

サッカー放送映像のダイジェスト化に向けたキーフレーム検出

Key Frame Detection from Broadcasted Soccer Video for Making a Digest

田村 徹†
Tohru Tamura

陳 曉琴†
Xiaoqin Chen

1. まえがき

近年、デジタル技術の発展に伴って膨大な映像データを蓄積、閲覧できるようになった。一方で映像データの利用者は自身が必要とする映像データを効率よく探し出すことが困難となっている。その対策として、映像データを整理・分類する研究が行われている。それらの研究では、映像をショットに分割し、ショットごとに代表的なフレームを抽出するなど物理的構造に基づく方法、映像内の物理的特徴量やオブジェクトに注目し、意味的なタグを付加する方法などがあり、一定の成果が期待されている¹⁾。さらに、映像情報に加え、音声信号、音響信号や視聴者の行動履歴など、複数の異なるメディアから得られるデータを利用する研究も行われている²⁾³⁾。これらの研究は映像情報を構造化することで、膨大な映像情報を効率的に利用することを主な目的としている。

ところで、映像記録メディアの大容量化、低価格化が進み、家庭でテレビ番組を大量に録画し、視聴者は好みの時間帯に録画した番組を視聴することができるようになった。しかし、大量に撮りためた録画映像をなかなか視聴する時間がないといった状況が考えられる。そこで、いったん録画した映像を短時間に要約するための研究がおこなわれている⁴⁾⁵⁾。それらの研究では、ドラマ映像や映画を対象として、物理的な構成要素、トラック構造と心理的な要素を利用した要約映像の生成手法を提案している。一方で、テレビで多く放送されるスポーツ映像を対象とした研究には、アメリカンフットボールの得点シーンをハイライトシーンとして検出する手法⁶⁾やサッカーのコナーキック、フリーキックやスローインなど特定のシーンを検出する手法に関する研究がある⁷⁾⁸⁾。しかし、いずれもスポーツ放送における限られたシーンの検出に主眼を置いており、試合の流れを維持しつつ視聴者が興味をもつ重要なシーンをあつめ要約するための手法というよりは、映像がもつ内容を分類・整理し、特定のシーンを効率よく検索・利用するための手法と考えられる。

そこで、本研究ではサッカー放送を録画した映像の重要な場面を集めてダイジェスト映像化する方法について検討した結果を報告する。本研究では、サッカー放送の映像データをそれが意味する内容ごとに分割・分類し、たとえば得点場面といった特定のシーンを効率よく検索・利用することを目的とするのではなく、録画された映像データをもとに視聴者

が試合内容を理解し、興味あるシーンを含んで十分に楽しめるダイジェスト映像を作成することを目的とした研究である。サッカー放送を選んだ理由としては、サッカーは世界的に普及し、かつ人気のあるスポーツであり数多くの試合が放送されること、試合時間が決まっており試合の最初から最後までほぼ編集なしに放送されることから、ダイジェスト化された映像ソースの利用価値が高いと考えたためである。

録画されたTV放送をもとに、重要な場面を集めたダイジェスト映像を作成する方法として、映像データからフレーム画像を順次取り出し、そのフレーム画像が重要な場面に踏まれるフレーム画像か否かを判定する。フレーム画像が重要な場面に含まれると判断された場合には、そのフレーム画像を含む前後のフレームを一定時間取り出し、つなぎ合わせて

