

サッカー放送映像のダイジェスト化に向けたショット境界の検出法 A Method of Shot Boundary Detection from Broadcasted Soccer Video for Making a Digest

田村 徹† 大崎義貴† 飛永由夏† 森山 剛†
Tohru Tamura Yoshitaka Osaki Yuka Tobinaga Tsuyoshi Moriyama

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。しかし、一方では目的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、代表的なフレームを抽出して映像データを分類する方法、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、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ゆっくりと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⁾やサッカーのコーナーキック、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、筆者らは、録画されたサッカーの試合映像から重要なシーンを検出して、つなぎ合わせたダイジェスト映像の作成手法について検討している。この際、重要なシーンを検出することが必要であり、重要シーンに含まれる特徴として画面内に現れるテロップの検出法について検討した^{9) 10)}。また、重要シーンを自然につなぎ合わせるためには、ショット境界を検出し、ショットの切り替わりに合わせて重要なシーンを繋ぎあわせる必要がある。

そこで、本研究では、サッカー放送を録画した映

像のショット境界の検出方法について検討する。サッカー放送を選んだ理由は、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含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ショット境界で一定時間取り出し、つなぎ合わせてダイジェスト映像化することを考え

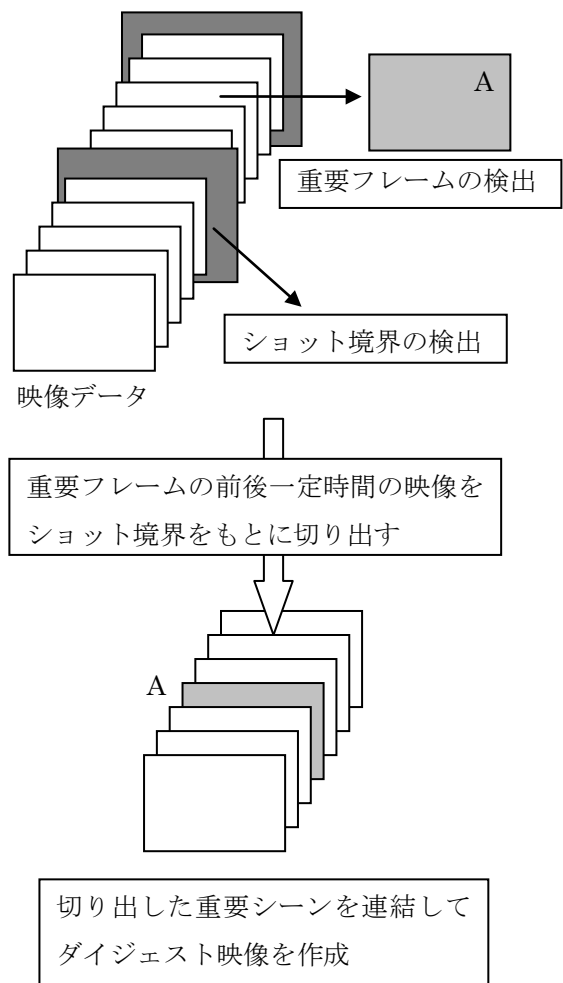


図1 ダイジェスト映像の作成手順

† 東京工芸大学大学院工学研究科

ている (図 1)。本報告では、図 1 のうち、ショット境界の検出方法について報告する。

2. ショット境界の検出法

サッカー映像のショット境界を検出する方法として、フレーム画像の色ヒストグラムを用いる方法が Ahmet らによって提案されている¹¹⁾。そこでは、2 枚のフレーム画像の色ヒストグラムを求め、その差をもとに 2 枚のフレーム画像の間にショット境界があるか否かを評価している。その際、フレーム内の緑色の領域の大小によってショット境界の判定に用いる色ヒストグラムの類似性の閾値を変えている。こうすることでフィールドを俯瞰するような緑の多いシーンと、たとえば、観客席など緑の少ないシーンによって異なる判定基準を用いたショット境界の検出法を提案している。また、新田らは、フレーム画面を 16 の小領域に分割し、2 枚のフレーム間で隣接するブロック間の色ヒストグラムの差分和をもとにアメリカンフットボールの試合映像のショット領域の検出方法を提案している¹²⁾。Ahmet らの方法の場合、シーンの大まかな内容、フィールドを映しているのか、それ以外の場所を映しているのか、を区別することで、ショットの検出精度の向上を図っている。しかし、この方法は、フレーム内の選手の動きやカメラのパンやズームといった動きを考慮していない。一方、新井らの手法の場合は、フレームを小領域に分割し、それぞれ隣接する領域との色ヒストグラムを比較することで、フレーム内の選手の動きやカメラアクションへの対応が図られている。



