広視域角多視点映像視聴のための
視聴者の嗜好に応じた自動視点推薦方法

冨安 史陽

概要

近年，視聴者が視点と視線を切替えることで被写体を様々な位置や角度から視聴できる「複数の視点を有する映像」のエンターテインメント利用に注目が集まっている．しかし，視点切替えは 7 次元もの変数選択を連続して実施する複雑な操作であり， 視点切替え自体が視聴者の負担となる．この問題に対し先行研究では，視点切替えの「簡便化」と「自動化」という 2 つの観点から個別に議論している．しかし，実際の視聴状況を考えると，両者を組合わせた最低限の視点切替えで常に望む視点からの映像を視聴することができる新たな視聴方式が必要である．

これに対し本論文では，人とシステムがインタラクティブに協調しながら視聴する視点を選択する協調視聴方式が，将来実現されることを想定している．そのためには，自由なタイミングで視点切替えに手動介入できる自動視点推薦機能，すなわち視聴者の嗜好を反映した任意の視点切替えから，適応的に推薦内容を更新できる自動視点推薦方法の実現が不可欠である．

そこで本論文では，フィールドを取囲む複数台の固定カメラでサッカーの試合を撮影した広視域角多視点映像を対象に，推薦視点系列の生成を多視点映像の視点切替えを表現した「重み付き有向グラフの最適経路探索問題」とみなし，これを解くための 2 つの基盤技術を提案して検証した．それらは，多視点映像を適切な時区間に逐次分割することによる過不足のないグラフのノード生成技術と，視点切替えを表すノード間のリンクのコストを決定するための属性に寄与する特徴点追跡技術である．

第 1 章では，複数の視点を有する映像のエンターテインメント利用に注目が集まっているという背景と映像視聴における問題を述べる．次に，映像視聴の支援を行う先行研究の課題とその解決に向けたアプローチとして前提とする協調視聴方式を概説し，本論文の目標について述べる．そして，各章の位置づけについて説明する．

第 2 章では，複数の視点を有する映像の編集と提示に関する先行研究を紹介する．映像の編集では，一般的な映像生成手順を列挙し，このうち映像撮影後の処理となる 2D 映像解析技術の 1 つである特徴点追跡手法について述べる．その後，映像分割手法および映像選択手法について述べる．映像の提示では，視聴インタフェースについて先行研究をまとめる．その後，本論文の位置付けおよび対象とするコンテンツについて概説する．

第 3 章では， 多視点映像を自動的に適切な時区間に 分割する映像分割手法

について述べる．本技術は，分割した映像をノードとし，視点切替えフレームをリンクとして，視点推薦に用いる重み付き有向グラフを作成するのに必要となる重要な技術である．

具体的には，まず 多視点映像の編集結果を収集し，視聴対象となる被写体に着目して分析を行った．その結果，一般的な視聴者は，ボールに着目して映像を視聴し，被写体であるボールとカメラとの幾何関係を考慮して視点を切替えることが分かった．そこで，この規則を満たす多視点映像分割アルゴリズムを開発し，実際のサッカーの試合映像を用いた評価実験より，提案手法が従来手法よりも視聴者の視点切替えと同じタイミングで過不足なく映像を分割できることを確認した．

第 4 章では，速度変化と姿勢変化が顕著なサッカー選手から抽出される画像上での位置変化と特徴量変化が顕著な特徴点を追跡する手法について述べる．本技術は，視点推薦に利用する重み付き有向グラフの作成において，グラフのリンクの重みを設定するために用いる選手の身体動作の変化量の計算に必要な技術である．具体的には，特徴点マッチングにより画像空間の広域を粗に探索し，その後，探索した特徴点を開始位置として Mean-Shift 探索により狭域を密に探索する．シミュレーション映像およびサッカーを撮影した実映像を用いた評価実験において，提案手法が従来手法よりも長時間正確に特徴点を追跡できることを確認した．

第 5 章では，本論文を総括する．本論文をまとめ，今後の研究の展望を述べる．

本論文の貢献は，広視域角多視点映像を一般的な視著者の基準に基づき過不足なく分割する手法と，映像中のサッカー選手の特徴点を高精度に長時間追跡する手法の開発により，多視点映像の視聴における視点系列の推薦を重み付き有向グラフの最適経路探索問題とみなす自動視点推薦方法を提案した点である．

目次

第 1 章　序論　1

1.1　背景　1

1.2　問題と目的　1

1.3　先行研究の概略と課題　4

1.4　アプローチと目標　7

1.5　本論文の構成　12

第 2 章　複数の視点を有する映像の編集と提示に関する先行研究　13

2.1　複数の視点を有する映像の編集　14

2.2　複数の視点を有する映像の提示　25

2.3　本論文の位置づけ　27

第 3 章　視点選択を考慮した広視域角多視点映像の分割　33

3.1　はじめに　33

3.2　被写体とカメラ間の関係性を用いた映像分割方法　36

3.3　映像分割の評価実験　50

3.4　第 3 章のまとめ　65

第 4 章　粗密探索に基づくサッカー選手の特徴点追跡　67

4.1　はじめに　67

4.2　特徴点マッチングと Mean-Shift 探索による特徴点の粗密探索　71

4.3　特徴点追跡の評価実験　78

4.4　第 4 章のまとめ　85

第 5 章　結論　89

5.1　まとめ　89

5.2　今後の研究課題　92

謝辞 97

参考文献 99

研究業績一覧 105

図目次

1.1 用語のイメージ 3
1.2 協調視聴方式における視点選択 8
1.3 映像分割の例 11
1.4 推薦視点系列の例 11
1.5 重み付き有向グラフの例 11
1.6 最適経路探索の例 11

2.1 複数の視点を有する映像をコンテンツ化するまでの 4 つのプロセスの関係性 14
2.2 本論文の位置づけ 31

3.1 赤羽におけるカメラとレンジセンサの配置 37
3.2 豊田におけるカメラとレンジセンサの配置 38
3.3 レンジセンサデータの例（赤羽） 39
3.4 多視点映像の例（赤羽） 40
3.5 カメラパラメータの算出 40
3.6 多視点映像編集用インタフェース [16] 42
3.7 ボールとカメラ間の関係 45
3.8 編集結果とボールとカメラ間の関係 46
3.9 視点切替えとボール位置 47
3.10 検出成功許容しきい値に対する再現率の変化 53
3.11 両時間幅 $T_e = T_i$ に対するカット数の変化 54
3.12 両時間幅 $T_e = T_i$ に対するリンク数の変化 54
3.13 多視点映像の編集結果と分割結果の比較 56
3.14 視聴対象領域の拡大に伴う，視聴対象フレーム数の増加とフレーム一致率 58
3.15 4 つのカット抽出条件におけるカット数と再現率の変化 61
3.16 4 つのカット抽出条件の例 62

3.17 新たな編集映像データセットに対するカット抽出精度の再現率とカット数の変化 65

4.1 特徴点追跡処理の流れ 71
4.2 特徴点マッチング候補検出点の絞込み 74
4.3 追跡点と各画素との特徴ベクトル間類似度分布 76
4.4 追跡対象画像 78
4.5 全画像シーケンスの各手法別特徴点追跡成功率 80
4.6 追跡成功許容範囲 4 画素における追跡対象画像の変化パターン毎の追跡成功率 81
4.7 20 フレーム置きのサッカー映像の追跡結果 84
4.8 サッカー映像の追跡成功特徴点数の変移 85

表目次

1.1 用語の定義 2
1.2 視点切替えの性質とそれに応じた先行研究との関係 4
1.3 複数の視点を有する映像の視点推薦方法のまとめ 6

2.1 提案する特徴点追跡手法の位置づけ 19
2.2 本論文の映像分割手法の位置づけ 22
2.3 本論文の「撮影」,「編集」,「提示」 27

3.1 測定機器の仕様 37
3.2 抽出したシーンの性質の集計 41
3.3 編集結果における選択ショットのフレーム長の分布とボール含有率 43
3.4 再現率とカット数とリンク数 55

4.1 各特徴点の名称と表記方法 72
4.2 特徴点マッチングによる Mean-Shift 探索の初期値 75
4.3 追跡対象画像の変化パターンと各フレーム間の変化量 79
4.4 追跡成功率の代表値 87
4.5 追跡成功率の代表値（続） 88

第 1 章

序論

1.1　背景

我々は，テレビジョンを通してサッカーや野球，水泳，マラソンなど様々なスポーツを観戦することができる．さらに近年では，DAZN（ダゾーン）[1] やスポーツナビ [2]，スカパー [3]，WOWOW[4] などのスポーツ動画像配信サービスも盛んになってきており，国内外の多くのスポーツを観戦することができるようになった．

一般的なスポーツ中継では，複数のカメラで試合会場を取り囲み，全てのカメラで同時に試合を撮影する．そして，テレビジョン放送局の編集者が，撮影した全ての映像の中から最適な視点の映像を選択し，繋ぎ合わせることで一本の放送映像を編集する．したがって，我々は，第三者によって既に編集された映像をテレビジョンやインターネットを通じて視聴している．

これに対し，視聴者が，試合を撮影した全ての素材映像を受け取り，視点と視線の選択権を獲得することができれば，自らの嗜好に基づき視点と視線を切替えながら映像を視聴することができるようになる．これにより，従来の編集された映像を視聴するのとは異なり，自身の好みの位置や角度から試合を楽しんだり，特定の選手に注目して映像を視聴することが可能となる．

1.2　問題と目的

第 1.1 節で述べるように，複数の視点を持つ映像の視聴において，視聴者が視点と視線を切替えられるというのは，被写体を好きな位置や角度から視聴できるという点において利点となる．しかし，従来のテレビジョン放送のような既に編集された映像を視聴してきた視聴者にとっては，複数の視点を有する映像を視聴するために視点を切替え続けること自体が大きな負担となる．

表 1.1　用語の定義

Term	Definition
ショット	各カメラの全映像 (各視点映像)
シーン	実世界の意味的にまとまりのある状況 またはそれを規定する時区間
シーン映像	シーンを説明する編集された映像
カット	ショットを分断する区切り (必ずしもカメラ間で同期しているわけではない)
クリップ	全てのシーン映像から構成されるまとまり
視点切替 フレーム	編集途中あるいは最終的なシーン映像で 視点映像を切替えるカット位置のフレーム
部分ショット	カットで分断されたショットの一部分
選択ショット	視聴対象となった部分ショット

複数の視点を有する映像の視聴における視点切替えとは，知覚・認知・行動からなる人の決定プロセスにおいて，各視点から撮影した映像（以降，視点映像と表記）を視聴することで被写体の空間を知覚し，どの視点が自分の嗜好に最も合っているのかを判断することで次の行動を認知し，瞬時にその視点を選択するという行動を行う 作業である．一般的に，実世界の 3 次元空間を画像の 2 次元空間に投影する際の画角を固定と考えると，視点（x, y, z の 3 次元），視線（x 軸回転，y 軸回転, z 軸回転の 3 次元），時間（t の 1 次元）の計 7 次元の変数セットから，瞬時に最適な組合せを決定する作業とみなすことができる． この様子を図 1.1 に示す．また，関連する用語の定義を表 1.1 に示す． 図 1.1 は，4 階層の画像ストリームからなり，映像をクリップ，シーン映像，ショット，フレームの順に細分化している [5]．この定義に従うと，視聴者の視点切替えは，各視点映像（ショット）を意味のある単位に分割するフレーム（カット）で部分映像（部分ショット）に分割し，複数の部分ショットを組合わせてシーン映像を生成することと説明できる．さらに，視点切替えを行いながら映像を視聴することは，複数のシーン映像を組合わせることで放送映像（クリップ）を生成することと等しくなる．

ここで，視点切替えの問題について分析すると，次の 2 つに分類される．

- 視点選択の難解性
  - 視点と視線の操作性

    ある時刻の視点切替えについて考えると，視聴者は，時間 t を除いた残りの

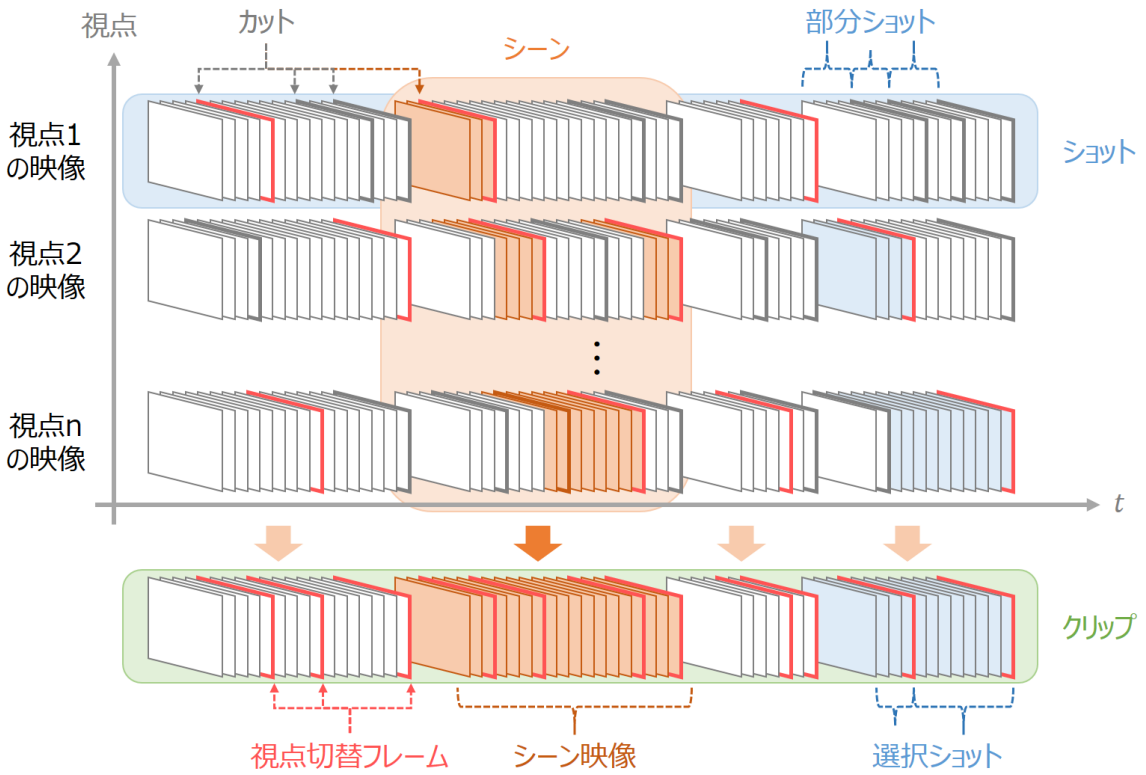


図 1.1　用語のイメージ

6 変数（視点と視線）を制御して，視聴する視点映像を決定しなければならない．しかし，視聴者が，一瞬で 6 つもの変数を同時に 制御することは困難である．

– **被写体の視認性**

被写体の視認性とは，被写体を見たときに，被写体そのものや被写体同士の関係性について，正しく理解できるかどうかの度合の事である．特に複数の視点を有する映像では，視点と視線の組合せが無限に存在するため，現在視聴している視点映像以外の視点からの被写体の 3 次元的な構造と 2 次元的な映像との関係性を把握することは困難である．そのため，視点映像を決定する作業は，視聴者にとってかなり負荷の高い作業となる．

