

スポーツ中継映像データベースのカメラワーク情報による特徴解析手法

Sports Video Database Analyzing Method Using Camerawork Information

高木 真一[▼] 服部 しのぶ[◆]
 横山 和正 小舘 亮之[▲]
 富永 英義

Shin'ichi TAKAGI Shinobu HATTORI
 Kazumasa YOKOYAMA Akihisa KODATE
 Hideyoshi TOMINAGA

本稿では、映像より抽出したメタデータであるカメラワーク情報を統計的に解析することで、映像の特徴解析を行う手法について検討する。対象映像をテレビ放送のスポーツ中継番組とし、カメラワーク発生確率とカメラワークトランジションというメタデータ解析のための新しいパラメータを提案する。このパラメータを利用することで、競技種類ごとに統計的性質に有意な差が生じ、また色情報の特性が大きく異なる複数の映像でも、同一競技であれば、同一の特性が得られることを実験によって示し、カメラワーク情報が映像特徴解析のためのメタデータとして有効性があることを示す。

In this paper, we propose a new video metadata analyzing method based on camerawork information. We introduce 2 new features: Camerawork Extraction Ratio and Camerawork Transition. In a grounding experiment on numbers of live broadcast sports programs recorded in a video database, different characteristics were obtained for each sport statistically, which proved our system's effectiveness and applicability to automatic video genre classification systems with a large video database.

1. はじめに

マルチメディアコンテンツに関する内容情報(メタデータ)の記述方式であるMPEG-7[1]が標準化され、数多くの記述子が規定されたが、記述されたメタデータの有用な利用方法については標準化後の重要な検討課題となっている[2]。また、MPEG-7の記述ツールは、その多くを人手で与えるようなテキスト情報を利用したハイレベルな記述ツールであり、必ずしもコンテンツの特徴を明示的な語句によって表現できるとは限らない。そのため、コンテンツのストリーム信号から抽出したローレベルな記述データを検索などのアプリケーションで直接的に利用すること以外に、テキスト情報では表現で

きないような特徴情報も、ローレベルな記述データ群を解析することで定義・表現することは重要な研究課題と言える[4]。そこで本稿では、MPEG-7 Visual Part において規定されているローレベルなメタデータ記述ツールの一つであるカメラモーション記述子(Camera Motion Descriptor)に注目する。テレビ放送スポーツ中継番組を対象とし、カメラワーク情報を時系列に発生するデータとしてとらえ、統計的な解析を試みることで、各種スポーツ競技映像群からスポーツ競技ごとに有意に異なる特性が分析できることを実験によって示す。さらに、色情報の特性が大きく異なる複数の映像でも、同一競技であれば、映像メタデータ解析によって同一の特性が得られることを実験によって示し、カメラワーク情報が映像特徴解析手法のためのメタデータとして、有効性があることを示す。

2. スポーツ中継映像とカメラワーク情報

スポーツ中継放送映像は、カメラ設置位置の制約により、画像内に映る選手やグラウンド上のオブジェクトの位置などに制約がある。従って特定のスポーツを特徴付ける代表的な画像が存在しそれをもとにシーンを構造化する手法が検討されている[5]。しかし代表的な画像は、その構図や色情報は試合の開催会場や試合の開催時間帯・開催時の天候等などの環境により大きく変化する。一方スポーツ競技は、競技の種類ごとに一定のルールが存在し、試合はルールに従って展開されるため、撮影される映像のカメラワーク情報は、ある程度パターン化されていると考えられる。このパターンは、ルールの違いから発生するものであり、同一競技であれば、試合の開催会場、開催時間帯、開催時の天候等の撮影環境には依存しにくい。このMPEG-7によるカメラワーク情報の記述データを利用することの利点の一つとしてスポーツの競技判定が考えられるが、映像中の特定のカメラワークがあるところは、ハイライトシーンの候補である傾向が強く、一般個人撮影の旅行記録映像などに対しても編集支援環境が構築できるといった様々な応用が考えられる[3][6]。