図 2 類似した背景で別ショットの例

しかし、フレームを小領域に分割し、すべての小領域ごとにその周辺小領域との間で色ヒストグラムの類似度を計算する必要があり、計算コストの面で検討の余地がある。また、フレームの分割方法は固定されており、フレーム内の選手の大きさや位置と小領域が必ずしも一致するとは限らない。

そこで、本研究ではフレーム全体の色ヒストグラムの類似性と、フレーム内でもっとも大きな領域の色を除いた残りの色ヒストグラムの類似性の両者を利用したショット境界の検出法を提案する。たとえば、図 2 のような場合では、緑のフィールドが多くを占め、フィールド全体の色ヒストグラムは高い類似性を持つ。一方で、フレーム内でもっとも大きな領域、図 2 の場合にはもっとも大きな領域は緑色でフレーム画像の背景を構成している、を除いた残りの色ヒストグラムの類似度は低い値となる。図 2 の場合は上下のフレーム間にショット境界があり、フレーム全体の色ヒストグラムともっとも大きな領域を占める色を除いた色ヒストグラムの類似度を用いることで、ショット境界の検出を正しく行うことができる。同一色で大きな領域は、フレーム画像の背景を構成する場合が多く、フレーム内を背景と前景に分離することができ、選手がフレーム内で移動した場合とショットが切り替わって異なる選手を映した場合を区別でき、カメラのパンやズームといった動きにも対応することができる。また、フレーム全体の色ヒストグラムを一度計算するだけで、2 つの類似度を簡単に求めることができ、計算コストも少なく済むといったメリットが考えられる。

3. 色ヒストグラムと類似性

3.1 色ヒストグラムの計算法

ここでは、フレーム画像から色ヒストグラムを求める方法について述べる。まず、フレーム画像の各画素値 (R, G, B) から、その画素の色相値 (H) に以下の式を用いて変換する。

$$H = \begin{cases} 60 * \frac{G - B}{\max(R, G, B) - \min(R, G, B)} + 60 & \text{where } \max(R, G, B) = R \\ 60 * \frac{B - R}{\max(R, G, B) - \min(R, G, B)} + 180 & \text{where } \max(R, G, B) = G \\ 60 * \frac{R - G}{\max(R, G, B) - \min(R, G, B)} + 300 & \text{where } \max(R, G, B) = B \end{cases}$$

変換した H の値域を 16 等分して 16 色の色相とした。フレーム画像の各画素から H の値を計算し、色相ごとに度数を求め、最大度数を 255 に正規化した。図 3 にフレーム画像とその色ヒストグラムの計算結果を示す。図 3 の上がフレーム画像であり、図 3 の下がフレーム画像から計算した色ヒストグラムである。

3.2 色ヒストグラムの類似性評価

2枚のフレーム画像の色ヒストグラムの類似度が高い場合には2枚のフレーム画像は同一ショット、すなわち、その2枚のフレームの間にショット境界はない。類似度が低い場合には、その2枚のフレーム画像は異なるショット、すなわち、その2枚のフレームの間にショット境界が存在すると判定する。本研究では、2枚のフレーム画像の色ヒストグラムの類似性の評価値として、フレーム1の各色相の度数 x_i とフレーム2の各色相の度数 y_i との間の相関係数を用いることにした。相関係数 r の計算は次式を用いて行った。ただし、 $\bar{x}$ 、 $\bar{y}$ は、 x_i 、 y_i の平均値、 N は色相の数である。

$$S_x = \sqrt{\frac{1}{N} \sum_i (x_i - \bar{x})^2} \quad S_y = \sqrt{\frac{1}{N} \sum_i (y_i - \bar{y})^2}$$

$$S_{xy} = \frac{1}{N} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$r = \frac{S_{xy}}{S_x S_y}$$



図3 フレーム画像（上）と色ヒストグラムの計算結果（下）

4. ショット境界の判定法

4.1 2つの色ヒストグラム

ショット境界の判定は、1秒間隔離れた2枚のフレーム画像の色ヒストグラムの相関係数を用いて行

う。図4に1秒間隔離れた2枚のフレーム画像とそのフレーム画像から計算した色ヒストグラムを示す。図4の例では、2枚のフレーム画像は異なるショットに属しており、2枚のフレーム画像の間にショット境界が存在することになる。