- **視点選択の時間的連続性**

3D 空間における被写体の変化とともに，映像上での被写体の見え方も変化するため，視聴者が望む視点映像を常に見続けるためには，視点選択を時間的に連続して実施しなければならない．

したがって，上記問題を解決する「複数の視点を有する映像を視点切替えしながら不便

表 1.2　視点切替えの性質とそれに応じた先行研究との関係

必須機能	課題	解決方法		
		大分類	小分類	先行研究
視点と視線の選択	視点選択の難解性	手動視点選択の簡便化	視点と視線の操作性向上	[6, 7, 8]
			被写体の視認性向上	[9, 10, 11]
	視点選択の時間的連続性	自動視点推薦による代替化	履歴に基づく手法	[12]
			ルールに基づく手法	[13, 14, 15]
			機械学習に基づく手法（履歴 + ルール）	[16, 17, 18, 19]

なく視聴できる方法」の開発が目的となる．

1.3　先行研究の概略と課題

第 1.2 節で挙げた「複数の視点を有する映像を視点切替えしながら不便なく視聴できる方法」の開発という目的に対し，先行研究では視点選択の難解性に対して「手動視点選択を簡便化するユーザインタフェースの開発」を行い，視点選択の時間的連続性に対して「視点選択を代替する自動視点推薦システムの開発」を行うという 2 つの方法で解決を図っている．第 1.2 節の性質と先行研究との関係を表 1.2 に示す．以下，各先行研究の詳細については，第 2 章にて述べる．

1.3.1　手動視点選択を簡便化するユーザインタフェースの先行研究

先行研究では，視聴者の視点選択の複雑さを解消するために，様々なユーザインタフェースの開発を行っている．これらは，表 1.2 で示したように，「視点と視線の操作性を向上する方法」と「被写体の視認性を向上する方法」の 2 つの方法に分類される．

- **視点と視線の操作性を向上する方法** [6][7][8]

  視点と視線の計 6 次元の変数決定において，視点と視線を直感的に制御する様々な方式やユーザインタフェースが研究開発されている．

  最も身近な例としてゲーム（RPG，シューティング，スポーツ）が挙げられる．家庭用ゲーム機であれば，ゲームの仮想空間内に定点設置された仮想視点間を，ゲーム機のコントローラのボタンを押すことで切替えることができる．また，コントローラのジョイスティック操作により，ゲームの 3 次元空間内に設置したキャ

ラクタや仮想カメラを視聴者の意思通りに移動させることで視点と視線を切替える方法もある．さらに近年では，スマートグラスやヘッドマウントディスプレイが登場し，装着者の身体動作に基づいて視点や視線を変更する方法に注目が集まっている．

次に身近な例として，多視点映像の実応用例にあたる遠隔会議システムや監視カメラシステムが挙げられる．遠隔会議システムは，遠隔地に設置したカメラで会議参加者を撮影し，ディスプレイに表示する視点映像をスイッチングすることで視点を切替えるシステムである．また，監視カメラシステムは，監視対象の施設内に設置したカメラの映像をスイッチングすることで視点を変更したり，制御可能な雲台を利用することでジョイスティック操作などによる視線の変更が可能である．

また，製品として利用されている例として，様々な乗り物のシミュレータ（飛行機，自動車，電車，船，宇宙船）が挙げられる．これらは，運転操作を模擬するため本物同様の操作部を作成し，機器の操作に応じて視点や視線が変化する仕組みとなっている．

さらに，自由視点映像を題材に視点や視線の変更を研究した例として，普段我々が日常的に行う「覗き込む」や「回り込む」といった視認動作により視点切替えを行うユーザインタフェース [6][7] や，視点位置と視線方向を 2 マーカで直感的に指定するユーザインタフェース [8] などがある．

- **被写体の視認性を向上する方法** [9][10][11]

  被写体の 2 次元的な視認性と 3 次元的な視認性を高める様々なユーザインタフェースが開発されている．被写体の 2 次元的な視認性を高める方法としては，特定の被写体を常に画像中央に表示する視聴方式 [9] や，全視点のサムネイル映像を同時に表示するユーザインタフェース [10] がある．また，被写体の 3 次元的な視認性を高める方法としては，被写体とカメラとの関係性を俯瞰図に表示するインタフェース [11] が存在する．

1.3.2 視点選択を代替する自動視点推薦システムの先行研究

先行研究では，視聴者の視点選択の回数を減らすために，様々な視点推薦システムの開発を行っている．これらは，映像生成システムの一部である．映像生成では，ある方針に従って複数の撮影済みの映像を整理し一本の映像にまとめる「映像編集」と，映像の中心となる大切なポイントを取りまとめる「映像要約」の 2 つが議論されている．これらは主にオフラインでの処理であり，映像の編集及び要約と映像の視聴は異なる時間に異なる場所で行われることが一般的である．

表 1.3 複数の視点を有する映像の視点推薦方法のまとめ

手法	視聴対象	抽出方法	利点	欠点
履歴	複数人	視聴履歴	システム構成が簡易	膨大な視聴履歴が必要
ルール	複数人	画像特徴量	全映像に適用可能	個人に応じた推薦苦手
機械学習	個人 複数人	視聴履歴 画像特徴量	個人に応じた推薦可	膨大な視聴履歴が必要

特にサッカーなどのスポーツ中継を視聴する場合を考えると，複数の視点を有する映像の視聴における視点推薦は，映像の時間的連続性を担保しつつ，複数の視点から視聴者に推薦する視点系列を算出する作業とみなせるため，前者の「編集」に分類される．しかし，視聴者の自由な視点切替えを可能とするためには，オンラインでの処理が必要となる．その上で，映像から如何にして重要なシーンを抽出するかが問題となる．ここで課題となるのは，誰にとって重要なのか，そしてそれを如何にして抽出するのかの 2 点である．

これに対し先行研究では，「履歴に基づく手法」と「ルールに基づく手法」，「機械学習に基づく手法」の 3 つの手法で解決を図っている．各手法の特徴を，表 1.3 にまとめる．

- **履歴に基づく手法** [12]

  1 つの映像を複数の視聴者が視聴することを前提に，複数の視聴者が過去に視聴した履歴を参考にして，新たな視聴者が同じ映像を視聴する際に，映像の重要なシーンを提示することを目指している．視聴履歴の解析技術の開発により，複数人が視聴したフレームを連続的につなぎ合わせることで，一般的に興味度の高い視点系列を推薦することが可能である．しかし，視聴履歴の少ない映像では，視点の推薦精度が低下する．
- **ルールに基づく手法** [13, 14, 15]

  画像から抽出した画像特徴量を用いて，映像文法などを参考に全視点を数値評価し，点数の高い視点を選択することで，映像の重要なシーンを抽出することを目指している．画像特徴量の算出さえできれば，如何なる映像に対しても視点推薦が可能である．しかし，映像に対する個人の嗜好の違いなどを考慮した視点推薦には不向きである．
- **機械学習に基づく手法** [16, 17, 18, 19]

  履歴に基づく手法とルールに基づく手法を組合わせた手法である．具体的には，画像特徴量に視聴者の嗜好を表現する重みを付加し，この重みを視聴履歴から求める．そして，これに基づき全視点を数値評価して，点数の高い視点を重要なシーン

として抽出する．全映像に共通の画像特徴量を利用し，多量の視聴履歴から重みさえ求めることができれば，個人と複数人とに関係なく視点推薦することが可能である．しかし，重みを求めるためには膨大な視聴履歴が必要となる．

1.3.3　先行研究の課題

以上より，「複数の視点を有する映像を視点切替えしながら不便なく視聴できる方法がない」という問題の解決に対し，先行研究では「手動視点選択」と「自動視点推薦」を個別の研究として独立に議論している．これらは一見して排他的であるが，視聴者の視点切替えが簡便になっただけでは，視聴者は常に視点を選択し続けなければならず，視聴者の視点切替えに応じて推薦する視点系列が変化しなければ，視聴時における視聴者の自由な視点切替え の効果を損なうことになる．したがって，複数の視点を有する映像を実際に視点を切替えながら視聴するためには，両者を同時に解決する必要がある．

そこで，「手動視点選択」である視聴者の視点切替えに応じて，「自動視点推薦」としてシステムが視聴者の嗜好の変化を読み取り，視点切替え以降のフレームに対して，推薦する視点系列をリアルタイムに適宜変更する手法が必要である．本論文では，視点切替えを学習データとみなし視聴者の映像に対する嗜好を表現する重みを適宜更新する学習ベースの自動視点推薦方法によりこれを解決する． これにより，視聴者は常に自身の嗜好に沿った映像を不便なく視聴することが可能となる．

1.4　アプローチと目標

第 1.4 節では，「複数の視点を有する映像を視点切替えしながら不便なく視聴できる方法がない」という課題を解決するためのアプローチを示す． 本論文では，複数の視点を有する映像を 視聴する新たな視聴方式として，視聴者による「手動視点選択」とシステムによる「自動視点推薦」の両者を同時に実現する人とシステムの協調視聴方式を想定する．

以降は，第 1.4.1 項で人とシステムの協調視聴方式の全体像を説明する．次に，第 1.4.2 項で協調視聴方式の構成要素の一つである「自動視点推薦部」における推薦視点系列の生成を，重み付き有向グラフの最適経路探索問題として解く方法について説明する． その後，第 1.4.3 項で本論文の目標をまとめる．

1.4.1　人とシステムの協調視聴方式

人とシステムの「協調作業」とは，人とシステムが目的を共有し，共に協力し合いながら行う作業である [20]．先行研究では，飛行機や電車などの乗り物の操縦やプラントの監

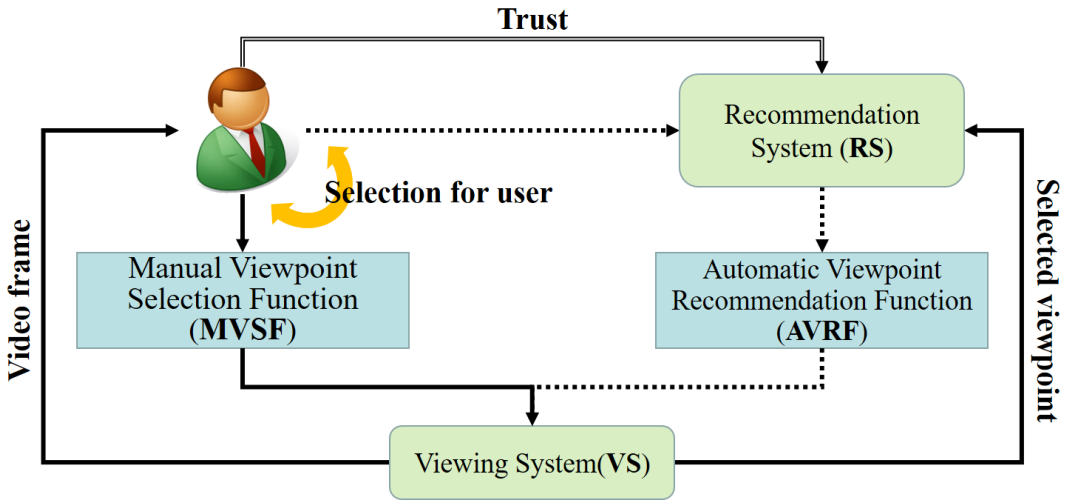


図 1.2 協調視聴方式における視点選択

視といった人の手動制御とシステムの自動制御が混在する作業において議論されている．これに対し本論文では，視聴者が自身の嗜好に応じて視点を切替える「手動視点選択」とシステムが視聴者の視点切替えの負担を軽減するために視聴すべき視点を推薦する「自動視点推薦」とを同時に実装することで，複数の視点を有する映像の視聴を，視聴者と視点推薦システムが，互いの行動を理解しつつ視聴すべき視点を選択する協調作業であるとみなす．そして，これを人とシステムの協調視聴方式と呼び，以下のように定義する．

図 1.2 に，協調視聴方式における各フレームの視点選択の様子を示す． 図 1.2 より，協調視聴方式は，視聴者が視点を選択する「手動視点選択部」と視点推薦システムが視点を選択する「自動視点推薦部」，選択された視点からの映像を表示する「映像表示部」の 3 つからなる．

- **手動視点選択部 (MVSF: Manual Viewpoint Selection Function)**
  視聴者がユーザインタフェースを利用し，映像またはその他のセンサデータを確認することで，2D 的および 3D 的に被写体を理解し，視聴者の嗜好に応じた視点を選択する．
- **自動視点推薦部 (AVRF: Automatic Viewpoint Recommendation Function)**
  自動視点推薦システム (RS: Recommendation System) が，視聴者の視点選択結果および映像内容を表現するアノテーション情報に従い，視聴者の嗜好に応じた視点を選択する．
- **映像表示部 (VS: Viewing System)**
  手動視点選択部または自動視点推薦部で選択された視点の映像 (Video frame) を画面に表示することで視聴者に提供し，映像に付与されたアノテーション情報

(Selected viewpoint) を自動視点推薦システムに提供する．これらは，視聴者と自動視点推薦システムが次のフレームにおいて視点選択する際の参考となる．