3. カメラワーク情報の解析パラメータ

ここではカメラワーク情報に関する2つの映像解析パラメータを提案する。カメラワーク抽出方法としては、本稿では対象映像フォーマットをMPEGとし、動きベクトル情報をベースとしたカメラワーク抽出手法[7]を適用した。本方式では映像ストリームのGOPごとに、カメラ固定(FIX)、パニング(PAN)、ズーム(ZOOM)の他に、手ぶれなどの動きの非常に激しい状態であるSHAKEをカメラワークとして抽出できる。また、いずれのGOPのカメラワークもFIX、PAN、ZOOM、SHAKEの4種類のいずれかに判定される方式となっている。

カメラワーク発生確率

本検討で適用しているカメラワーク抽出法では、MPEGストリーム中のGOPを単位として、時系列順にカメラワーク情報の抽出を行っている。そのため、映像ストリーム中のGOP個数分だけカメラワークが発生する。よって各映像中におけるカメラワーク $x \in \{FIX, PAN, ZOOM, SHAKE\}$ の発生確率 w_x の算出式を以下のように定義する。

$$w_x = \frac{\text{映像ストリーム中の}x\text{の発生個数}}{\text{映像ストリーム中の全GOP数}} \times 100(\%)$$

▼ 学生会員 早稲田大学大学院国際情報通信研究科
takagi@tom.comm.waseda.ac.jp

◆ 早稲田大学理工学部電子・情報通信学科

sally.yokoyama,tominaga@tom.comm.waseda.ac.jp

▲ 早稲田大学国際情報通信研究センター

akikod@giti.waseda.ac.jp

カメラワーク・トランジション

新たな概念としてカメラワーク・トランジションを提案する。これは、カメラワークの種類だけ状態が存在し、時間の経過によって変化していくカメラワークの状態遷移を示すものである。その様子を図1.に示す。本稿では、ストリームの先頭GOPから順に抽出されるカメラワークXにおいて、あるn番目のGOPで抽出されるカメラワークX_nとn+1番目のGOPで判定されるカメラワークX_{n+1}が異なれば、カメラワークの状態遷移が発生する。n番目とn+1番目でカメラワークに変化がなければ遷移は発生しない。FIX, PAN, ZOOM, SHAKEの4種類の状態があるので、FIXからPANへの遷移、FIXからZOOMへの遷移、PANからFIXへの遷移など合計12種類の遷移が発生する。本検討ではカメラワーク・トランジションに関する最初の検討であるので、遷移の方向性までは考慮に入れず、カメラワークAからBへの遷移とカメラワークBからAへの遷移を区別せずに、両者をカメラワークAB間の遷移とみなす。カメラワーク・トランジションの遷移発生確率T_{AB}の算出式は以下のように定義できる。

$$T_{AB} = \frac{\text{カメラワークAB間の遷移発生回数}}{\text{映像ストリーム中の全遷移発生回数}} \times 100(\%)$$

A, B ∈ x

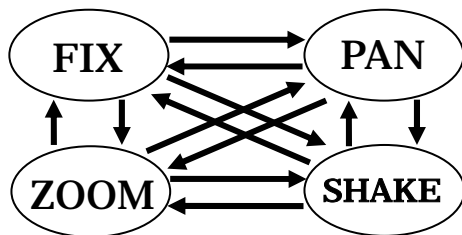


図1 カメラワーク・トランジション
Fig.1 Camerawork Transition

4. 実験に使用する映像データセット

本稿では、相撲、テニス、野球、サッカー、アメリカンフットボール、バスケットボールの6競技9種類のスポーツ映像を対象とした。2002年に放送された各競技15試合ずつのスポーツ中継放送番組に対し、試合開始から15分間の映像をMPEG1に符号化して用いた。但し、大相撲に関しては幕内後半最初の取組から15分間を用いた。そのうち、大相撲、テニス、野球に関しては、異なる大会やリーグの映像をそれぞれ準備した。具体的には大相撲に関しては名古屋場所(NGY)、秋場所(AKI)、テニスに関しては全英オープン(UKO)と全仏オープン(FRO)、野球に関しては日本のプロ野球(NPB)と米国のメジャーリーグ(MLB)を対象とした。サッカーは2002年ワールドカップ大会(WCS)、アメリカンフットボールは米国NFL(NFL)、バスケットボールは米国NBA(NBA)を対象とした。また全てのスポーツ中継番組映像にはCMは含まれてない。なお、MPEG符号化パラメータは、ビットレート1.5Mbps、フレームレート30fps、GOP構造N=15・M=3、画像サイズ352画素×240画素である。