ここで、連続する2枚のフレーム間でショット境界の判定を行わないのは、例えば、ショットが数フレームの間で徐々に切り替わるような場合にも対応可能であり、すべての連続するショット間で色ヒストグラムの相関係数を計算する場合と比較すると計算コストを低く抑えることができるためである。一方このままでは、正確なショット境界をフレーム単位で決定することはできないが、必要となれば、ショット境界が検出された場合のみ、その間の連続する2枚のフレーム画像をとりだし、同様の判定を行うことで、より正確なショット境界の位置を検出することも可能であると考えている。

本研究では、色ヒストグラムとして、フレーム画像全体の色ヒストグラム（全体的色ヒストグラムと呼ぶことにする）と第1フレームの最大色相を除いた残りの色ヒストグラム（ここでは部分的色ヒストグラムと呼ぶことにする）の2つの色ヒストグラムを用いてショット境界の判定を行う。図4の例では、部分的色ヒストグラムとは、第1フレームの最大色相、図4の例ではヒストグラムのもっとも左端のbinを両フレームのヒストグラムから除いた15色相を用いることになる。



図4 1秒間隔の2枚のフレーム画像（上）とその色ヒストグラムの計算結果（下）

4.2 相関係数の計算

ここでは、全体的色ヒストグラムの相関係数（全体色の相関係数）と部分的色ヒストグラムの相関係数（部分色の相関係数）を用いて、ショット境界の有無を判定するために、複数のフレーム画像対から2つの相関係数の計算を行う。

フレーム画像対として、異なる3試合のサッカー放送を録画したものから、試合ごとにショット境界を含む50組と、含まない50組、3試合で150組ずつ、計300組を用意した。フレーム画像対の時間間

隔は 1 秒とした。図 5 にフレーム画像対の例を示す。上がショット境界を含む場合、下はショット境界を含まない場合の例である。

図 6 にショット境界を含まないフレーム画像対 150 組とショット境界を含むフレーム画像対 150 組から計算した相関係数の値を示す。横軸が全体色の相関係数、縦軸が部分色の相関係数の値である。●がショット境界を含まないフレーム画像対の結果であり、○はショット境界を含んだフレーム画像対の結果である。ショット境界を含まないフレーム画像対は、図の右上、つまり全体色の相関係数、部分色の相関係数ともに高い値をしめす場合が多いことがわかる。一方、ショット境界を含んだフレーム画像対の場合には、左下、つまり全体色の相関係数、部分色の相関係数ともに低い値を示すものが多い傾向はみられるが、図の全体に広く分布している。



図 5 フレーム画像対。ショット境界を含む（上）、ショット境界を含まない（下）

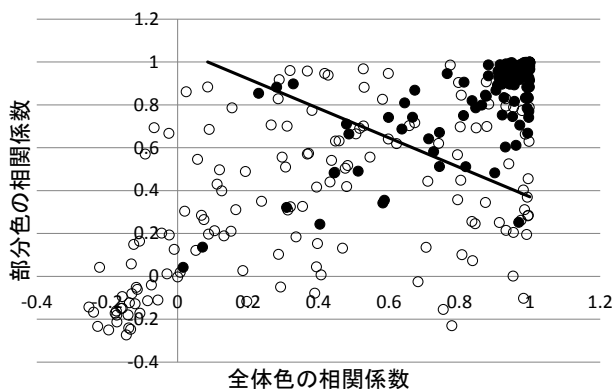


図 6 相関係数の計算結果

- : ショット境界を含まない
- : ショット境界を含む
- : 判別式

提案手法の特徴を明らかにするために、図 6 の全体色の相関係数の値と部分色の相関係数の値と実際のフレーム画像の特徴について検討する。

まず、全体色の相関係数が高く、部分色の相関係

数が低くなる場合、つまり図 6 の右下の領域は、背景が似ているが、前景が異なるような場合が典型的な例となる。サッカー放送の場合では、図 2 に示すように、ショット境界を含むフレーム画像対が似たような構図でありながら、異なる人物をとらえたような場合で、図 2 の場合では、全体色の相関係数が 0.99、部分色の相関係数が 0.20 であった。このような場合、全体色の相関係数だけで、ショット境界の判定を行うことはできないが、部分色の相関係数と合わせて判定することで、ショット境界を正しく検出できる可能性がある。一方、図 7 のように同じショットでありながら、フレーム内に別の人物が入ってくるような場合にも、全体色の相関係数は高いが、部分色の相関係数は低くなる傾向を示す。図 7 の場合では、全体色の相関係数が 0.82、部分色の相関係数が 0.51 であった。このような場合には、誤検出の可能性が出てくる。