手動視点選択部と自動視点推薦部は， 視聴者の視点切替えを介して相互に連携し合う．つまり，複数の視点を有する映像 の視聴において，視聴者は自身の嗜好に応じて視点を切替え，システムはその視点切替えに応じて視聴者の嗜好の変化を読み取り，以降のフレームで視聴者に推薦する視点を更新する．視聴者は推薦された映像を視聴することで，自身の嗜好に沿った視点推薦が行われたかを判断することができるため，次の視点切替えを行うための材料とすることができる．これにより，視聴者は，常に自身の望む視点からの映像を見続けることができるようになるため，各自の嗜好に基づき視点を切替えながら不便なく複数の視点を有する映像 を視聴可能となる．

したがって，人とシステムの協調視聴方式を実現するためには，視聴者の視点切替えに応じた推薦視点系列を生成可能な新たな自動視点推薦部の開発が必要である．

1.4.2　自動視点推薦部

第 1.4.1 項で述べたように，人とシステムの協調視聴方式として，視聴者が不便なく視点を切替えながら常に自身の望む視点からの映像を視聴するためには，視聴者の視点切替えに応じた推薦視点系列を生成可能な自動視点推薦部が必要である．自動視点推薦では，まず，複数の視点を有する映像を，視点切替えを行う可能性のある適切なフレームで部分ショット群に分割する必要がある．そして，これらの異なる視点の部分ショット群から，視聴者の嗜好に基づいて最適な視点を選択する必要がある． つまり，視点切替えなどの視聴者の操作を視点推薦システムへの入力と捉え，画像数値評価関数における視聴者の嗜好を表現する重みを更新することができる機械学習ベースのフィードバックシステムとしなければならない．これにより，視聴者の嗜好の変化をシステムが理解し，これを反映した推薦視点系列の生成が可能となる．

本論文では，推薦視点系列の生成を，重み付き有向グラフの最適経路探索問題として解く．さらに，評価関数の重みの更新を，システムの推薦視点と視聴者の選択視点との差を埋めるように重みを更新する線形計画問題として解く．これにより，従来の機械学習ベースの視点系列生成方法とは異なり，少数のデータセットから重みの更新を行うことが可能となる．

本論文で目指す自動視点推薦部は，次の 4 つのステップからなる．

Step1　映像の分割

複数の視点を有する映像では，視点と視線の組合せが無限に存在する．各視点映像は，画角を固定と考えると，3 次元空間の一部を切り取った映像となる．この時，

被写体は 3 次元空間中を移動するため，ある視点映像には被写体が映り，別の視点映像には被写体が映らないということが発生する．そのため，視聴対象である被写体を常に視点映像に映すような視点系列を生成するために，視聴対象が映るフレームを事前に把握し，意味のある連続したフレーム群として映像を部分ショットに分割する．

Step2 重み付き有向グラフの作成

抽出した部分ショットから，視聴対象が常に映像に映る全視点系列を表現した重み付き有向グラフを作成する．

初めに，抽出した部分ショットを有向グラフのノードとし，部分ショットが時間的に連続するノード間をリンクで接続する．このリンクが，視点切替えを表現している．

最後に，各リンクに，小さい方が視点切替えし易いとする視点切替えのコストを設定する．コストは，視点切換えを評価した数値と視聴者の嗜好を表す重みの線形和とする．例えば，視点切替えの評価値は，画像上における視聴対象の画像特徴量の組合せとして表現し，視聴対象である被写体が画像中心に映っている，大きく映っている，正面を向いている，時間的に変化しているといった場合に映像の見栄えが良いとして小さなコストを設定する．また，視点切替え前後で視点映像の類似度が高く，視聴者の視認負荷が小さい場合もコストを小さく設定する．

Step3 重みの更新

映像視聴開始時は，視聴者の過去の重みを読み出すか，または事前に設定した初期値を利用する．

映像視聴中は，視聴者の視点切替えが発生した際に，重みを更新する．システムが推薦する視点映像と視聴者が選択した視点映像の差を埋めるように重みを更新する．具体的には，システムが推薦する視点映像のノードへのコストよりも，視聴者が選択した視点映像のノードへのコストが小さくなるように，重みの総和が 1 になるという条件の下で，線形計画法により最適な重みの組合せを求める．

Step4 視点系列の推薦

Dijkstra のアルゴリズムを用いて，作成した重み付き有向グラフに対し最適経路探索を行う．

視聴対象である被写体毎に専用の重み付き有向グラフが作成されるため，被写体に応じた専用の有向グラフを用いて最適経路探索を行う．視聴者が視点切替えを行った先の視点のノードを開始点として，そこから一定時間後までの最適経路を算出する．

生成された最適経路に基づき，推薦視点系列を作成し推薦する．

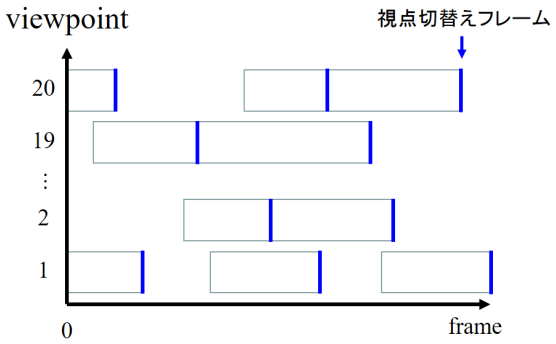


図 1.3　映像分割の例

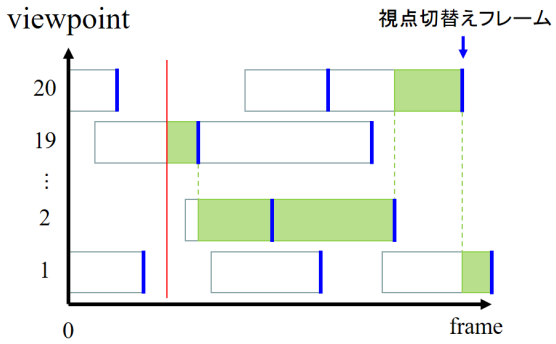


図 1.4　推薦視点系列の例

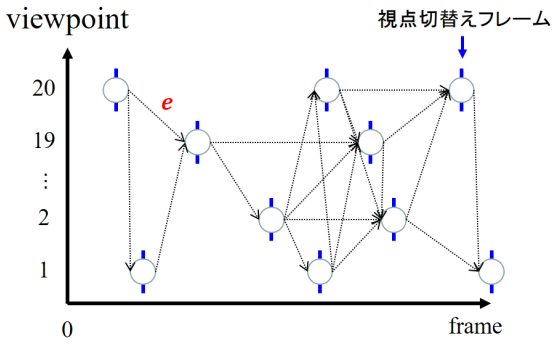


図 1.5　重み付き有向グラフの例

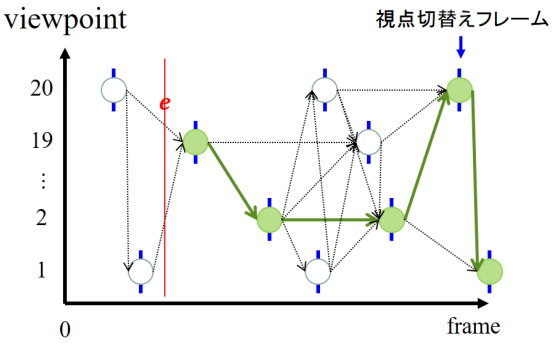


図 1.6　最適経路探索の例

図 1.3 から図 1.6 に，20 視点分の映像を持つ多視点映像を分割し，重み付き有向グラフを作成して，最適経路探索の後，推薦視点系列を生成する様子を示す．これらは全て，縦軸に視点数 (viewpoint) を，横軸にフレーム番号 (frame) をとる．

図 1.3 は，多視点映像から部分ショットを抽出する映像分割の例である．視聴対象である被写体が映る連続したフレーム群を長方形で表す．また，青色の縦線が，視点切替えを行うフレームである．この青線で区切られた長方形が一つの部分ショットである．視聴対象が映る連続したフレーム群においても，映像内容によっては複数の部分ショットに分割される．

図 1.5 に，図 1.3 から作成した重み付き有向グラフを示す．ノードを円で，リンクを点線の矢印で表す．また，リンクのコストを e で記す．青色の縦線は，図 1.3 と同じ視点切替えを行うフレームであり，ノードの位置を表す．この有向グラフは，視聴対象が映る部分ショットの集合からなるため，このリンクを辿ることによって，常に視聴対象が映る視点系列を算出することが可能である．

図 1.6 に，図 1.5 の重み付き有向グラフから算出した最適経路の例を示す．これは，赤色の縦線が示すフレームにて視点切替えが発生したと仮定した場合の最適経路探索の結果を示している．図 1.6 の一連の緑色の太い実線矢印と緑色のノードが最適経路である．

図 1.4 は，図 1.6 の最適経路から算出した推薦視点系列を緑色の長方形で表現している．実際には，赤色の縦線で示す視点切替えフレームの次のフレームから視点系列を推薦する．

1.4.3　本論文の目標

本論文では，第 1.4.2 項で記した自動視点推薦部の実現にあたり，次の 2 つの研究を目標とする．

- **映像分割**

  重み付き有向グラフの生成にあたり，複数の視点を有する映像の分割方法について第 3 章に記す．これは，複数の視聴者の視点切替えに対応した映像分割手法が存在しないため必須の研究課題であると考えている．
- **特徴点追跡**

  視点切替えを表すノード間のリンクのコスト設定に寄与する画像特徴量の 1 つである被写体の身体動作の時間的な変化量の算出に必要な特徴点追跡方法について第 4 章に記す．視聴者の視認性を反映する 被写体の様々な画像特徴量のうち，被写体の位置や大きさ，向きなどについては，被写体の 3 次元的な位置座標と各視点の位置座標の関係から算出できる．しかし，身体動作などの時間的な変化量は映像からしか算出することができないため必須の研究課題であると考えている．

1.5　本論文の構成

本論文の構成は，以下である．

第 2 章では，複数の視点を有する映像をコンテンツとしてサービス化するために必要な技術について，映像の撮影・伝送・編集・提示の 4 つの観点から整理し，特に本論文で対象とする「編集」と「提示」について関連研究をまとめ，本論文の位置づけを示す． 第 3 章では，多視点映像の重み付き有向グラフの作成において，グラフのノードとリンクを決定するために，カメラと被写体の幾何関係から多視点映像を過不足なく分割する手法について述べる．第 4 章では，多視点映像の重み付き有向グラフのリンクのコスト決定にあたり，映像中のスポーツ選手の身体動作の変化量を利用するために，スポーツ選手の特徴点を高精度に追跡する手法について述べる．第 5 章では，本論文のまとめと協調視聴方式の実現に向けた今後の研究課題について述べる．

第 2 章

複数の視点を有する映像の編集と提示に関する先行研究

複数の視点を有する映像を，社会的なエンターテインメント向けコンテンツとするためには，「映像撮影技術」，「映像編集技術」，「映像伝送技術」，「映像提示技術」の 4 つの技術が必要である．第 2 章では，本論文が対象とする「映像編集技術」及び「映像提示技術」の先行研究をまとめ，本論文の位置づけを示す．

- 映像撮影技術
  撮影とは，複数のセンサを用いて被写体を記録することである．どのような被写体をどのような機材でどのように撮影するかという観点が存在する．
- 映像編集技術
  編集とは，撮影した複数の映像から新たな映像を生成する作業のことである．一般的な映像生成の手順を紹介し，その後，撮影後の処理となる 2D 映像解析の特徴点追跡手法，映像編集の映像分割手法と映像選択手法について，先行研究をまとめる．
- 映像伝送技術
  伝送とは，複数の視点を有する映像を視聴者に提供することである．
- 映像提示技術
  提示とは，視聴者に複数の視点を有する映像を提示することである．「視聴ユーザインタフェース」の観点から，先行研究をまとめる．

上記の 4 つの技術は，厳密に独立しておらず，相互に関係し合っている．この関係性を，図 2.1 に示す． 例えば，「撮影」と「編集」の関係では，映像を編集しやすいようにカメラの台数や配置を考慮して撮影を行うことがある．また，「伝送」と「提示」の関係では，視聴者が自らの嗜好に基づき視点切替えを行いながら映像を視聴できるようにする

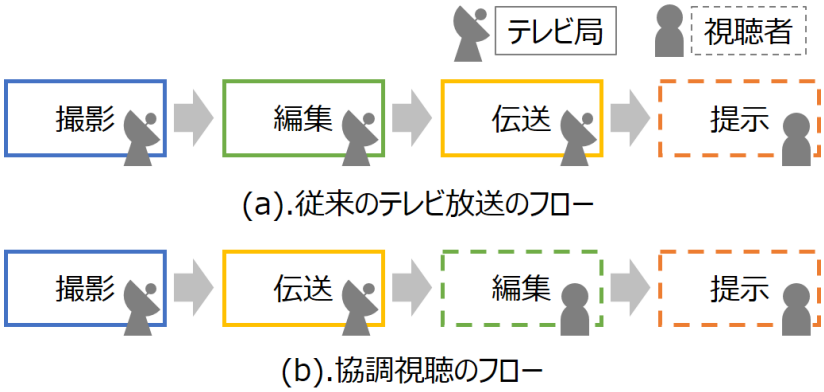


図 2.1　複数の視点を有する映像をコンテンツ化するまでの 4 つのプロセスの関係性

ために，視点切替えに応じて伝送帯域を効率的に制御し，視聴者が選択している視点は高解像度な映像を，他の視点は低解像度な映像を伝送することで全視点の映像を同時に伝送するものがある．さらに，「編集」と「提示」の関係では，複数の視点を有する映像の視聴における視聴者の連続した視点切替えを減らすために，映像から抽出した画像特徴量を利用して自動的に視点系列を推薦するものがある．

従来のサッカー中継では，図 2.1(a) に示すように「撮影 → 編集 → 伝送 → 提示」の順番で処理が進む．具体的には，フィールドを取り囲むように設置した複数のカメラで異なる方向からサッカーの試合を同時に撮影し，それらをテレビジョン放送局の編集者が一本の映像に編集してから，テレビジョンの電波帯に載せて配信し，我々はその映像をテレビジョンを通して視聴する．一方，本論文で対象とする複数の視点を有する映像においては，視聴者が自らの意思に基づき視点を変更できるようにするために，図 2.1(b) に示すような「撮影 → 伝送 → 編集 → 提示」の順番となる．具体的には，同様に複数のカメラで撮影した全ての映像を視聴者に伝送し，視聴者側で映像の編集を行いつつ視聴する．