5. カメラワーク情報の統計的解析手法

5.1 カメラワーク発生確率

競技ごとにFIX, PAN, ZOOM, SHAKEのそれぞれのカメラワーク発生確率W_xが算出されるが、放送番組のカメラ

操作の最も基本であるFIXの発生確率W_{FIX}に関するデータを表1.に示す。NGYやAKI等の記号は、各スポーツを表す記号である。ここでは、平均値 \bar{x} 、標本分散値 s^2 、歪度 a_3 、尖度 a_4 を計算する。その算出式は以下のように表される。

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad s^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1} \quad a_3 = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{Ns^3}$$

$$a_4 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{Ns^4} \quad \text{但し } s = \sqrt{s^2}$$

表1. 各競技のFIX発生確率に関するデータ
Table 1 Probability of Camerawork "FIX"

	\bar{x}	s^2	a_3	a_4
NGY	83.991	12.401	-0.732	3.775
AKI	83.640	14.133	-0.818	2.799
UKO	69.712	51.158	-0.143	2.097
FRO	65.980	23.449	-0.021	2.146
NPB	70.092	21.630	-0.480	2.214
MLB	71.478	21.664	0.316	2.037
WCS	30.377	11.569	-0.634	2.782
NFL	43.996	24.635	0.364	2.158
NBA	49.442	27.454	0.330	2.274

本検討では、各競技(大会・リーグ)15試合ずつのデータを使用したので、N=15でありx_iは各試合のFIXの発生確率W_xを示す。表1より、競技の種類によってFIX発生確率の平均値に明確な差が生じていることがわかる。狭い土俵を距離が離れたところから撮影している相撲はFIX発生確率が高く、広いグラウンドのボールを常に追跡しながらカメラ操作をしているサッカーはFIX発生確率が低くなっているなど、競技の性質による撮影方法違いがカメラワーク発生確率に反映されている。また、歪度 a_3 、尖度 a_4 を算出した。歪度は、分布の対称性を示すもので、値が0に近いほど左右対称の発生分布と判断でき、尖度は、分布のスジの広がりを示し、正規分布に近いほど値が3に近くなる特性を持つ。若干値に開きがあるものの、おおよそ±1.000以下の誤差に収まっており、比較的正規分布に近い値を示している。以下の解析はカメラワークの発生確率が正規分布に従っていると仮定する。

5.2 競技間の統計的差分の解析

表1.では、競技ごとに数値に違いが発生しているように見えるが、競技間でカメラワークの発生率に統計的に有意な差があるかを検討する必要がある。ここでは一元配置の分散分析を行うことによって各競技のカメラワーク発生確率の母平均の差の検定を行う。具体的には全競技のデータを同時に比較することのできる方法として一般的な多重比較を用いたテューキーの方法による検定を行う。

各データ群の母分散の一様性の検定

テューキーの方法による検定を行うには、各データ群の等分散性が前提となるので、ここではまず、9個の各スポーツ種類群の母分散の一様性の検定をハートレイ検定によって証明する。まず以下の仮説と対立仮説をたてる。

仮説H₀:各スポーツ競技間の母分散は互いに等しい。

対立仮説H₁:少なくとも一つの母分散は異なる。

表1より検定統計量F_{max}は以下のように表せる。

$$F_{\max} = \frac{\max S_i^2}{\min S_i^2} = \frac{51.158}{11.569} = 4.422$$

有意水準を 5% とすると、各競技の繰返し数は 15、スポーツ競技数は 9 であるので、 $F_{\max} = 4.422 < F_{(9,15)}(0.05) = 5.38$ となり、仮説 H_0 を棄却することができない。よって表 1. の等分散性が示される。