次に、全体色の相関係数が低く、部分色の相関係数が高くなる場合、図 6 の左上について考える。このようになるのは、図 2 や図 7 の逆で背景が異なり、前景が似ている場合が定型的な例となる。サッカー放送の場合では、図 8 の上に示すように、同じショットでカメラがパンしたり、ズームしたりする場合に、全体色の相関係数が低くなるが、部分色の相関係数は高い値となる傾向がある。図 8 の上の場合には全体色の相関係数が 0.28、部分色の相関係数が 0.88 であった。この場合、全体色の相関係数のみによってショット境界の判定を行うと誤検出となるが、



図 7 同じショットでありながら、部分色の相関係数が低くなる例



図 8 全体色の相関係数が低く、部分色の相関係数が高い例。ショット境界を含まない（上）、ショット境界を含む（下）。

部分色の相関係数を合わせて利用することで、正しい判定を行える可能性がある。一方、図8の下例はショット境界を含むフレーム画像対であり、全体色の相関係数が0.32、部分色の相関係数が0.96となる。この場合には、全体色の相関係数のみによって判定すれば正しく判定できるが、部分色の相関係数を合わせて使うことで、未検出となる可能性がある。

また、図6を見るとショット境界を含まない画像対でありながら、全体色、部分色の相関係数がともに極端に低い場合がある。これは、図9に示すように選手が比較的大きめに写った状態で、ボールを追ってカメラが高速でパンするような場合である。図9はゴールキックのシーンでボールを追ってカメラが高速でパンしている。このような場合には、前後のフレームで前景も背景をもともに変化しており、全体色も部分色もともに前後のフレーム間で大きく変化することとなり、相関係数の値がともに低くなる。図9の例では、全体色の相関係数が0.02、部分色の相関係数の値が0.04で図6のもっとも左下に位置する黒丸の画像対である

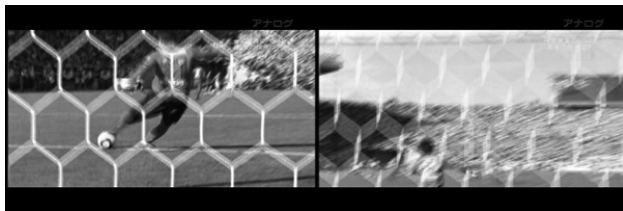


図9 同じショットで全体色、部分色の相関係数がともに低くなる場合

最後に、ショット境界を含みながら、全体色、部分色の相関係数がともに高い値を示す場合である。これは、図6の右上の領域にある白丸である。このようなフレーム画像対としては、似たような構図から、同じチームの異なる選手のアップ、ミドルショットを切り替える。あるいは、フィールド全体を俯瞰するようなロングショット間でショットを切り替える場合などが、典型的な例となる。これらを、同一ショット内の2つのフレーム画像対と区別して判定することは、色ヒストグラムを利用した、ショット境界検出では難しい。

4.3 ショット境界の判定基準

ここでは、図6に示した、相関係数の計算結果をもとに、ショット境界の判定基準について検討する。本研究では、図6に示した、300のフレーム画像対の相関係数の値を説明変数とし、ショット境界を含む表1判別分析の結果

	判別係数	標準判別係数	
全体色の相関係数	-2.898	-1.165	
部分色の相関係数	-4.217	-1.630	
定数項	4.464		
	F値	p値	判定
全体色の相関係数	25.14	0.000	【**】
部分色の相関係数	49.19	0.000	【**】
定数項			

むか否かを目的変数とした線形判別分析を行った。分析結果を表1に示す。全体色の相関係数、部分色の相関係数ともにショット境界の有無の判定に対して有意である。

そこで、表1の判別係数をもとに、判別式を以下のように定めた。

$$y = -2.898r_1 - 4.217r_2 + 4.464 \dots (1)$$

ここで、 r_1 は全体色の相関係数、 r_2 は部分色の相関係数である。 y の値が正となるときは、2枚のフレーム画像対の間にショット境界があると判定し、 y の値が負の場合には、ショット境界がないと判定する。判別式を図6に直線で示す。直線より上の領域はショット境界がないと判定する領域、直線より下はショット境界があると判定する領域となる。図6の300対のデータに対し、下式のように適合率と再現率を計算すると、適合率は0.90、再現率は0.76となった。

$$\text{適合率} = \frac{\text{正検出数}}{\text{正検出数} + \text{誤検出数}}$$

$$\text{再現率} = \frac{\text{正検出数}}{\text{ショット境界を含む画像数}}$$

ここで、正検出とは、ショット境界を含むフレーム画像対が式(1)によってショット境界があると正しく判定された場合である。つまり、図6の白丸うち直線の下領域に位置する場合が正検出となる。一方、誤検出とはショット境界が存在しないフレーム画像対でショット境界があると間違って判定される場合であり、図6の黒丸のうち直線の下領域に位置するときに誤検出となる。また、ショット境界を含む画像数とは、ショット境界を含むフレーム画像対の数であり、150である。