2.1　複数の視点を有する映像の編集

一般的な映像生成手順は，以下のとおりである．

1. 前処理
   - 撮影機器の設置 ：カメラの位置と角度を決定
   - カメラ校正 ：カメラのレンズ特性とカメラ間の幾何関係を導出
   - 時刻同期 ：全カメラの撮影開始を統一
     あるいは後処理として，撮影映像の開始時間と終了時間を統一
   - 色調調整 ：視点切替えに備え映像間の色味を統一

2. 撮影
3. 後処理
    - ノイズ除去 ：インタレースや白飛び，黒潰れを除去
    - **2D 映像解析 ：各映像から被写体情報を抽出**
    - 3D 映像解析 ：カメラ間の幾何関係から被写体空間を三次元復元
4. 映像編集
    - **映像分割 ：意味のある最小単位（部分ショット）に映像を分割**
    - **映像選択 ：全視点から最適な部分ショットを選択**
    - 映像結合 ：選択した部分ショットを一本の映像に結合

第 2.1 節では，第 2.1.1 項で「後処理」の「2D 映像解析」に関して，「特徴点追跡手法」の先行研究をまとめる．第 2.1.2 項で「映像編集」の「映像分割」に関して，「映像分割手法」について先行研究をまとめる．第 2.1.3 項で「映像編集」の「映像選択」に関して，複数の視点映像から視聴すべき視点映像を推薦する「視点推薦手法」について先行研究をまとめる．

2.1.1 特徴点追跡の先行研究

特徴点追跡は，物体追跡手法の一種である．以下では，第 2.1.1.1 で物体追跡と特徴点追跡の関係について述べ，第 2.1.1.2 で特徴点の特徴量について簡単に説明する．次に，第 2.1.1.3 で特徴点マッチングによる特徴点追跡の先行研究を，第 2.1.1.4 で Mean-Shift 探索による特徴点追跡の先行研究を第 2.1.1.5 で Kalman-Filter と Particle-Filter による特徴点追跡の先行研究を 紹介し，第 2.1.1.6 で本論文の特徴点追跡手法の位置づけをまとめる．

2.1.1.1 物体追跡の先行研究

物体追跡の研究 [21] は，特徴点ベースの手法 [22, 23, 24, 25, 26, 27]，ブロブベースの手法 [28, 29, 30]，カーネルベースの手法 [22, 23, 28, 29, 31, 32, 33, 34, 35, 36, 37] の 3 つに分類される．

特徴点とは，画像の局所領域から得られる特徴量が周辺画素よりも顕著な画素のことである．そのため， 特徴点ベースの手法は，画像から特徴点を抽出し，特徴量の類似度に基づいて 特徴点を追跡し，複数の特徴点の追跡結果を統合することで物体追跡する手法である．追跡物体の局所特徴量が変化しにくい場合に適している．

ブロブ（Blob：Binary Large Object）とは，同じ論理状態のピクセルが隣接している領域のことである．そのため， ブロブベースの手法は，追跡対象の輪郭やシルエット情報などのブロブ を抽出し，この類似度に基づき物体を追跡する手法である．追跡対象の形状

が変化しない剛体物体を追跡する場合に適している．

カーネルベースの手法は，追跡対象を包含する画像領域をカーネル関数で変換して追跡対象モデルを作成し，このモデルの類似度に基づき物体を追跡する手法である．追跡対象の形状変化に対応して追跡するためには，追跡対象モデルの適宜更新が必要であり，カーネル関数に入力する追跡対象を包含する画像領域の適切な更新が必要となる．そのため，形状変化の激しい対象には不向きである．

2.1.1.2　特徴点追跡の先行研究における特徴量について

特徴点追跡において，色情報の一つであるカラーヒストグラムを特徴量として利用する多くの先行研究が存在する [22, 23, 29, 30, 32, 33, 34, 35, 36, 37]．カラーヒストグラムとは，特徴点を中心とする一定領域内の画素値を集計したものである．カラーヒストグラムの色の表現方法は，HSV 色空間や RGB 色空間など様々である．カラーヒストグラムを利用する特徴点追跡手法は，ヒストグラム間類似度に基づく各画素の重み分布に対して探索を行う．そのため，特徴点を中心とする一定領域内における追跡対象の形状変化に対しては，生成されるカラーヒストグラムが変化しないため 頑健であるという利点を持つ一方で，照明変化などによる輝度値の変化に対しては，生成されるカラーヒストグラムが変化するため 弱いという欠点を持つ．

また，特徴点追跡において，輝度勾配情報の一つである Scale Invariant Feature Transform(SIFT)[38] や Speeded Up Robust Features(SURF)[39] を特徴点の特徴量として利用する先行研究が存在する．これらは，特徴点を中心とした局所領域を細分化し，各領域の輝度勾配を方向毎に集計した値を特徴ベクトルとしたものである．これらを利用する特徴点追跡手法は，特徴ベクトル間類似度に基づく各輝度勾配方向の重み分布に対して探索を行う．そのため，大域的な輝度勾配の変化，平面的な回転変化，スケール変化に対して強いという利点を持つ．一方で，オクルージョンや影などによる局所的な輝度勾配の変化に弱いという欠点を持つ．

2.1.1.3　特徴点マッチングによる特徴点追跡の先行研究

広域に対し粗な特徴点の探索を行う手法として，特徴量の類似度に基づき複数画像間の対応点を探索する特徴点マッチングと呼ばれる手法がある [22, 23, 40, 41]．特徴点マッチングは，3 次元測距 [40] やパノラマ画像合成 [41]，物体追跡 [22, 23] などに利用される．

3 次元測距では，幾何関係が既知である 2 台のカメラで撮影した画像から特徴点を検出し，この特徴点群に対して特徴点マッチングを適用することで，被写体の同一箇所を表す特徴点対を算出する．そして，カメラ間の距離が既知であることを利用して，三角測量の原理から被写体までの距離を算出する．パノラマ画像合成では，カメラを微小移動させながら連続撮影した画像から特徴点を検出し，この特徴点群に対して特徴点マッチングを適

用することで，被写体の同一箇所を表す特徴点対を算出する．そして，全ての特徴点対が上手く重なるように画像を変形しながら重畳することで，パノラマ画像を合成する．物体追跡では，時間的に連続撮影した画像から特徴点を検出し，この特徴点群に対して特徴点マッチングを適用することで，被写体の同一箇所を表す特徴点対を算出する．そして，複数フレームにわたり同一の特徴点を探索することで，特徴点を追跡することができる．

特徴点マッチングの欠点として，検出した全ての特徴点群に対して特徴量の比較を行うと，計算量の増加と，類似した特徴量を持つ異なる特徴点との誤対応の増加につながる可能性がある．さらに，被写体の見え方の変化によって，算出される特徴点の位置や特徴量が変化するため，正確な特徴点の対応付けが困難な場合もある．

2.1.1.4 Mean-Shift 探索による特徴点追跡の先行研究

狭域に対し密な特徴点の探索が可能な手法として，Mean-Shift 探索がある．Mean-Shift[42] は，密度関数の極値探索問題についてカーネル密度推定を用いた外乱にロバストな統計的データ解析手法である．Mean-Shift の特徴点追跡への応用では，追跡する特徴点と類似した特徴量を持つ画素を，極値探索問題として探索する．Mean-Shift 探索は，計算コストが低く高速に計算できるという利点を持つ一方で，全探索範囲に対して大域解の算出を保証しておらず，容易に局所解に収束するという欠点を持つ．これは，Mean-Shift 探索の開始位置と追跡対象である特徴点の位置との間に，追跡対象である特徴点と類似した特徴量を持つ追跡対象でない他の画素が存在する場合に，この画素を探索結果として出力してしまうということである．

3 次元空間を移動する被写体を撮影した映像に対する特徴点追跡では，被写体が画像上を平行移動する場合（画像空間の移動）と奥行方向に前後移動する場合（スケール空間の移動）の両者を考慮する必要がある．

画像空間に対する Mean-Shift 探索では，追跡対象である特徴点が画像上を大きく移動すると，Mean-Shift 探索の開始位置である前フレームの特徴点位置と移動後の特徴点位置との距離が長くなり，特徴点と類似した特徴量を持つ異なる局所解に収束する可能性が高くなる．この欠点を解決するために，Kalman-Filter や Particle-Filter を利用して，特徴点の移動先を予測し Mean-Shift 探索の初期位置を変更することで，目標画素周辺から探索を開始する手法がある．これらについては，第 2.1.1.5 で述べる．

スケール空間に対する Mean-Shift 探索では，カラーヒストグラムのようなスケールを考慮しない特徴量の特徴点を追跡する場合と，SIFT や SURF のようなスケールを考慮する特徴量の特徴点を追跡する場合とで，大きく二つの方法がある．前者では，追跡対象のスケール方向の変化に対して，大きさを変えた複数の探索窓を用意し，各 Mean-Shift 探索の結果から最適な探索結果を選択する必要がある．Collins[30] は，追跡対象のスケール方向への変化にロバストな追跡手法として，スケール空間に Mean-Shift 探索を適用する

手法を提案している．後者 [23, 24, 25] では，特徴量自体が被写体のスケールを考慮しているため，複数のスケールに対して特徴量を計算し，追跡対象である特徴点の特徴量との類似度分布に対して Mean-Shift 探索を適用する．都築ら [24] は，特徴点の特徴量として SIFT を利用することで，スケール変化を考慮した Mean-Shift 探索による特徴点の追跡手法を提案している．

2.1.1.5　Kalman-Filter と Particle-Filter による特徴点追跡の先行研究

特徴点の追跡において，特徴点の移動量を予測し，位置を推定する手法として，Kalman–Filter[43, 44] や Particle–Filter[45] を用いる手法がある．

Kalman–Filter とは，逐次ベイズフィルタの一種であり，システムが線形モデルでかつ観測値に対するノイズが白色正規分布に従うといった仮定を必要とする．しかし，この仮定を許容できるシステムにおいては実効性が高く，様々な場面で既に実用化されている．Kalman–Filter と Mean–Shift を併用する多くの研究では，Kalman–Filter による予測位置を Mean–Shift 探索の開始位置とする [22, 23, 28, 29, 31, 32, 33]．しかし，Kalman–Filter は，追跡対象の速度や方向が急に変わると予測に失敗する．すると Mean-Shift 探索の初期位置として不適切な画素位置から探索を開始するため，局所解に陥る可能性が高くなる．特にスポーツ選手の腕や脚などはこれらの変化が大きく，画素位置の厳密な予測が困難である．

また，Particle–Filter も逐次ベイズフィルタの一種であり，非線形なモデルや非ガウス性のノイズが混入するシステムにも適用可能である．Particle–Filter は，「分布」を推定する仕組みであり，実用上は唯一の推定値を算出するために，一般的に「重み付き平均」をよく利用する． Particle-Filter と Mean-Shift を利用する研究では，Particle-Filter による予測位置を Mean-Shift 探索の開始位置とする [34, 35]．Kalman-Filter の場合と同様に追跡対象が大きく移動する場合を考えると，広範囲を探索する必要があり，Particle の数を増やす必要がある．すると，計算量の増加や追跡点以外の局所解に陥る可能性の増加につながる．逆に計算量を抑えるためには Particle の数を制限する必要があり，Particle が Mean-Shift 探索の最適な初期位置にばら撒かれない可能性が高くなる．

2.1.1.6　本論文の特徴点追跡手法の位置づけ

スポーツ選手は，一般的な歩行者に比べて身体動作が大きく，大域特徴量の変化が大きい．しかし，手や足といった各身体部位は剛体であり，各部位の位置変化や姿勢変化はあるものの，部位自体の形状変化は少ない．したがって，特徴点といった局所特徴量の方が，ブロブやカーネルといった大域特徴量よりもスポーツ選手の追跡に適していると考え，本論文では特徴点ベースの手法を採用する．なお，特徴量としては，追跡対象の画像上における 2 次元的な回転変化，拡大変化，輝度変化に頑健な SURF 特徴量を用いる．

表 2.1 提案する特徴点追跡手法の位置づけ

	Coarse	Fine
Local	–	Mean-Shift search
Global	Feature point matching	**Proposed method**

本論文では，スポーツ選手の緩急のある動作による特徴点の急激な速度変化や，スポーツ選手の 3 次元的な姿勢変化による特徴点の特徴量の変化に対応する必要がある．そこで，広域に対し粗な特徴点の探索を行うことができる特徴点マッチングと狭域に対し密な特徴点の探索が可能な Mean-Shift 探索を組合わせる．提案手法の位置づけを表 2.1 にまとめる．表 2.1 より提案手法は，特徴点マッチングと Mean-Shift 探索の利点を生かし，欠点をなくした手法となっている．

従来手法では，特徴点の大きな移動に対し，Kalman-Filter などによる特徴点の移動予測を用いることで Mean-Shift 探索の初期位置を変更していた．しかし，スポーツ選手の緩急のある動作に対しては，予測に失敗することが予想される．そこで，本手法では，特徴点マッチングにより，Mean-Shift 探索の初期位置を変更する．また，単純に画像全体に特徴点マッチングを行うと，計算量と誤対応の増加を招く．そこで，Kalman-Filter による特徴点の移動予測を用いて，特徴点マッチングの対象領域を限定する方法としている．

2.1.2 映像分割に関する先行研究

映像分割は，映像要約 [46, 47, 48, 49, 50] や映像編集 [51, 52, 53, 54] に関する多くの先行研究で議論されている．先行研究では，スポーツ [46, 47, 48, 49, 50]，保育 [51]，講義 [52]，料理 [53]，ライフログ [54] など様々な映像を対象としている．

以下では，第 2.1.2.1 で映像要約の映像分割，第 2.1.2.2 で映像編集の映像分割に関して先行研究を紹介し，第 2.1.2.3 で本論文の映像分割手法の位置づけについて述べる．

2.1.2.1 映像要約における映像分割の先行研究

本論文では，1 本の映像から映像内容をよく表現した部分ショットを抽出し元の映像よりも短い一本の映像を生成することを「映像要約」と呼ぶ．以下，「映像要約における映像分割」の先行研究を紹介する．