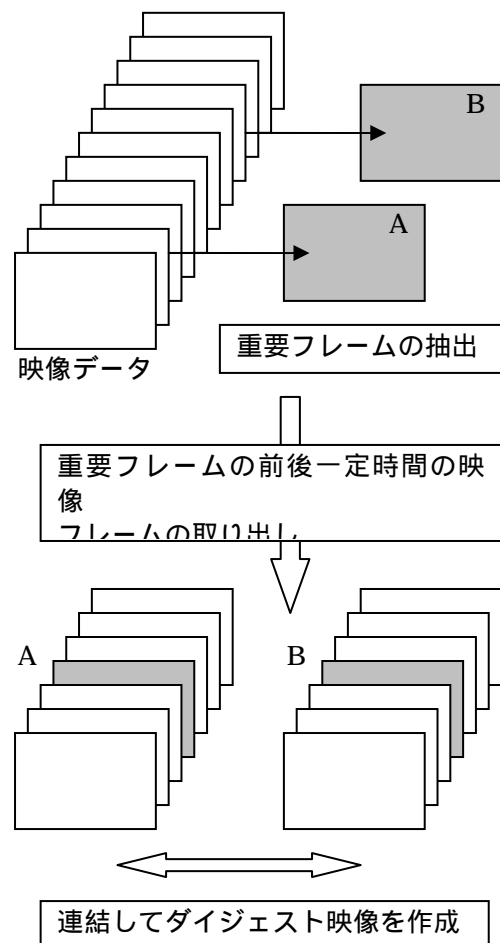


図1 ダイジェスト映像の作成手順

† 東京工芸大学大学院工学研究科

ダイジェスト映像化することを考えている(図1)。本報告では、図1のうち、映像データから重要場面に含まれるフレームを検出する手法について報告する。

報告では、まず、サッカー映像において重要と思われる場面の決め、重要な場面を代表するフレームを持つ共通の特徴について述べる。次に、重要なフレームを持つ共通な特徴検出法と検出実験の結果について述べ、最後に、まとめと今後の課題について述べる。

2. 重要な場面とフレーム特徴

サッカー映像をダイジェスト化するにあたり、まずサッカー放送の中でどのような場面が重要となるかを検討する。サッカーは得点を競うスポーツであることを考えると、得点に関係するシーンは最も重要と考えられる。しかし、サッカーの試合をダイジェスト化するためには、得点シーンだけでは不十分であることが予想される。そこで、サッカーの試合を録画した映像から、得点シーンのみを編集した映像を用意し、11名の被験者に編集した映像を視聴させた。その後、編集した映像に対して、ダイジェスト映像として見たいシーンについてコメントを収集



図2 重要な場面の例。
上：得点シーン、下：選手交代

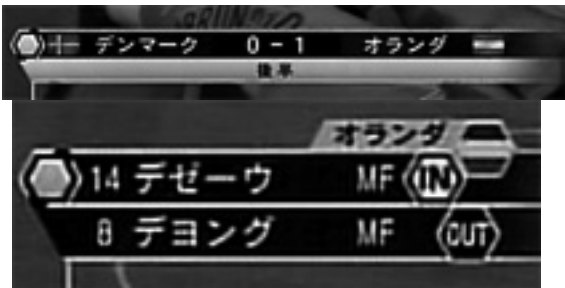


図3 重要な場面にインポーズされた部分の拡大

したところ、「試合の開始、終了の場面」「選手がシュートした場面」「選手の紹介」「フリーキックやコーナーキック」「選手交代」「反則」などがダイジェスト映像に必要な重要な場面であることがわかった。

次に、上述した重要な場面に通ずるフレーム画像の特徴について録画映像をもとに調べたところ、重要な場面の多くで画面下部に得点や選手の背番号などを示す数字がインポーズされることがわかった。実際の例を図2に示す。図2の例は上が得点シーン、下が選手交代のシーンであり、いずれも画面の下部にインポーズが挿入されていることがわかる。図3に重要な場面にインポーズされた部分を拡大して示す。得点シーンでは両チームの得点が数字で表示され、選手交代のシーンでは選手の背番号が数字で表示されている。そこで、本研究では重要な場面を示すフレームの特徴を「画面の下部に数字が存在する」とし、この条件をみたすフレームを検出することとした。実際にダイジェスト映像を作成する際には、検出したフレームの前後のフレームを切り出してつなぎ合わせることで、ダイジェスト化を行うことができると考えている。

3. フレーム内特徴の検出

3.1 特徴検出法

本研究の目的は、録画された映像ソースから重要な場面を検出してつなぎ合わせ、ダイジェスト化した映像を生成することである。そのためには、重要な場面に含まれるフレームがもつ共通の特徴を精度よく検出する必要がある。ここでは、前章で述べたフレーム内特徴を検出する方法について述べる。

前章で示したように、検出対象となる共通な特徴は数字であり、ある程度の大きさの違いはあるものの、画面に対し垂直に正対している。また、対象となる数字は活字であり、手書き文字のように複雑な形状をしていない。以上の点をふまえ、本報告では数字の検出法としてテンプレートマッチング法を用いることとした⁹⁾。テンプレートは数字の0から9に対応する10種類を用意した。テンプレートは、

$$NCC(a, b) =$$

$$\frac{\sum_{i=0}^{H_t-1} \sum_{j=0}^{W_t-1} |I(a+i, b+j)T(i, j)|}{\sqrt{\sum_{i=0}^{H_t-1} \sum_{j=0}^{W_t-1} I(a+i, b+j)^2 \times \sum_{i=0}^{H_t-1} \sum_{j=0}^{W_t-1} T(i, j)^2}}$$

$I(a, b)$: Source Image, $T(i, j)$: Template Image

H_t : Template Image Height

W_t : Template Image Width

... (1)