各データ群のチューキーの方法による多重比較

まず、検定する競技(大会・リーグごと)の総数を a とし、各競技を区別するために 1 から整数の番号をふる。ここでは、 $NGY=1, AKI=2, UKO=3, \dots, NBA=9(a=9)$ となる。また、競技ごとの実験に使用したデータセットの映像の個数を n_i とし、競技 i の j 番目の映像データからのカメラパラメータ発生確率を x_{ij} と表現する。全体のデータの変動(全変動) S_T 、競技間のデータの変動(水準間変動) S_A 、競技内のデータの変動(水準内変動) S_E は、次のように表せる。

$$S_T = \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij}^2 - \frac{(\sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij})^2}{N}$$

$$S_A = \frac{(\sum_{j=1}^{n_1} x_{1j})^2}{n_1} + \frac{(\sum_{j=1}^{n_2} x_{2j})^2}{n_2} + \dots + \frac{(\sum_{j=1}^{n_a} x_{aj})^2}{n_a} - \frac{(\sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij})^2}{N}$$

$$S_E = S_T - S_A$$

水準間変動と水準内変動の自由度はそれぞれ $a-1, a(n-1)$ であるので、水準内変動の平均平方 V_E は次のように表せる。

$$V_E = S_E / a(n-1)$$

そこで、スチューデント化された範囲の上側 $100\alpha\%$ $q(a, a(n-1); \alpha)$ を統計数表より求め、競技 i_1 と競技 i_2 のカメラパラメータ発生確率の平均値を $\bar{x}_{i_1}, \bar{x}_{i_2}$ とすると、

$$|\bar{x}_{i_1} - \bar{x}_{i_2}| \geq q(a, a(n-1); \alpha) \sqrt{\frac{V_E}{n}}$$

ならば、2 つの競技 i_1 と競技 i_2 間の母平均に有意な差があるということになる。本稿では、一般的な値である $\alpha=0.05$ とし、有意水準 5% で検定を行った表 1 と統計数表より S_T と S_A を計算し、 V_E を求めると、

$$q(a, a(n-1); \alpha) \sqrt{V_E/n} \approx 5.587 \dots \dots \dots (1)$$

を得る。

表 3 に表 1 より計算した 2 つの競技 i_1 と競技 i_2 間のカメラワーク発生確率の平均値の差を示す。各欄の絶対値を式(1)の結果と比較し、式(1)より大きければ 2 つの競技間の母平均に有意な差があることを示す。この結果より、大相撲名古屋場所と大相撲秋場所間、全英オープンテニスと全仏オープンテニス間、プロ野球と米メジャーリーグ間の同一競技内では母平均データに有意な差がないと判定されていることがわかる。つまり、カメラワーク情報の統計的解析によるデータ区別の結果とスポーツ競技の種類が一致していることがわかる。しかし、米アメフトと米バスケットなどの一部の異なる競技間などで、データに有意な差がないと判定されている。

5.3 カメラワークトランジションを利用した解析

表 2. に、カメラワーク・トランジションにおける FIX と PAN の間の遷移発生確率を示す。ここでは 5.2 において米アメフトと米バスケット間で、カメラワークの FIX 発生確率に差がないため、データに有意な差がないとされてしまっているが、

それらがカメラワーク・トランジションを利用したデータ解析をすることで、両者を統計的に区別できることを示す。そのためには、等分散性があり、母平均に有意な差が生じていればよい。

表 2 カメラワークトランジションに関する実験データ
Table 2 Transition Ratio for "FIX-PAN"

	\bar{x}	s^2	a_3	a_4
NFL	14.802	27.674	0.682	2.432
NBA	44.760	51.636	0.607	3.407

カメラワークトランジションの等分散性

まず、両者の母分散 $\sigma_{NFL}^2, \sigma_{NBA}^2$ に関して有意水準 0.05 で母分散の検定を行う。まず次の仮説・対立仮説を立てる。

仮説 $H_0: \sigma_{NFL}^2 = \sigma_{NBA}^2$ 対立仮説 $H_1: \sigma_{NFL}^2 \neq \sigma_{NBA}^2$