Rehman ら [46] は，サッカー映像を対象とした映像要約について幅広く調査を行っている．彼らは，映像分割手法 の 1 つとしてカットの抽出方法を挙げ，カット抽出における低レベル特徴量 (色，形，質感，物体) と高レベル特徴量 (プレーの再生/停止) の利用につい

てまとめている．一般的にはこれらを複合したマルチモーダル情報の利用が有用であるとされる．また，Bach ら [47] は，野球映像を対象に，画像特徴量やカメラの移動特徴量に基づき複数の隠れマルコフモデルを用いて各フレームのシーン認識を行い，シーンの切り替わりをショットの区切りとして映像を分割している．山田ら [48] は，サッカーの試合内容に関するアナウンサーと解説者のコメントに基づき，サッカー番組のセグメントメタデータを自動付与する技術を提案している．各コメントに基づきイベントが発生した区間を抽出することで，映像を分割している．新田ら [49] は，音声ストリームの写しであるクローズドキャプションと呼ばれるテキスト情報に基づくセグメント分割と画像特徴量に基づくセグメント分割の結果を時間的に対応づけることで，アメリカンフットボールや野球の映像の正確な映像分割を行っている．Nitta ら [50] は，MPEG-7 を利用して意味内容から野球映像の木構造を作成し，この木構造に基づき映像を分割している．

これらは，編集済みの放送映像を対象としており，アクティブなカメラ撮影や映像編集によるズーム・視線・視点の変化に伴う画像情報の変化，及び映像の内容を表現した音声やテキストなどの付加情報を利用して映像を分割している．

2.1.2.2　映像編集における映像分割の先行研究

本論文では，複数の映像から主要部分を抽出し 1 本の映像を生成することを「映像編集」と呼ぶ．以下，「映像編集における映像分割」の先行研究を紹介する．

石川ら [51] は，幼稚園に設置した 7 台の固定カメラの映像から，園児の親に提供する園児の 1 日をまとめたダイジェスト映像を生成している．園児に持たせた無線タグより測定した園児の室内位置情報を利用して園児が映るカメラを特定し，画像解析により画像中の園児の位置を特定する．また，保育園の 1 日のスケジュールに基づき，映像をイベントごとに分割する．これにより，視点の選択と映像の分割を行っている．丸谷ら [52] は，遠隔講義や講義アーカイブで用いられる講義映像を，大学の講義を固定カメラで撮影した多視点映像から自動生成しようとしている．講義内容の理解度が高い講義映像が望ましく，理解度を高めるのに適した撮影対象が映る視点映像を選択するため，映像中の講義状況の認識を行っている．彼らは，隠れマルコフモデルを用いて講義状況を認識することで，「語りかけ」,「スライド説明」,「板書説明」などの状況を認識し，この状況変化のタイミングで映像を分割している．熊野ら [53] は，放送の多チャンネル化に対して映像編集作業の効率化が必要であると考えており，映像文法と画像処理を利用してカットとカメラワーク情報を算出することで映像の自動分割を行っている．Sumi ら [54] は，個人の経験を共有するという観点から，室内に設置した固定カメラ映像とイベント経験者の一人称映像および映像と共に測定した位置や視線などのセンサ情報を利用して，個人が体験したイベントの説明映像を編集している．展示会場における来場者の行動に応じて，展示物への

訪問，展示者との会話などのイベントを認識し映像を分割している．Muramatsu ら [16] は，サッカーを撮影した多視点映像中の物体の幾何情報と視聴者の視点選択結果から視聴者個人の視聴傾向を Support Vector Machine で学習し，他のサッカーを撮影した多視点映像に対して視点系列の推薦を行っている．各時刻における全視点のフレームを識別器で数値評価し，評価の高いフレームの視点を推薦対象とする．これを全時刻のフレームに実施し，推薦する視点が切り替わるタイミングで映像を分割している．また，Wang ら [17] も同様に，サッカー映像中の物体の幾何情報を用いて視点を数値的に評価し，最大評価値のフレームを繋ぎ合わせることで視点系列を推薦している．彼女らは，同じ多視点映像を視聴した複数人の視聴履歴から学習を行っており，推薦する視点が切り替わるタイミングで映像を分割している．

これらは，複数のカメラで撮影した素材映像から，画像情報やセンサ情報を用いて映像を分割し，映像内容の時間的な連続性を守りつつ，撮影した被写体やイベントの内容をよく表す 1 本の映像を生成しているとまとめることができる．

2.1.2.3 本論文の映像分割手法の位置付け

本論文では，「サッカーを固定カメラで連続的に撮影した広視域角多視点映像を，一般的な視聴者が自らの嗜好に基づき視点切替えをする可能性のある全てのフレームで，過不足なく分割する」ことを目指している．これは，広視域角多視点映像から個別の視聴者が望む一本の映像を実時間で編集するために必要な研究であり，「映像編集における映像分割」に分類できる．

また，第 2.1.2 項と第 2.1.2.1，第 2.1.2.2 で取り上げた先行研究との関係を表 2.2 に示す．表 2.2 は，次に示す二つの指標にしたがい研究を分類したものである．横軸には，映像要約および映像編集により生成した映像の視聴者を示す．縦軸には，映像要約および映像編集にて映像分割をする際の基準の所在を示す．具体的には，個人の感性や個人の視聴履歴といった個人に特化した基準で映像を分割するものと，映像文法や複数人の視聴履歴といった複数人に共通した基準で映像を分割するものとに分類する．

第一象限は，複数人向けの映像を生成するために，個人の基準に従い映像を分割する方法である．これは，我々が普段テレビジョンや映画で目にする映像の生成方法であり，特定の個人が複数人の嗜好を想定し，自身の判断に従い 映像を分割するものである．

第二象限は，個人向けの映像を生成するために，個人の基準に従い映像を分割する方法である．これは，特定の個人が視聴することを前提に，その個人が望む観点や個人の映像視聴履歴に従って映像を分割するものである． 例えば，石川ら [51] は，園児の親に提供する園児の 1 日のダイジェスト映像を生成するために，その園児が映像に映るという観点に基づいて映像を分割している．Sumi ら [54] は，一人称映像を交えた個人の活動記録

表 2.2　本論文の映像分割手法の位置づけ

基準＼視聴者	個人	複数人
個人	園児要約映像生成手法 [51], 個人活動記録映像生成手法 [54], 個人向け視点推薦手法 [16]	手動編集手法（TV, 映画）
複数人	提案手法	スポーツ映像の要約手法 [46, 47, 48, 49, 50], 講義編集手法 [52], 調理手順編集手法 [53], 複数人向け視点推薦手法 [17, 18]

映像を生成するため，素材映像を個人活動のイベント毎に分割している．Muramatsu ら [16] は，個人の嗜好に沿った多視点映像の視点推薦を行うため，学習結果から個人の視聴傾向に基づき多視点映像を数値評価することで，評価値の高い視点を推薦し，この視点の変化を映像の分割タイミングとしている．

第四象限は，複数人向けの映像を生成するために，複数人の基準に従い映像を分割する方法である．これは，不特定多数の人が視聴することを目的に，共通ルールである映像文法や複数人の映像視聴履歴から算出した共通の嗜好に基づいて映像を分割するものである．例えば，スポーツ映像の要約 [46, 47, 48, 49, 50] では，画像特徴量や音声，テキストなどの情報を利用して，複数人に共通した基準で映像を分割している． 具体的には，画像特徴量の変化量の大きいフレームで映像を分割したり，録音された歓声の大きなフレームを見どころとし，それ以外で映像を分割する．丸谷ら [52] は，複数の講義受講者向けに，講義内容を最も良く表す講義映像を生成するために，多視点映像から講義状況を認識し，状況変化のタイミングで映像を分割している．熊野ら [53] は，放送の多チャンネル化に対し，映像編集の効率化を目指し，映像文法という体系化された共通の基準を用いることで映像の分割を行っている．Wang ら [17] は，複数人に共通した嗜好に沿って多視点映像の視点推薦を行うため，学習結果から一般的な視聴傾向に基づき多視点映像を数値評価することで，評価値の高い視点を推薦し，この視点の変化を映像の分割タイミングとしている．さらに，Wang ら [18] は，視聴者を視聴傾向毎にグループ分けし，グループ毎に学習することで映像の分割精度を向上している．

第三象限は，個人向けの映像を生成するために，複数人の基準に従い映像を分割する方

法である．本論文はこの第三象限に位置づけられる．これは，複数の視聴者が各自の嗜好に応じて視点を切替えながら映像を視聴することを前提に，個人の嗜好に応じた映像を推薦するため，複数人の基準を網羅するように映像を分割するものである．第三象限に分離される先行研究はなく，本論文において新たな技術開発が必要不可欠である．

2.1.3 映像選択の先行研究

複数の視点を有する映像を提示するにあたり，視聴者の視点切替えを代替する手法として，視点や視線を自動的に推薦する研究が行われている．第 1.3.2 項に記すように，視点や視線を自動推薦する関連研究は，「履歴に基づく手法」と「ルールに基づく手法」,「機械学習に基づく手法」の 3 つに分類することができる．

- **履歴に基づく手法**

  1 つの映像を複数の視聴者が視聴することを前提に，新たな視聴者の視聴において，複数の視聴者が過去に視聴した履歴から映像の重要なシーンを抽出して提示する手法である．多くの視聴者が視聴したフレームを連続的につなぎ合わせることで，一般的に興味度の高い視点系列を推薦することが可能である．一方で，視聴履歴の質と量に依存するため，視聴履歴が少ない映像では，視点系列の推薦精度が低下する．

  Mase ら [12] は，多視点映像の視聴履歴を用いて視聴支援を行うシステムを開発した．このシステムでは，視聴履歴の要約をヒートマップで表現し，多くの視聴者が視聴した視点をハイライト表示することで，見どころを簡単に提示している．また，多くの視聴者が視聴した視点をつなぎ合わせることで，視点系列を作成している．

- **ルールに基づく手法**

  画像から抽出した画像特徴量を用いて，映像文法などを参考に映像を数値評価し，評価の高い視点や視線の映像を選択することで，映像の重要なシーンを推薦する手法である． 画像特徴量の算出さえできれば，如何なる映像に対しても同基準で視点推薦が可能である．しかし，コンテンツ毎，被写体毎，個人の嗜好毎など，映像に応じて異なるルールの生成が必要である．

  Shen ら [13] は，Quality-of-View（QoV）と呼ぶ，画像の数値評価方法を提案した．これは，画像に映る被写体の位置，角度，顔の向きなどの画像特徴量に基づき，画像の良さを数値評価する手法である．また，Z.Wang ら [14, 15] は，陸上の短距離走を撮影した映像に対し，選手の視認性を高めるため，表示する画面上での選手のサイズが常に同じ大きさになるように，画像を拡大縮小して表示する手法を提案

している．これにより，画像空間において仮想的に視点や視線を変化させたような映像を推薦している．

- **機械学習に基づく手法**

  機械学習に基づく手法は，履歴に基づく手法とルールに基づく手法を組合わせた手法である．具体的には，画像特徴量に対し，視聴者の嗜好を表現する重みを定義し，視聴者の視聴履歴を正解データとしてこの重みを学習する．そして，視点系列を推薦する際には，画像から得られる画像特徴量に事前学習しておいた重みを掛け合わせることで各画像を数値評価する．そして，最も点数の高い視点を時系列に沿って順次選択することで推薦する視点系列とする． 視聴履歴から個人と万人とに関係なく重みさえ求めることができれば，対象に応じた視点系列を生成することができる．一方で，正確な重みを求めるためには，履歴に基づく手法と同様に，膨大な量の視聴履歴が必要となる．

  Muramatsu ら [16] は，サッカーの試合を撮影した多視点映像を対象に，個人の視聴者向けの視点系列を生成する方法を提案している．画像上での選手位置や大きさといった画像特徴量を用いて，被験者の多視点映像編集履歴を正解データとし，SVM により視聴者の視聴傾向を学習している． Wang ら [17] は，サッカーの試合を撮影した多視点映像を対象に，画像上での選手位置や構図といった画像特徴量を用いて，複数人の被験者の多視点映像編集履歴を正解データとすることで，複数人に共通した視聴傾向を学習している．視点系列を推薦する際には，学習器で全視点の映像を評価し，評価値の高い視点を推薦する． Wang ら [18] は，サッカーの試合を撮影した多視点映像を対象に，視聴者の視聴履歴から視聴者を分類する手法を提案し，グループ毎に学習することで視点推薦の精度向上を図っている． 北原ら [19] は，サッカーの試合を撮影した自由視点映像を対象に，視聴者の嗜好に応じて仮想視点の位置を自動的に決定する手法を提案している．視聴者が映像を視聴する前に，シュートなどのサンプルイベントを撮影した複数の自由視点映像を視聴させる．イベントの被写体と視聴者が映像視聴時に設定した仮想カメラ位置との関係性を特徴量として，視聴者の視聴履歴を正解データとすることで，視聴者の視聴傾向を学習している．そして，実際に映像を視聴する際には，サンプル映像で視聴者が選択した被写体とカメラ位置との関係性に最も近い仮想カメラ位置を視聴者に提示する．

以上より，複数の視点を有する映像の自動視点推薦について，様々な研究が行われている．履歴に基づく手法は，複雑な処理を必要としない代わりに，膨大な量の視聴履歴を必要とする．ルールに基づく手法は，映像毎に映像の特性に応じた新たなルールが必要となる．一方で，機械学習に基づく手法は，全映像に共通したルールを定め，ある単位で共通

の重みを生成することで，汎用性高く視点系列を生成することができる．しかし，履歴に基づく手法 と同様に，重みを算出するためには，膨大な視聴履歴が必要となる．そこで，少量の視聴履歴から重みを算出可能な機械学習に基づく自動視点推薦方法が必要となる．

2.2　複数の視点を有する映像の提示

第 2.2 節では，複数の視点を有する映像の「提示」に関して，「視聴インタフェース」についてまとめる．

視聴インタフェースとは，複数の視点を有する映像を視聴するためのユーザインタフェースの事である．映像表示部と「視点」または「視線」の切替え部，あるいはその両方から構成される．視点や視線を視聴者の意思に基づき切替えることができ，選択した視点や視線からの映像を視聴することができる．これにより被写体を好きな位置や角度から視聴することができる．