縦約 30 画素、横約 20 画素の 2 値画像で背景が黒、数字が白である。検出すべき数字は、ある程度の範囲で大きさがことなる場合がある。そこで、異なる大きさの数字の検出にはテンプレートの大きさを拡大・縮小することで対応した。拡大・縮小はテンプレートの縦横比を保ったまま、比率を 0.6 から 2.0 まで 0.2 刻みで変えた。また、今回検出すべき数字には、暗い背景に明るく表示される場合と、明るい背景に暗く表示される場合とがある。ここでは、フレーム画像の濃淡を反転させることで両極性の数字検出を行った。また、類似度の評価としては、式 (1) に示した正規化相互相関を用いて行った。

図 4 にフレーム内特徴の検出法を示す。まず、数字検出を行うフレーム画像を読み込む。次に前処理として数字の探索領域の決定、鮮鋭化、ノイズ除去を行った。まず、読み込んだフレーム画像を分割し、画像の上半分と画像両端を画像幅の 10% 分の幅で除いた領域を探索領域とした。これは、画像の上部や両サイドは検出すべき数字が表示されることのない領域であるためである。次に、フレーム画像のカラー情報を利用して、緑色の領域と肌色の領域を探索領域から除いた。緑色のフィールドと選手の顔や肌を探索領域なら除くことで、誤検出の可能性を低減することが狙いである。残った領域にラベリング処理を行い、ソース画像面積の 0.5% より小さな領域を消去した。これは、検出すべき数字はソース画像中である程度以上大きな領域の内部に存在しているため、孤立した小さな領域を除くことで、誤検出の可能性を低減することが期待されるためである。図 5 にフレーム画像から上半分、両端の幅 10% の領域、

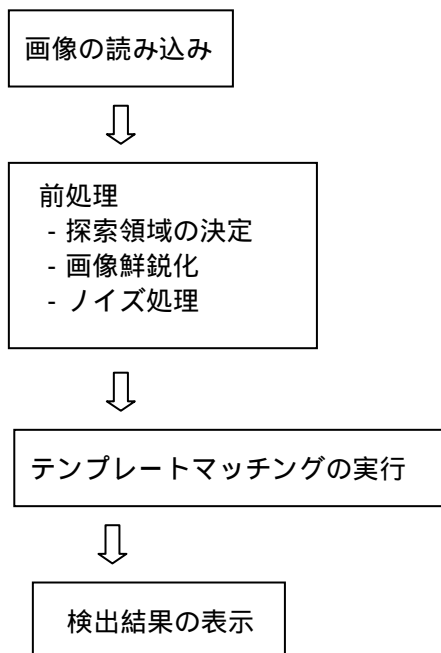


図 4 数字検出処理の流れ

グリーン領域を消去して残った探索領域の例を示す。図 5 の下の画像の白で示した部分に対応するフレーム画像の領域が数字を探索する領域となる。図 5 の下の画像にラベリング処理をして、面積が小さい部分を消去することで、グラウンドの広範囲を俯瞰したような場面の探索領域をほぼなくすることができる。この処理によって、選手の体の一部を数字と誤認識する可能性を低減することができる。また、図 2 の下をフレーム画像として上半分、両端の幅、グリーン領域を消去した画像を図 6 に示す。図 6 を見ると、ラベリング処理後に消去する領域の大きさを適当に設定することで、検出すべき数字が存在する領域は消去せずに、誤検出の原因となる選手の領域を消去できることがわかる。本報告では、フレーム画像面積の 0.5% 以下の領域を消去することとした。