NFL と NBA の標本分散を s_{NFL}^2, s_{NBA}^2 とすると、検定統計量 $T(s_{NFL}^2, s_{NBA}^2)$ は、以下のように表せ、自由度 $(n_{NFL}-1, n_{NBA}-1)$ の F 分布に従う。

$$\frac{s_{NFL}^2}{s_{NBA}^2}$$

表 2 と統計数表より、

$$T(s_{NFL}^2, s_{NBA}^2) = 0.536 \quad F_{14,14}(0.975) = 0.339$$

$$F_{14,14}(0.025) = 2.949$$

を得られ、棄却域 R に対し、 $T(s_{NFL}^2, s_{NBA}^2) \notin R$ であるので、仮説を棄却できず、両者の母分散に有意な差がないことが確認できる。

カメラワークトランジションの母平均の検定

両者の母平均 μ_{NFL}, μ_{NBA} の差の検定を有意水準 0.05 で行う。

仮説 $H_0: \mu_{NFL} = \mu_{NBA}$ 対立仮説 $H_1: \mu_{NFL} \neq \mu_{NBA}$

母平均の差の統計検定量 $T(\bar{x}_{NFL}, \bar{x}_{NBA}, s_{NFL}^2, s_{NBA}^2)$ は、以下のように表せ、自由度 m の t 分布に従う。

$$T = \frac{\bar{x}_{NFL} - \bar{x}_{NBA}}{\sqrt{\frac{s_{NFL}^2}{n_{NFL}} + \frac{s_{NBA}^2}{n_{NBA}}}}$$

$$m = \frac{(\frac{s_{NFL}^2}{n_{NFL}} + \frac{s_{NBA}^2}{n_{NBA}})^2}{\frac{s_{NFL}^4}{n_{NFL}^2(n_{NFL}-1)} + \frac{s_{NBA}^4}{n_{NBA}^2(n_{NBA}-1)}}$$

で近似する。m が整数でない時は、その最も近い整数を m と定める。表 2 と統計数表より、

$$T = -13.028 \quad m \approx 26 \quad t_{26}(0.025) = 2.056$$

であるので、

$$|T| \geq t_{26}(0.025)$$

となり仮説 H_0 が棄却され、両者の母平均に有意な差があることが確認できた。つまりこの米アメフトと米バスケットの映像では、FIX のカメラワーク発生確率だけでは、ほぼ同じ特性になっているが、カメラワークの状態遷移に関して差異があることを示している。以上により、カメラワーク発生確率だけでは特性に差が生じなくても、カメラワーク・トランジションを利用してデータ解析することによって、より詳細な特性

まで分析できることが確認できた。

5.4 色情報の異なる複数同一競技試合の解析

ここでは、カメラパラメータ情報に関して、スポーツ競技が同一であれば試合の会場の違いや試合開催の時間帯・天候等に影響されることなく、安定してその競技の特性が解析できることを示す。競技の開催会場による色情報が大きく異なる例として、テニスを取り上げて検証をした。テニスの場合、全英オープンテニスは、緑色の芝コートで行われるのに対して、全仏オープンテニスは、赤土のクレコートで行われるため、両者の主要色は大きく異なる。その様子を図2に示す。FIX カメラワーク発生確率を、6.3 におけるカメラワーク・トランジションの場合と同じように、母分散と母平均の差を有意水準 0.05 で検定を行った。すると、母分散・母平均共に

仮説： $\sigma_A^2 = \sigma_B^2$ 仮説： $\mu_A = \mu_B$

を棄却不可能となり、分析した両者のカメラワーク情報に有意な差がないことが確認できる。つまり会場や時間帯等の撮影環境の違いがスポーツ中継番組映像のカメラワーク情報の統計的解析結果に影響していない。



図2： 競技開催場所による主要色の異なる場合
左：(緑色の芝コート) 右：(赤土のクレコート)
Fig. 2 Difference of Dominant Color Information