第 1.3.1 項に記したように，先行研究は，視聴者の視点選択の複雑さを解消するために，「視点と視線の操作性を向上する研究」と「被写体の視認性を向上する研究」の 2 つに分類することができる．

2.2.1　視点と視線の操作性を向上する研究

東海ら [9] は，多視点映像を視聴する際に，視聴対象である被写体を常に画面の中央に固定表示する釘付け視聴方式を提案した．本方式は，被写体が画像中央に一定サイズで表示されるように，フレームごとに画像の拡大縮小と平行移動を行う．これは，被写体に視対象点を固定することで，視点（x 座標，y 座標，z 座標）と視線（x 軸回転，y 軸回転，z 軸回転）の計 6 自由度の選択操作のうち視点の 3 自由度の選択を，画像の拡大縮小と平行移動を自動的に行うことで仮想的に代替しているとみなすことができる．

N.Inamoto ら [6] は，ヘッドマウントディスプレイ（HMD）を用いてサッカーの自由視点映像を Augmented Reality(AR) として視聴可能な「箱庭スタジアム」と呼ぶシステムを作成した．視聴者が，HMD を着用し，実世界の机上に設置したサッカーフィールドの模型を見ると，視聴者の頭部位置と角度に応じて，そこから見えるサッカーの試合映像が重畳されて見える．これにより視聴者は，サッカーの試合の映像を好きな位置や角度から視聴することができる．この視点選択の操作性は，サッカーフィールドを覗き込むという頭部ジェスチャを利用しており，視聴者にとってとても直感的である．しかし，映像を拡大して視聴したい場合など視点の位置や視線の角度によっては，フィールド模型に対する視聴者の姿勢維持が困難な場合がある．

Zhenli ら [7] は，リアルタイムに 3D 映像を伝送するシステムを開発した．このシステ

ムは，ディスプレイに対する頭部位置から見える被写体の映像を表示する．そのため，視聴者が頭部を移動すると常にその位置からの映像が表示されるため，被写体を 3D 映像として視聴することが可能である．利点として，被写体に対し「回り込む」といった視聴者にとって直感的な操作により視点を切替えることが可能である．しかし，社交ダンスのような狭視域角映像を対象としており，サッカーのような広視域角映像では，被写体とカメラとの距離が遠くなるため，頭部移動程度の視点変化では効果が薄いと考えられる．そのため，広視域角映像の場合は，大きな距離を移動するような視点切替え方法が必要であると考えられる．

渡邊ら [8] は，サッカーの試合を撮影した自由視点映像を対象に，仮想カメラの視点と視線を直感的に選択可能なインタフェースを開発した．机上に俯瞰サッカーフィールドの模型を用意し，この上に視点位置を表現するマーカと視線方向を表現するマーカを配置することで視点と視線を変更することができる．これにより，3D 空間において視点と視線を表現する 6 次元の変数選択を直感的かつ正確に行うことができる．

以上より，先行研究では，視点と視線の計 6 次元の変数選択において，画像の平行移動と拡大縮小を疑似的な視点切替えとみなし自由度を下げる方法や，視点と視線を直感的に制御する様々なユーザインタフェースが研究されている．特に後者については，普段我々が日常的に行う「覗き込む」や「回り込む」といった視認動作により視点切替えを行う方法や，広視域角映像に対し俯瞰映像を利用することで，視点位置と視線方向を直接的に指定する方法が有効であるとまとめることができる．

2.2.2　被写体の視認性を向上する研究

東海ら [9] は，複数の視点を有する映像を視聴する際に，視聴対象である被写体を常に画面の中央に固定表示する釘付け視聴方式を提案した．本方式は，被写体が画像中央に一定サイズで表示されるように，フレームごとに画像の拡大縮小と平行移動を行う．これにより，視点を切替えても常に視聴対象が画面の中央に表示されるため，視聴者は視聴対象を見失うことなく映像を視聴することができる．

丸谷ら [10] は，web 配信可能な多視点映像ストリーミング視聴システムを開発した．このシステムのインタフェースは，視聴映像と共に，俯瞰フィールド上に撮影カメラとそのカメラで撮影されたサムネイル映像を表示している．これにより，3D の被写体が 2D の各視点画像上でどのように見えるのかという 2D と 3D の関係性を容易に知ることができる． しかし，自由視点映像のように視点数が増えた場合に，全視点のサムネイル映像を表示・確認することはできないため，見どころとなる視点を自動で選出しサムネイル表示する新たな手法が必要となる．

S.Gondo ら [11] は，スポーツの戦略分析を支援するシステムを開発した．本システム

表 2.3　本論文の「撮影」,「編集」,「提示」

分類	項目	本論文の設定	新規
撮影	撮影する被写体	サッカー の試合	
	撮影する機材	複数台のカメラ（2D）とレンジセンサ（3D）	
	撮影する方法	フィールドを取り囲む固定カメラ配置	
編集	アノテーション情報	2D：画像上での選手の身体動作の変化量 3D：フィールド上での選手位置含む幾何関係	第 4 章
	映像編集	視聴者の嗜好に応じた視点推薦 （重み付き有向グラフ作成の映像分割方法）	第 1.4 節 第 3 章
提示	視聴 UI	俯瞰映像 UI と頭部移動 UI による視点選択	

は，俯瞰視点と選手視点のプレーの様子を動画を用いて同時に確認することができる．これにより，分析者は，試合内容を容易に理解するとともに，選手目線での議論が可能となる．

以上より，先行研究では，被写体の視認性を向上させる様々なユーザインタフェースが研究されている．広視域角映像の視聴において，被写体を画像中央に固定表示するなど 撮影画像中の被写体を強調表示することで 2D の視認性を高める方法や，サムネイル画像の利用により 2D–3D の視認性を高める方法，俯瞰視点画像の利用により 3D の視認性を高める方法などが有効であるとまとめることができる．

2.3　本論文の位置づけ

第 2.3 節では，第 2.1 節と第 2.2 節を踏まえ，本論文の位置づけを記す．

初めに，「撮影」,「編集」,「提示」の 3 つに関して，表 2.3 に本論文の設定を改めてまとめる．「伝送」については，本論文では議論しないため省略する． 表 2.3 より，本論文は，フィールドを取り囲むように複数台のカメラとレンジセンサを固定配置し，これを用いてサッカーの試合を撮影した広視域角多視点映像を対象に，映像から抽出した選手の身体動作の変化量やレンジデータから抽出したフィールド上での選手位置などの情報を元に，視聴者の嗜好の変化に応じた視点を推薦しつつ，専用の UI を用いて視聴者が不便なく自身の嗜好に基づき視点を切替えながら本映像を視聴することを目指す．特に，ローカルな環境に全視点映像を保持した状態で視聴支援を行う「自動視点推薦システム」の構築を目指して研究を行った成果を記す．

以下では，本論文における研究対象コンテンツの選択理由と先行研究との関係性を示す．

2.3.1　研究対象コンテンツの選択理由

2.3.1.1　視点の連続性

複数の視点を有する映像は，視点数の違いによって「多視点映像（離散視点）」と「自由視点映像（連続視点）」の 2 つに分類される．

- **多視点映像（離散視点）** [7, 9, 10, 12, 13, 14, 15, 16, 17, 18, 46, 47, 48, 49, 50, 51, 52, 54, 55, 56]

  多視点映像とは，複数のカメラで異なる方向から被写体を同時に撮影した映像のことである．カメラを用いて撮影した映像であるため，視点と視線は有限である．そのため，3 次元空間において，視点と視線のペアが離散的に分布する映像となる．

- **自由視点映像（連続視点）** [6, 8, 11, 19, 40, 41, 57, 58, 59, 60, 61]

  自由視点映像とは，任意の視点から任意の視線で被写体を視聴できる映像のことである．自由視点映像の生成には，大きく分けて「画像ベースの手法（2D）」[57, 58, 59],「ビルボードベースの手法（2.5D）」[19],「モデルベースの手法（3D）」[60],「光線ベースの手法（4D）」[61] の 4 つの手法が存在する．3 次元空間において，視点と視線は無限であり，視点と視線が連続的に変化する映像となる．

自由視点映像は，視点と視線の自由度が高い点において，多視点映像よりも優れている．しかし，Computer Graphics を用いて作成した自由視点映像では，リアリティが足りない．また，リアリティを求め実写の多視点映像から正確で綺麗な自由視点映像を作成するためには，映像中からの正確な被写体抽出や，オクルージョンによる画像欠損の補間，表情等の細部の再現性向上など解決しなければならない課題が多い．その上，任意の視点映像を生成するためには，未だ高性能な計算機と膨大な処理時間が必要である．したがって，視聴者が視点を切替えながら実時間で自由視点映像を視聴する環境の構築は困難である．

一方，多視点映像は，自由視点映像に比べ，視点切替えにより選択可能な視点数も視線数も格段に少なくなる．しかし，複数台のカメラで撮影した映像をそのまま視聴者に提供するだけで，視聴者は視点を切替えながら映像を視聴することができる．また，4K や 8K などの高解像度映像を撮影するカメラも将来的には安価になると予想されるため，数十台の 8K カメラで被写体を撮影する未来が予想できる．したがって，様々なコンテンツに対し，質の高い映像を撮影でき，容易に展開できるため多視点映像はとても利便性が高いと考えられる． また，視点を選択するという本課題は，いずれの映像形式でも不変で共通で

ある．将来的に自由視点映像が高速に計算可能となり主流になったとしても本課題は有効である．

これらの点を踏まえて，社会的実現性が高いと考えられる多視点映像を本論文では選択する．

2.3.1.2　視線の動的性

複数の視点を有する映像の撮影方法の違いによって「移動カメラ撮影」と「固定カメラ撮影」の 2 つに分類される．

- **移動カメラ撮影** [14, 15, 41, 46, 47, 48, 49, 50, 54, 55, 56]

  移動カメラ撮影とは，カメラの視線方向を被写体の動きに応じて変化させながら撮影する方法である．多視点映像の撮影では，全てのカメラが特定の被写体を追従するように視線を変えながら撮影することで，全ての視点映像にその被写体を映すことが可能である．これにより，視聴者が視点切替えを行っても，特定の被写体を見失うことなく様々な角度から視聴することができる．

  テレビジョン放送のサッカー中継映像では，特定のサッカー選手を追いかけるようにカメラマンが手動で視線を変更しながら撮影を行っている．また，撮影画像から特定の被写体を画像処理により抽出し，その被写体を追従するようにカメラ雲台のパン・チルト・ロールを自動制御しながら撮影する多視点映像撮影システムも研究されている [55, 60]．また，複数カメラ間で情報交換することで多数の対象を同時追跡可能なシステムも研究されている [56]．
- **固定カメラ撮影** [6, 7, 8, 9, 10, 12, 13, 16, 17, 18, 19, 40, 51, 52, 57, 58, 59, 60, 61]

  固定カメラ撮影とは，カメラの視線方向を変化させずに撮影する方法である．したがって，被写体が画角外に移動することがあるコンテンツの場合は，必ずしも全ての視点映像に特定の被写体が映るとは限らない．しかし，視線方向を固定することで，カメラと撮影対象である空間との関係も固定されるので，その空間内にいる被写体位置を把握する上で，視聴者の負荷が低いという利点がある．また，空間内における被写体位置を考えることで，複数の被写体が存在するコンテンツの場合，被写体間の関係性も理解しやすい．

本論文では，複数人の視聴者が個人の嗜好に応じて被写体を選択し，その被写体に注目しながら視点を切替えつつ映像を視聴することを想定している．後者の固定カメラ撮影では，画像を平行移動や拡大縮小し被写体を画面中央に表示する釘付け視聴方式 [9] を併用したり，画像上の被写体をハイライト表示することで，被写体の視認性を高めることが可能である．したがって，本論文では，被写体間の関係性を理解しやすい固定カメラ撮影を選択する．これにより視点切替えという本質的な課題に集中することができる．

2.3.1.3　被写体の広域性

本論文では，撮影対象である被写体が存在し得る空間（以降，被写空間と表記）の規模に応じて，複数の視点を有する映像を「狭視域角映像」と「広視域角映像」の 2 つに分類する．

- **狭視域角映像** [7, 13, 14, 15, 51, 52, 54, 57, 58, 59, 60, 61]

  規模の小さな被写空間を近景から撮影した映像のことである．スポーツでは，レスリングや相撲などの狭いフィールドでプレーするものを分類する．被写空間を取り囲むようにカメラを配置して撮影する場合，比較的カメラを密に配置することが可能である．そのため，視点間で視野の重複が起こりやすい．これに加えて，被写体の数が少ないスポーツが多く，被写体の移動範囲も狭いので，視点切替えにおいて被写体を見失うことが少なく，視認性が高いといえる．
- **広視域角映像** [6, 8, 9, 10, 11, 12, 16, 17, 18, 19, 46, 47, 48, 55, 56]

  規模の大きな被写空間を遠景から撮影した映像のことである．スポーツでは，サッカーや野球などの広いフィールドでプレーするものを分類する．被写空間を取り囲むようにカメラを配置して撮影する場合，比較的カメラを粗に配置して撮影することになる．そのため，視点間で視野の重複が少ない．これに加えて，被写体の数が多いスポーツが多く，被写体の移動範囲も広いので，視点切替えにおいて被写体を見失うことがあり，視認性が低いといえる．

以上より，広視域角映像は狭視域角映像に比べ被写体の視認性が低いため，視点切替えを行う際に被写体を見失いやすく，視点切替え先の画像における被写体間の関係性を想像しにくい． そのため，視点切替えにおける視聴支援の必要性が高い．したがって，本論文では，広視域角映像を研究対象コンテンツとする．

2.3.1.4　被写体の種類と数

サッカーは，選手が手以外の部位でボールを操作し，敵陣のゴールへ入れるスポーツである．以下，2 つの理由から，サッカーを視聴する上で，視聴者の映像に対する嗜好の違いが表れやすいと考え選択する [6, 8, 9, 10, 11, 12, 16, 17, 18, 19, 46, 48].