5. 判別精度の確認実験

ここでは、式(1)を用いたショット境界の判定精度について検討する。判定精度を検討するために、サッカー放送を録画した2試合の映像から、1分間のシーンを各試合15シーン、2試合で30シーン、すなわち30分の映像を用意した。映像データの先頭フレームから順に1秒間隔で2つのフレーム画像対を抜き出し、全体色の相関係数と部分色の相関係数をもとに、式(1)を用いてショット境界の有無を判定した。使用した映像ソースは式(1)を決定する際に使用した試合とは異なる試合の映像である。

表2に結果を示す。ショット境界の数は試合1が179、試合2が174、2試合合計で353であった。そのうち式(1)によって正しく検出された数（正検出）は、試合1が133、試合2が137、ショット境界ではないところでショット境界と判定した誤検出は試合1で4、試合2で13であった。また、ショット境界の検出に失敗した未検出数は試合1で46、試合2で37であった。

表2 ショット境界の検出結果

	境界数	正検出	誤検出	未検出
試合1	179	133	4	46
試合2	174	137	13	37
計	353	270	17	83

表3に適合率と再現率の計算結果をまとめる。試合によって、多少の違いがあるが、全体では、適合率が0.94、再現率が0.76となった。式(1)を決定した300組のデータから得られた適合率、再現率と比較すると、再現率はほぼ同じ値となったが、適合率がやや高くなった。これは、式(1)を決定する際に、誤検出を起こしそうな画像対を多めに選んだことが影響しているものと考えている。本研究の目的はダイジェスト映像の制作であり、ショット境界ではないところにショット境界を検出して重要なシーンをショットの途中でつなぎあわせてしまう不自然さを少なくすることが必要となる。そこで、誤検出を少なくすることは未検出を少なくするより重要な課題と考えて、ショット境界の判別式をたてる際に誤検出を起こしやすそうなフレーム画像対をあらかじめ多めにいれておくようにした。

表3 適合率と再現率

	適合率	再現率
試合1	0.97	0.74
試合2	0.91	0.79
計	0.94	0.76

6. 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、ダイジェスト映像を作成することを目的としたショット境界の検出法について検討した。ショット境界判定は、フレーム画像全体の色ヒストグラムとその色ヒストグラムの最大度数のビンを除いた色ヒストグラムから計算される全体色の相関係数と部分色の相関係数を用いて行った。まず、ショット境界を含む150組のフレーム画像対とショット境界を含まない150組のフレーム画像対から計算した全体色の相関係数と部分色の相関係数の値から線形判別分析によって、ショット境界判別式を求めた。その判別式によって、実際に録画された映像ソースを用いてショット境界の判別実験を行ったところ、再現率0.76、適合率0.94という良好な結果を得た。

今後の課題としては、筆者らがすでに提案している、重要フレームの検出法と合わせて、実際のダイジェスト映像の作成と評価を行うことである。

参考文献

- Y. Li、田中譲、“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)
- 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・境界ソサイエティ大会講演集、SS-68-69
- 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)
- 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-II No.6、pp.1122-1131、(2001)
- 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)
- 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-II No.11、pp.1692-1700、(2002)
- 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)
- 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)
- 田村 徹、陳 曉琴、“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”、第10回情報科学技術フォーラム講演論文集、第3分冊、pp.447-452 (2011)
- 田村 徹、陳 曉琴：サッカー放送映像のダイジェスト化に向けたテロップ領域の検出、2012年電子情報通信学会総合大会情報・システム講演論文集2、pp.34 (2012)
- Ahmet Ekin, A. Murat Tekalp and Rajiv Mehrotra, “Automatic Soccer Video Analysis and Summarization”, IEEE TRANS. IMAGE PROCESSING, VOL. 12, NO. 7, pp. 796-806, (2003)
- 新田直子、馬場口登、北橋忠宏、“放送型スポーツ映像の構造を考慮した重要シーンへの自動アノテーション付け”、電子情報通信学会論文誌、Vol.J84-D-II No.6、pp.1838-1847、(2001)