6. まとめと今後の課題

本稿では映像より抽出したメタデータであるカメラワーク情報を統計的に解析することで映像の特徴解析を行う手法を検討した。2 つ新しいパラメータを提案し、このパラメータをスポーツ中継映像に適用し映像の特徴解析に関する有効性の検証を行い、カメラワーク情報が映像メタデータ解析手法のための情報として、有効性があることが確認できた。スポーツ中継以外の映像での実験やFIX 以外に関するパラメータに関する検討などが今後の課題である。

【文献】

[1] ISO/IEC 15938 International Standard,"MPEG-7: Multimedia Content Description Interface," 2002.
[2] 堀修,"MPEG-7 の概要と役割,"オーディオビジュアル複

合情報処理シンポジウム 2001~MPEG-7 最新動向と将来展望~,pp.3-10,Sep., 2001
[3] 遠藤 齊, 片岡良治, "カメラモーションに基づく類似動画画像検索," 信学技報, DE99-55, pp.147-152, Jul., 1999.
[4] 馬場口登,"メディア理解による映像メディアの構造化," 信学技報, IE99-18 PRMU99-42 MVE99-38(1999-07), pp.39-46, Jul., 1999.
[5] 杉山善明, 渋谷亜紀, 有木康雄,"DCT特徴に基づくTVスポーツ映像の自動判別,"信学技報, IE97-28, PRMU97-59, MVE97-44, pp33-38, 1997-07
[6] 金田瑞規, 土橋健太郎, 高木真一, 小館亮之, 富永英義, "映像編集支援のためのインデキシングパラメータ抽出手法の検討," 信学技報, IE2001-128, pp.13-18, 2001-12
[7] 土橋健太郎, 小館亮之, 西塔隆二, 富永英義,"手ぶれを考慮した MPEG2 からのカメラワーク検出の検討," 信学技報, CS2000-110 IE2000-120, pp.1-6,2000-12.

高木 真一 Shin'ichi TAKAGI

早稲田大学大学院国際情報通信研究科博士後期課程在学中。1999 同大学大学院理工学研究科修士課程修了。画像処理に関する研究・開発に従事。電子情報通信学会学生会員。日本データベース学会学生会員。

服部 しのぶ Shinobu HATTORI

早稲田大学大学院理工学研究科修士課程在学中。2002 早稲田大学理工学部電子・情報通信学科卒業。映像のメタデータ応用に関する研究・開発に従事。電子情報通信学会学生会員。

横山 和正 Kazumasa YOKOYAMA

早稲田大学大学院理工学研究科修士課程在学中。2002 早稲田大学理工学部電子・情報通信学科卒業。映像のメタデータ応用に関する研究・開発に従事。電子情報通信学会学生会員。

小館 亮之 Akihisa KODATE

早稲田大学国際情報通信研究センター客員助教授。1997 早稲田大学大学院理工学研究科博士後期課程修了,工学博士。主に画像処理・コンテンツ利用に関する研究・開発に従事。情報処理学会会員。電子情報通信学会会員。

富永 英義 Hideyoshi TOMINAGA

早稲田大学理工学部電子・情報通信学科教授。1964 早稲田大学大学院理工学研究科修士課程修了。工学博士。情報通信網・マルチメディアシステムなどの研究・開発に従事。情報処理学会会員。画像処理学会会員。電子情報通信学会副会長。

表3 各競技の一元配置分散分析(多重比較)による結果

Table 3 Multiple Comparison

	AKI	UKO	FRO	NPB	MLB	WCS	NFL	NBA
NGY	0.351	14.279	18.011	13.899	12.513	53.615	39.995	34.549
AKI	---	13.928	17.660	13.548	12.162	53.263	39.643	34.198
UKO	---	---	3.732	0.380	1.766	39.335	25.715	20.27
FRO	---	---	---	-4.112	-5.498	35.603	21.983	16.538
NPB	---	---	---	---	-1.386	39.715	26.095	20.650
MLB	---	---	---	---	---	41.101	27.481	22.036
WCS	---	---	---	---	---	---	13.620	-19.065
NFL	---	---	---	---	---	---	---	-5.445