- **被写体の種類が多い**

  ボールや選手（22 人=11[人/チーム]×2[チーム]）といった被写体の種類と数が多く，サッカーフィールド上に広く分布しているため，視聴者の個性や好みが多様化しやすい．
- **視聴の見どころがフィールド全体に分布**

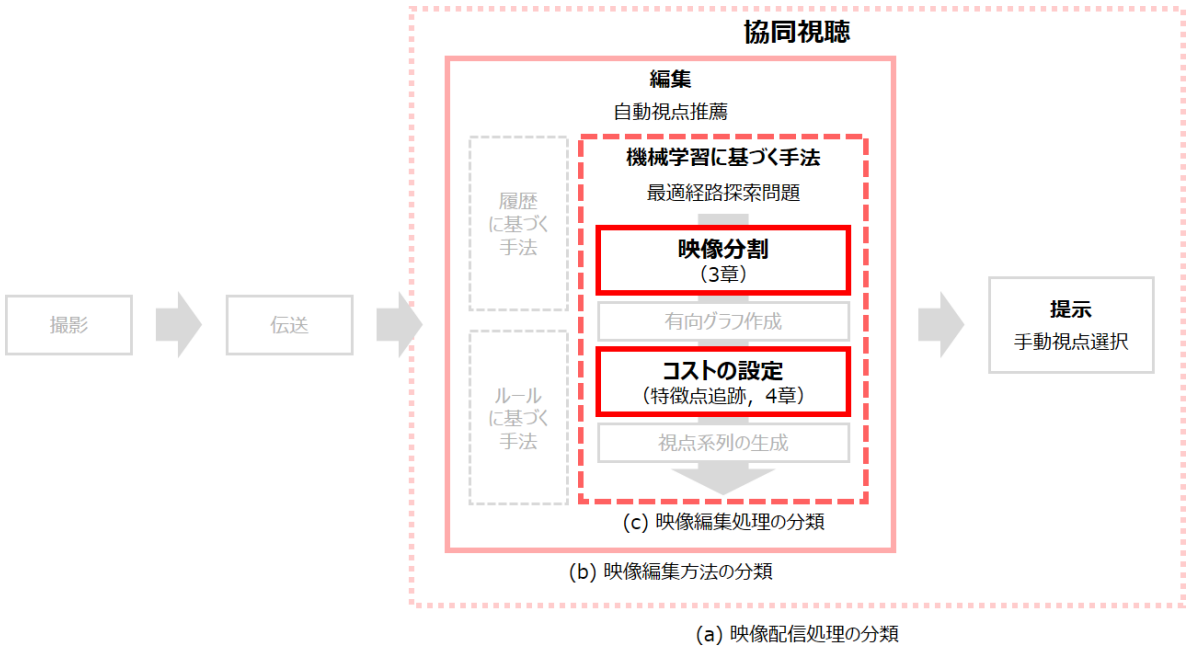


図 2.2　本論文の位置づけ

サッカーは，ボールの所有者が変わることで，試合の攻守が頻繁に入れ替わる．また，選手は，ボールがフィールド上を大きく移動することに備えフィールド全体に分散して位置しており，実際にボールが大きく移動すると，それに応じて選手全員が行動する．そのため，プレーの中心がフィールド内を移動すると共に，それに応じて全ての選手のプレーが変化するため視聴の見どころ が常にフィールド全体に分布する．

2.3.2　本論文と先行研究との関係性

本論文と先行研究との関係性を図 2.2 に示す．前述したように，本論文では主に視聴支援を目的とし，第 1.4 節で述べた複数の視点を有する映像の新たな視聴方式として人とシステムの協調視聴方式を想定している．そして，協調視聴方式の自動視点推薦技術の開発の一環として，第 3 章で映像編集技術のうち映像分割手法を，第 4 章で映像からの情報抽出技術のうち特徴点追跡手法を述べる．したがって，図 2.2 に示すように，本論文は「編集」と「提示」を同時に行う協調視聴方式に対し，「編集」部を実現するための基礎研究であると位置付けられる．本論文の新規ポイントについては，表 2.3 の最右列にも示す．

第 3 章

視点選択を考慮した広視域角多視点映像の分割

3.1　はじめに

3.1.1　背景

我々は，テレビジョンやインターネットを利用し，日常的に映像を視聴している．普段目にする映像の多くは，撮影されたままの元映像ではなく，何かしらの編集を施されたものである．一般的に映像編集とは，一台以上のカメラで被写体を複数回撮影し，撮影した映像群から必要な部分を切り出し，つなぎ合わせることで，新たに 1 つの映像を作成する作業である．そして，画質や音質の調整を行い，場合によっては字幕を挿入し，Computer Graphics などの特殊効果を適用することで，映像の質を向上させる．このように映像編集は，特別な装置やソフトウェアを用いて行う専門的な知識や技術を必要とする作業であり，多くの時間と労力を必要とする．そのため，映像編集の負担を軽減することを目的に様々な研究が行われている．

第 3 章では，映像の編集作業のうち，映像を意味的に連続するフレーム群に分割するカット抽出 [5] について述べる．カットとは，映像を分割する区切りの事である．カット抽出は，編集済みの映像を要約する研究 [46, 47, 48, 49, 50] や素材映像から新たに映像を編集する研究 [51, 52, 53, 54] において議論されている．前者は元映像から主要部分を抽出することに主眼を置き，後者はこれに加え時間的制約を考慮する．各々細かな違いはあるものの全体としては，1 本の映像を生成するために，編集対象である映像に適した情報を利用してカットを抽出する研究であるとまとめることができる．例えばカット抽出に用いる情報には，アクティブなカメラ撮影や映像編集によって変化する画像情報 [47, 51, 52, 53] や，アナウンスや解説などの音声・文書情報 [48, 49]，被写体のコンテキ

スト情報 [50]，映像撮影時に一緒に測定したセンサ情報 [51, 52, 54] などが挙げられる．

さらに，多視点映像の先行研究では，映像分割に関して，多視点映像の視聴支援として視聴者に視点系列を推薦する研究 [17] にて議論されている．この研究では，多視点映像を，時間的連続性を保ちつつ意味のある連続したフレーム群に分割し，最適なフレーム群を選択・結合することで，推薦視点系列を生成している．

3.1.2　目的と研究課題

本論文では，広視域角多視点映像の視聴支援として，映像に対する視聴者の嗜好に沿った視点系列を提示することで，視点選択の負担を軽減し，視聴者が映像視聴に集中できるようにしたい．そこで，広視域角多視点映像を分割し，分割した映像をノード，ノード間を視点切替えを表現するリンクで結んだ 重み付き有向グラフを作成することで，本グラフの最適経路探索問題として視聴者の嗜好との合致率が高い視点系列を算出する．第 3 章では，上記方針に対し，有向グラフの作成に必要な広視域角多視点映像の分割方法について記す．

本目的の達成には，サッカーを固定カメラで連続的に撮影した広視域角多視点映像を，一般的な視聴者が視点切替えをする可能性のある全てのフレームで，過不足なく分割するという課題を解決しなければならない．

多視点映像の視点数を n[個]，フレーム数を $f[frame]$ とすると，有向グラフのノード総数（映像の分割総数）は式 (3.1)，有向グラフのリンク総数（視点切替え総数）は式 (3.2) となる．この時，この有向グラフに対し，Dijkstra のアルゴリズムを用いて最適経路を探索する計算量は式 (3.3) となる．

$$m = n \times f \ \ [個] \tag{3.1}$$

$$e = n^2 \times (f-1) \ \ [本] \tag{3.2}$$

$$\mathrm{O}(e \times \log(m)) \tag{3.3}$$

したがって，視点数およびフレーム数の増加により，有向グラフの規模は容易に大きくなり，計算量は急激に増加する．しかし，視聴支援システムは，視聴者の映像視聴を妨げないように視点系列を提示する必要があり，推薦する視点系列の探索処理を実時間で完了しなければならない．そのためには，有向グラフの規模を小さくする必要があり，広視域角多視点映像を過剰に分割してはいけない．

また，開発する視聴支援システムを用いて，様々な視聴者に，各々の嗜好に沿った視点系列を提示したい．そのためには，視聴者が視点切替えをする可能性のある全てのフレームで映像を分割する必要がある．しかし，映像要約や映像編集に関する先行研究において，様々な映像分割法が提案されているが，多くの視聴者の視点切替えに対応できるよう

に映像を不足なく網羅的に分割することに着目した方法は存在しない．したがって，映像編集に関する専門的な知識を持たない一般的な視聴者が広視域角多視点映像に対してどのような嗜好を持ち，どのタイミングで，どの視点から，どの視点に視点切替えを行うかに関して知見が不足しており，これらを調査する必要がある．

以上より，広視域角多視点映像の過不足のない分割は，トレードオフの関係にあり，如何にして両立させるかが重要な点となる．視聴支援システムを実現する上で，本課題を解決しなければならない特有の課題であると考え，次の研究課題を設定する．

- 一般的な視聴者が視点切替えを行うフレームで映像を分割できること．
- 多くの視聴者の視点切替えに対応できるように不足なく映像を分割すること．
- 実時間での映像生成を目指し有向グラフの規模を小さくするため，過剰に映像を分割しないこと．

3.1.3 基本戦略

第 3.1.2 項に示す課題に対し，次に示す 3 つの基本戦略を立てる．

1. 視聴者の視聴傾向の分析
   多視点映像に対する視聴者の嗜好の変化を調査するために，被写体と複数の視点との関係を同時に考慮することができる幾何関係に着目して，視聴者の視点切替えを分析する．
2. 分析結果に基づく視点切替えフレームの抽出
   多くの視聴者の視点切替えに対応するために，視点切替え前後の視点と被写体との距離および角度に基づき，視点切替えの候補となるフレームを全て抽出する．
3. 視点切替候補フレームの統合
   細かく分割しすぎた映像を意味のある単位にまとめ直すために，時間制約を考慮して，前段で抽出した視点切替え候補フレームを統合する．

具体的には，視聴者の視点切替えの分析では，サッカーの試合を異なるカメラ配置で撮影した 2 種類の広視域角多視点映像を用意する．これらの映像を複数の一般的な視聴者に編集させ，時間的に連続した一本の映像を作成させる．そして，編集された映像に対し，同じ視点を選択し続けた時間，視点切替えを行ったフレーム，選択した視点映像に映る被写体と視点との幾何関係について分析する．次に，視点切替え候補フレームの抽出では，視点と被写体とを結ぶ直線とカメラ光軸とのなす角度に着目して視聴者が視聴対象とするフレームを抽出する．そして，視点切替え前後の 2 視点間の角度および視点と被写体との

距離に着目して，視聴対象フレーム群の中から視点切替え候補フレームを抽出する．最後に，視点切替え候補フレームの統合では，抽出された候補フレームに対し，連続する候補フレームの時間差が，前段の視点切替えで分析した同じ視点を選択し続けた時間以上となるように映像の時間的連続性を保証する範囲内でフレームを統合する．

以上より，広視域角多視点映像を視聴する一般的な視聴者の視点切替えを考慮し，過不足なく映像として意味のある単位での映像分割が可能となる．

本章の構成は次のようである．第 3.2 節で多視点映像を編集した映像の分析に基づく映像分割手法について提案し，第 3.3 節で評価実験について述べ，第 3.4 節でまとめる．

3.2　被写体とカメラ間の関係性を用いた映像分割方法

第 3.2 節では，提案する多視点映像の分割方法について記す．本論文では，複数の一般人に撮影条件の異なる 2 種類の多視点映像を編集させ，自由な視点切替えに関するデータを収集した．次に，俯瞰フィールド上における被写体とカメラとの関係 (距離と角度) に着目し，両データに共通する視点切替えの特徴について分析した．最後に，分析結果に基づき，多視点映像を過不足なく分割する方法を提案する．

以降では，第 3.2.1 項で多視点映像データセットについて説明し，第 3.2.2 項で多視点映像の編集方法について述べ，第 3.2.3 項で多視点映像の編集結果を分析し，第 3.2.4 項でボールに着目して多視点映像の視点切替えを分析し，第 3.2.5 項で分析結果に基づく映像の分割方法について提案する．

3.2.1　多視点映像データセット

本論文では，サッカーの試合を広視域角多視点映像のコンテンツとして，東京都赤羽スポーツの森公園競技場で行われた高校生による練習試合 (以降，赤羽と表記) と愛知県豊田スタジアムで行われた大学生による親善試合 (以降，豊田と表記) を，複数台のカメラとレンジセンサで撮影した．測定機器の仕様を表 3.1 に示す．カメラで，サッカーの試合映像を撮影し，レンジセンサで，フィールド上の選手の位置を測定する．撮影に用いたレンジセンサは，半円状に赤外線を照射し，その反射波がセンサに届くまでの時間を距離に変換するセンサである．

赤羽での撮影におけるカメラとレンジセンサの配置を図 3.1 に，豊田での撮影におけるカメラとレンジセンサの配置を図 3.2 に示す．　図 3.1 および図 3.2 は共に，メインスタンド向かって左奥のコーナーを原点とし，サッカーフィールドのサイドラインと平行な方向を x 軸，タッチラインと平行な方向を y 軸とする 2 次元フィールド座標系を表す．

表 3.1　測定機器の仕様

	Camera	Range sensor
Model	CASIO EX-F1	SICK LMS511
Frame Rate	30 fps	25 Hz
Resolution	1,920 pixel × 1,080 pixel	0.1667 °
Angle of view	36 mm	-5 °∼185 °
Distance	-	65 m
Number	20	2