フレーム画像は上述した処理を施した後に 8 ビットグレースケール画像に変換し、鮮鋭化とノイズ処

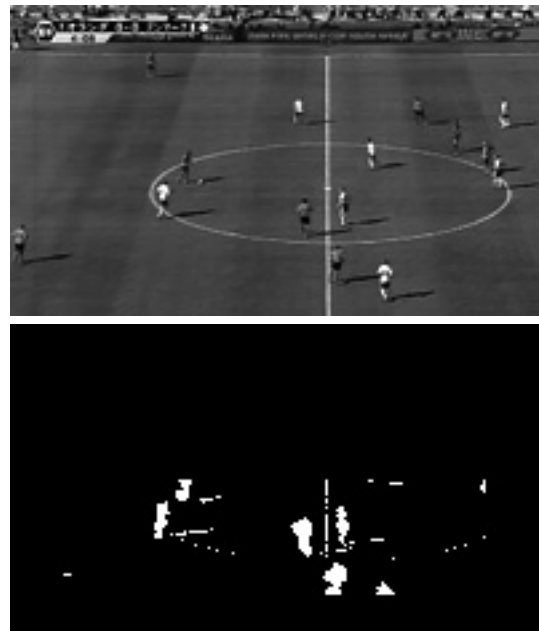


図 5 探索領域の決定。
上：フレーム画像、下：探索領域

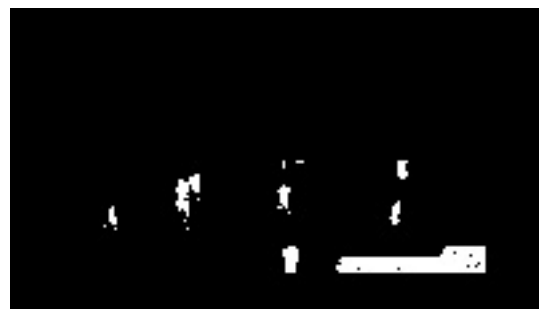


図 6 図 2 の下をフレーム画像とした場合の探索領域。

理を行った。鮮鋭化には 3×3 の8近傍高域強調フィルタを用いた。ノイズ処理は収縮・膨張処理によって行った。

3.2 フレーム画像内の数字検出実験

ここでは、録画された実際のサッカー放送から取り出したフレーム画像を用いて行った数字の検出実験について述べる。

(1) 実験画像

実験に使用した画像は異なるテレビ局が放送した2試合のサッカー放送から選んだ。試合ごとに、検出すべき数字を含んだ画像(key画像)を20枚。検出すべき数字を含まない画像(notkey画像)を40枚の計120枚の画像によって実験を行った。

図7に実験に用いたkey画像の例を示す。上下は異なる試合のメンバー紹介のシーンである。この画像は重要な場面として検出すべき数字が画像内に存在しており、key画像の例となる。key画像には、



図7 key画像の例
上：試合1、下：試合2

表1 実験結果

	未検出画像数	誤検出画像数	誤答率
試合1	0	3	0.05
試合2	1	6	0.12
計	1	9	0.08

表2 実験結果

	再現率	適合率
試合1	20/20(100%)	20/23(87%)
試合2	19/20(95%)	19/25(76%)
計	39/40(98%)	39/48(81%)

試合開始、終了、選手交代、フリーキック、反則、選手紹介などダイジェスト映像の要素として重要と考えられる場面が入っている。一方、notkey画像には、選手のアップ、選手や観客のミドルショット、フィールドを広く俯瞰するようなロングショットなど異なる構図から選んだ。図8にnotkey画像の例を示す。図8の上は選手のアップ、中がミドルショット、下がロングショットの例である。いずれも試合1からとった画像で、試合開始前の場面からの例である。

(2) 検出結果

前節で述べた検出手法をもちいて、検出実験をおこなった。実験では、key画像(20枚×2試合分)とnokey画像(40枚×2試合分)の検出を行った。結果を表1にまとめる。key画像のうち、検出すべき数字を検出できなかったkey画像の枚数を未検出画像数とした。試合1では20枚のkey画像すべてで数字を検出することができ、未検出画像数は0枚であった。試合2では1枚について検出に失敗した。



図8 notkey画像の例
上：アップ、中：ミドル、下：ロング

ただし、画像中の検出すべき数字すべてが正しく検出されたわけではなく、複数の検出されるべき数字があった場合、そのうち少なくとも1つが正しく検出された場合は正検出と判断した。画像中に検出すべき数字がない notkey 画像について、検出すべき数字ではない対象を数字としてあやまって検出した場合を誤検出として、その画像数を誤検出画像数とした。試合1では誤検出画像数が40枚中3枚、試合2では6枚であった。未検出画像と誤検出画像を合わせて誤答とし、その割合を誤答率とした。誤答率は全体で0.08であった。表2には表1に示した結果を再現率 = 正検出数/正答数、適合率 = 正検出数/検出数によって示した。

再現率は試合1が100%、試合2で95%であり、重要と思われる場面の検出はほぼ成功した。一方、適合率は試合1で87%、試合2では76%となり、ダイジェスト映像中に不要な場面が含まれることとなる。全体としての適合率81%がダイジェスト映像作成上どの程度問題となるかについては、実際にダイジェスト映像を作成し、そのダイジェスト映像を



図9 誤検出の例1

評価する必要がある、今後の課題と考えている。

(3) 誤検出画像の検討

適合率を向上し、ダイジェスト映像中に重要ではないシーンが含まれることを防ぐためには、誤検出画像を解析することが必要である。ここでは、誤検出画像を解析し、フレーム画像中のどのような部分が検出対象の「数字」と誤って認識されたのかを明らかにする。

