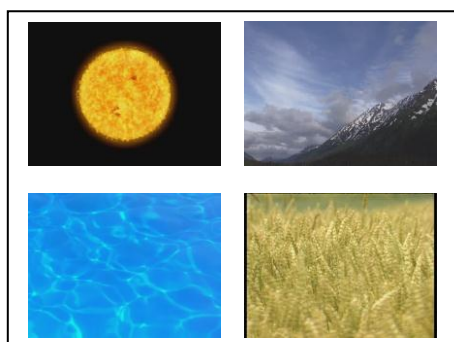


図 5 検索結果

(左から右の順で

d01.mov : 33258, d04.mov : 7393,
d05mov : 582, d02.mov : 10200)

7. 考察

7.1 検索速度

関連研究[1]では特徴量抽出のために要する時間が実時間の約 3% であった。本論文では、動画の大きさが 320*240 であれば実時間の約 8%、640*480 であれば実時間の 20% であった。しかし、[1]では特徴量をメモリ上に格納し、提案手法ではファイルに書き出しているのが概に提案手法が劣っているとは限らない。

同様に、検索に要する時間は[1]では 6 時間 15 秒に対して 0.84 秒であり、本論文では 81 分 31 秒に対し最長で 0.95 秒、最短で 0.44 秒であった。しかし、[1]では特徴量をメモリ上に格納してから測定しているのに対し、提案手法ではファイルからの読み取りの時間も加えているので、概に提案手法が劣っているとは限らない。

7.2 検索精度

関連研究[1]や[2]では、k02.mov に対しての検索で図 6 のような動画データを一致した動画データとして検索してしまうが、提案手法では図 5 の左上のみしか検出されなかった。関連研究[1]、[2]では輝度のみ使用しているので明るい部分が色に関わらず検出されてしまうが、提案手

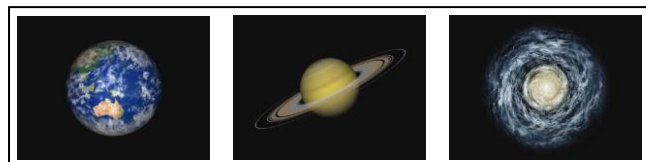


図 6 参考文献[1][2]での誤検出の例

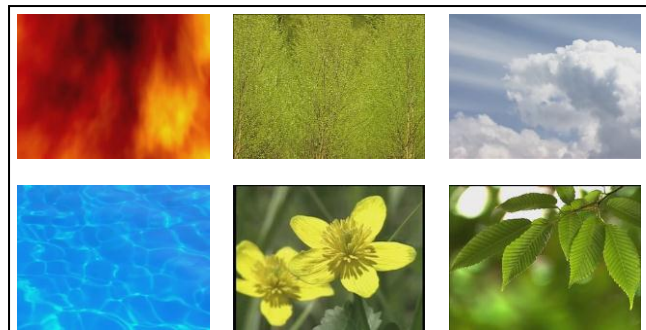


図 7 Key 動画の類似動画の一部

法では色差成分を使用しているので明確に区別することができた。これは提案手法の利点である。

DB 動画内には図 7 のような動画が含まれていたが、検索結果からは類似した動画としては検出されなかった。これは、U と V から特徴量を生成する方法とヒストグラムの比較の方法が原因であると考えられる。この点に関しては改良の余地があると考えられる。

8. おわりに

動画検索において類似した動画を検索することを目的として、新たな特徴量抽出法を提案した。フルカラー動画の色差から生成されるヒストグラムを用いて 1 フレーム毎に特徴量を作成し、その特徴量から生成されたヒストグラムを用いて比較を行うことによって既存の手法では検索できないものが検索可能になる、新たな検索手法の提案した。特徴量についての考察と、検索結果と検索時間についての考察も行なった。色差 U、V を使用して特徴量を取得し、新たな特徴量を用いて時系列アクティブ探索法を用いた結果、既存の輝度のみを用いた動画検索では誤検出してしまった動画を取り除くことに成功した。

今後は、人間が目視で確認したときに、より類似している動画を検出可能にするために、計算過程での定数の考察を行う。また、色差から生成される特徴量の生成方法についての見直しも行う。

参考文献

- 1) 柏野邦夫, 黒住隆行, 村瀬洋: ヒストグラム特徴を用いた音や映像の高速 AND/OR 検索, 電子情報通信学会論文誌, Vol.J83-D-II, No.12, pp.2735-2744, 2000.
- 2) 武小萌, 佐藤真一: 超高速 CM 検出に関する研究とその知識発見への応用, 電子情報通信学会 信学技報,

vol.111, pp.119-124, 2011.

- 3) 北川高嗣, 中西崇文, 清水康: 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, Vol.43, No.SIG 12, pp.38-51, 2002.
- 4) データクラフト, ”ビデオ素材辞典 Vol.1 CG - 光・炎・爆発”, 2002.
- 5) データクラフト, ”ビデオ素材辞典 Vol.2 CG - 宇宙・地球・惑星”, 2002.
- 6) データクラフト, ”ビデオ素材辞典 Vol.3 CG - ノイズ・合成パーツ”, 2002.
- 7) データクラフト, ”ビデオ素材辞典 Vol.4 四季・自然”, 2002.
- 8) データクラフト, ”ビデオ素材辞典 Vol.6 空と雲の情景”, 2002.
- 9) データクラフト, ”ビデオ素材辞典 Vol.7 水と波の表情”, 2002.
- 10) データクラフト, ”ビデオ素材辞典 Vol.8 四季の花”, 2002.
- 11) データクラフト, ”ビデオ素材辞典 Vol.9 爽やかな風景”, 2002.