誤検出にはフレーム画像中の数字ではない部分を数字と誤検出する場合と、検出した数字が検出すべき数字ではない場合の2つのパターンに大別される。図9に数字ではない部分を数字と誤認識した例を示す。図9上部のフレーム画像ではゴールネットが数字と誤って認識され、中央のフレーム画像では観客が来ていた服の模様を数字と誤検出していた。下図の左はゴールネットを数字として検出した部分の拡大図、下図の右に観客の来ていた服の模様を数字として検出した部分の拡大図をしめす。四角で囲んだ部分が検出部を示している。このような例は全部で7枚あり、ほかには選手のユニホームの模様や影を数字と誤検出する場面が見られた。図10はフレーム画像中の検出すべき数字ではない数字を検出した例を示す。上図が選手のユニホームの数字を誤検出したフレーム画像であり、下図に検出部分を拡大した図を示す。このように検出すべき数字ではない数字を誤って検出した例は他にもう1枚あり、観客の服の数字を検出していた。

このような誤検出を減らすことによって適合率の向上を図る必要がある。その方法としては、フレーム画像中にインポーズされ、検出すべき数字が含



図10 誤検出の例2

まれる領域をあらかじめ分離し、その領域内で数字の検出を行う方法が考えられる¹⁰⁾。また、検出すべき数字は連続する複数のフレーム内で同じ位置に表示されるのに対し、図9、図10に示した誤検出部分は連続するフレーム内の表示位置が変化する。この特性を利用し、複数のフレーム内の検出結果を統合利用して、誤検出を減らすことなどが考えられる。これらの手法と組み合わせ適合率の向上を図ることが今後の課題と考えている。

4. まとめと今後の課題

本研究ではサッカー放送を録画した映像をもとに重要なシーンを抜き出し、それらをつなぎ合わせることで視聴者が楽しむことができるダイジェスト映像を作成することを目的としている。ダイジェスト映像を作成する場合、視聴者が見たいシーンがダイジェスト映像に含まれていなければならない。そこで、まずサッカー放送の中で重要なシーンは何かについて調べ、重要なシーンに含まれるフレーム画像がもつ画像内の共通な特徴を見出した。次に、その特徴であるフレーム画像の下部に表示される数字の検出法について検討し、検出実験を行った結果、再現率98%、適合率81%といった結果を得た。

今後の課題としては、実際にダイジェスト映像を作成し、視聴者による評価を行うことが必要であると考えている。また、適合率を向上させるために、重要なフレームの検出法の改良を図ることも必要と考えている。

参考文献

1. Y. Li、田中讓、 “メタデータの管理に基づくビデオデータベースの構成”、情処学論、vol.39、no.4、pp.1137-1145、(1998)
2. 長谷山美紀、 “メディア横断型理論の構築と検索への応用”、電子情報通信学会基礎・境界サイエティ大会講演集、SS-68-69
3. Makoto YAMAMOTO, Miki HASEYAMA, “An Accurate Scene Segmentation Method Based on Graph Analysis Using Object Matching and Audio Feature”, IEICE TRANS, FUNDAMENTALS, Vol.E-A, No8 (2009)
4. 森山剛、坂内正夫、 “ドラマ映像の心理的内容に基づいた要約映像の生成”、電子情報通信学会論文誌、Vol.J84-D- No.6、pp.1122-1131、(2001)
5. 石井孝和、吉高淳夫、平川正人、市川忠男、 “映画の文法に基づくビデオ画像の内容検索”、情処学研報、97-DBS-111、pp.65-72、(1997)
6. 宮内進吾、馬場口登、北橋忠宏、 “テキスト・音声・画像の協調的処理による放送型スポーツ映像におけるハイライト検出とインデキシング”、電子情報通信学会論文誌、Vol.J85-D- No.11、pp.1692-1700、(2002)
7. 丸尾二郎、岩井儀雄、谷内田正彦、越後富夫、飯作俊一、 “サッカー映像からの特定イベントの抽出”、信学技報、PRMU99-41 (1999)
8. 三須俊枝、苗村昌秀、藤井真人、八木伸行、 “選手フォーメーション解析に基づくサッカーイベント判別法”、映像情報メディア学会誌、Vol.61、No.9、pp.1367-1375、(2007)
9. 田村秀行、 “コンピュータ画像処理 第8章 画像認識の手法”、オーム社
10. 新井啓之、桑野秀豪、倉掛正治、杉村利明、 “映像中のテロップ表示フレーム検出方法”、電子情報通信学会論文誌、Vol.J83-D- No.6、pp.1477-1486、(2000)