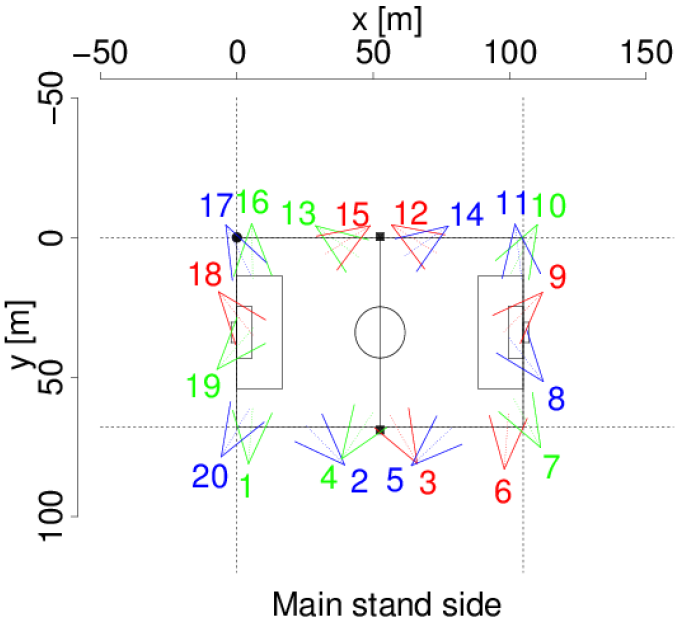


図 3.1　赤羽におけるカメラとレンジセンサの配置

フィールドサイズは，105 m ×68 m であり，図中にて原点を黒丸 (●) で描画する．レンジセンサは，両サイドラインとセンターラインの交点から，赤羽は 30 cm から 1 m 程度離した位置に，豊田は 5 m 程度離した位置に，それぞれ選手の腹部周辺に赤外線が当たるように地面から約 90 cm の高さに設定し設置した．図中では，レンジセンサを黒四角 (■) で描画する．カメラは，フィールドを取囲むように複数台設置した．図中では，各カメラの番号を数字で表記し，カメラ位置からカメラの光軸を点線で，カメラの画角を V 字の実線で描画する．赤羽は，メインスタンド側中央に位置する 6 台のカメラ (カメラ番号：1∼6) を 2 階くらいの高さに，残りのカメラをフィールド上に配置した．そのため，赤羽はフィールド近辺の低い位置からの撮影である．一方豊田は，全カメラをスタジアムの 2

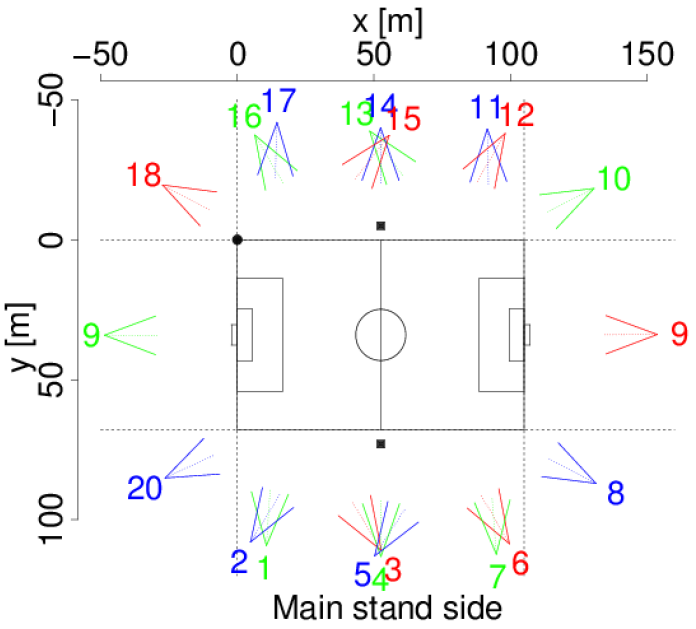


図 3.2　豊田におけるカメラとレンジセンサの配置

階席に配置した．そのため，豊田はフィールド遠方の高い位置からの撮影である．実際の撮影においては，各選手への事前注意と各撮影機材への保護人員の配備により，安全な環境で実施した．

取得した多視点映像とレンジデータから，フィールド座標系におけるボール位置 $p_b(t) = \{x, y\}$ を目視確認により求め，カメラパラメータとして 2 次元フィールド座標系に射影したカメラ位置 $p_c(n)$，カメラ光軸の単位方向ベクトル $\mathbf{L}_c(n)$，カメラの画角 $\alpha_c(n)$ を算出した．ここで，t はフレーム番号を表し，n はカメラ番号を表す．実際に取得したデータは，赤羽が 28,000 frame (約 15 分半) で豊田が 21,894 frame (約 12 分) であった．

図 3.3 に，赤羽で測定したレンジデータの例を示す．図 3.3 は，俯瞰フィールド座標系におけるカメラ位置，レンジセンサ位置，選手位置を表す．カメラの位置と向きは，図 3.1 と同様である．レンジセンサ (Range sensor) は，サイドラインとセンターラインの交点付近に配置した．フィールド上の円弧が各レンジセンサの測定範囲を表している．選手は複数の点集合として測定されるため，各レンジセンサが測定した選手の点群を描画している．また，図 3.4 に，赤羽で撮影した多視点映像の例を示す．図 3.4 は，20 視点分のスクリーンショットである．視点番号は，図 3.1 や図 3.3 の番号と対応する．

次に，具体的な処理手順を記す．

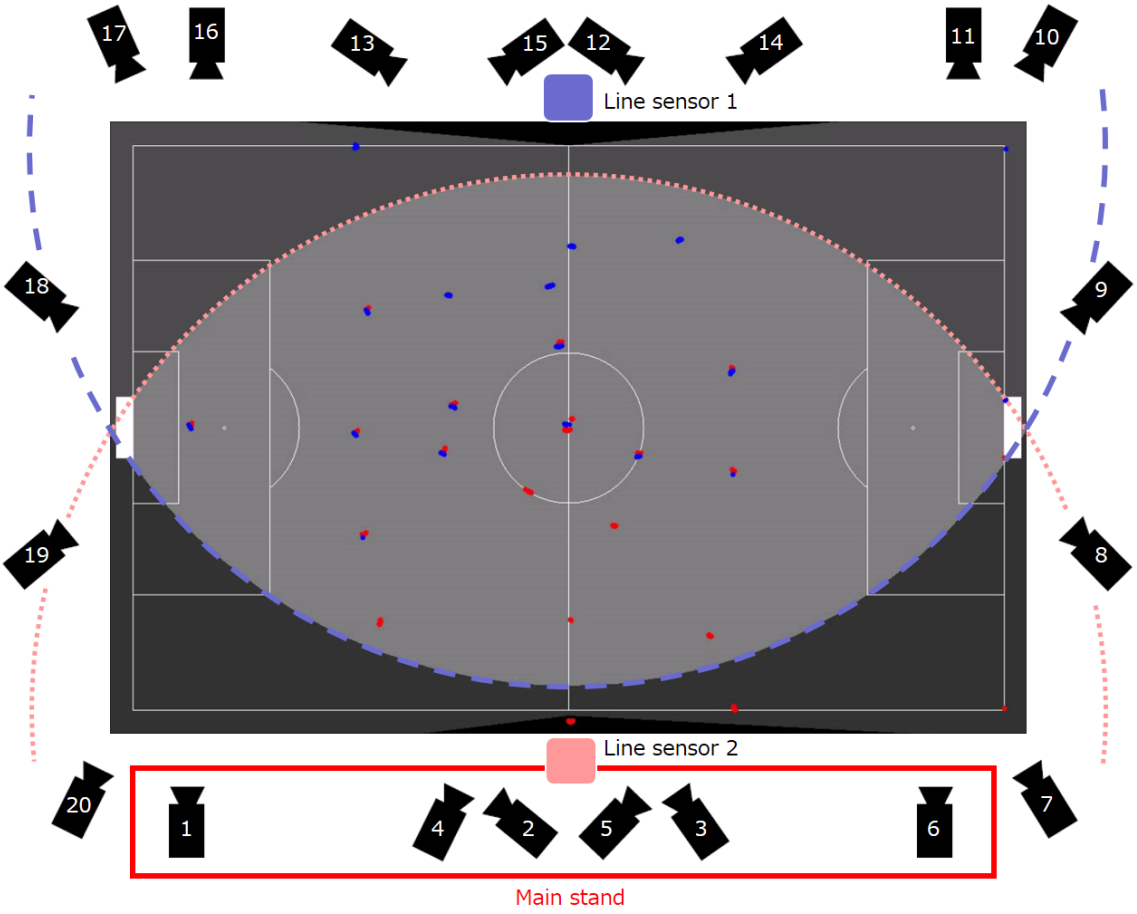


図 3.3　レンジセンサデータの例（赤羽）

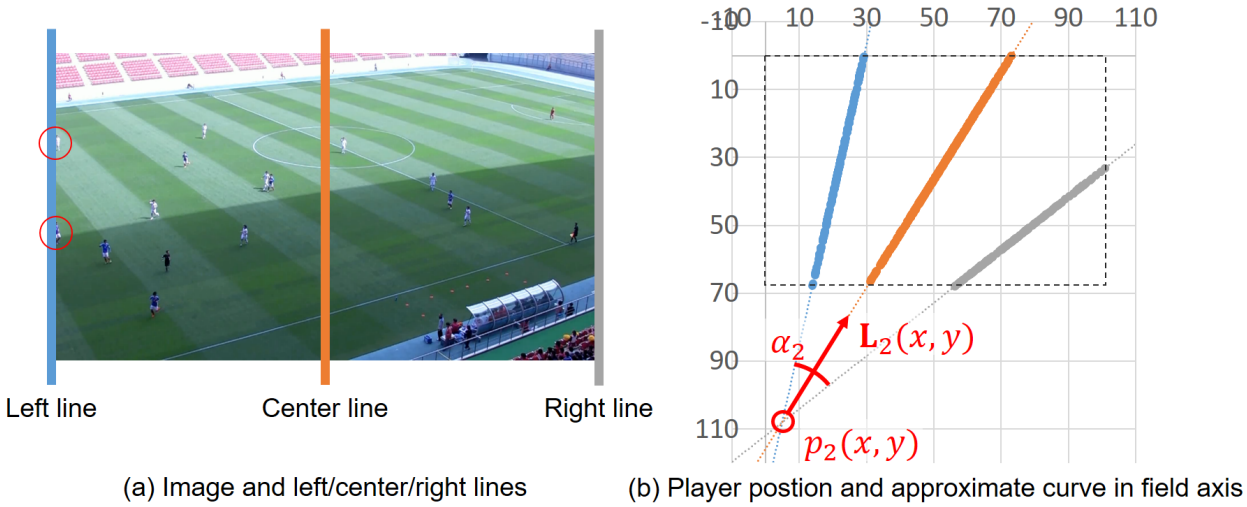
3.2.1.1　フィールド座標系におけるボール位置の算出

はじめに，ゴールポストやコーナーフラッグなどのフィールド上で位置が固定している物体を基準に，フィールド座標系と画像座標系の変換を行うホモグラフィ行列を算出する．次に，レンジデータからフィールド座標系での選手位置を算出する．その後，ホモグラフィ行列を用いてフィールド座標系での選手位置を画像座標系に投影し，画像座標系での選手位置を求める．最後に，画像上でボールが地面に触れる瞬間や選手が地面上のボールを蹴る瞬間を目視で確認し，その位置を手動で求め，ホモグラフィ行列を用いてフィールド座標系に逆投影する．投影点の間を線形補完し，フィールド座標系でのボール位置 $p_b(t)$ を算出する．

3.2.1.2　フィールド座標系におけるカメラパラメータの算出

カメラパラメータ算出の例として，豊田のカメラ 2 を図 3.5 に示す． 図 3.5 は，左側に

図 3.4　多視点映像の例（赤羽）



(a) Image and left/center/right lines　(b) Player postion and approximate curve in field axis

図 3.5　カメラパラメータの算出

カメラの撮影画像を，右側に 2 次元フィールド座標系を示す．

はじめに，図 3.5(a) に示すように，画像座標系での選手位置を利用して，画像の左端列 (Left line) 及び中央列 (Center line)，右端列 (Right line) を通過する選手を自動的に検出する．これにより，画像座標系で検出した選手に対応する，フィールド座標系での選手位置が得られる．図 3.5(b) は，図 3.5(a) の中央および両端の列を通過した選手のフィールド座標系での位置を点群で表している．次に，フィールド座標系での選手位置を表す点群に対して，列毎に最小二乗法を適用し 2 次の近似曲線を算出する．最後に，カメラパラメータ

表 3.2 抽出したシーンの性質の集計

Game		Akabane	Toyota
Scene		17	10
Length [frame]	Max	1071	1201
	Min	196	901
	Avg	658.1	979.0
	Std	301.1	102.2

として 3 本の近似曲線の交点をカメラ位置 $p_c(n)$，中央の近似直線を光軸の単位方向ベクトル $\mathbf{L}_c(n)$，両端の近似直線のなす角を画角 $\alpha_c(n)$ とする．

3.2.2 多視点映像の編集

視聴者が自らの嗜好に基づき視点を切替えながら多視点映像を視聴する際の視点切替えについて分析したい．そこで，一般的な視聴者が多視点映像を時間的に連続した一本の映像に編集するデータが必要となる．本論文では，第 3.2.1 項で述べた多視点映像データセットを用いて，Muramatsu ら [16] と Wang ら [17] が収集した多視点映像の編集結果を利用する．Muramatsu ら [16] は，赤羽の多視点映像データセットのうち全 20 視点分の映像を利用して編集結果を収集した．一方，Wang ら [17] は，豊田の多視点映像データセットのうちメインスタンド側の 13 視点分 (図 3.2 の 1∼10 と 18∼20) の映像を利用して編集結果を収集した．以下に，[16][17] による，編集作業の詳細を示す．

3.2.2.1 多視点映像の編集対象シーン

被験者の映像に対する興味の誘発と作業負担の軽減を目的に，撮影した多視点映像からシュートやパス，ドリブルなどの特徴的なプレーを行っているシーンを抽出する．これらは，一般的に多角度から視聴したいシーンであり，ボールがフィールド上を大きく移動するため，被験者が視点切替えを行いやすいシーンであると想定している．抽出した編集対象となるシーンの集計結果を表 3.2 に示す．上から順に抽出したシーンの総数 (Scene)，全シーンにおける各シーンのフレーム数の最大値 (Max)，最小値 (Min)，平均値 (Avg)，標準偏差 (Std) を表す．

3.2.2.2 多視点映像の編集用インタフェース

図 3.6 に多視点映像の編集で利用するインタフェース [16] を示す． 図 3.6 のインタフェースは，左上部に視点映像を表示し，映像の左右に配置した矢印ボタン ($\Leftarrow, \Rightarrow$) で視

図 3.6　多視点映像編集用インタフェース [16]

点を変更する．また，映像下部に配置したスクロールバーでフレームを変更する．その下に配置したボタンで映像の再生（▷），停止（□），記録（◯）などを行う．右上部には，選択した視点のリスト (Edit List) が表示され，自身の選択結果を確認することができる．右下部には，俯瞰フィールド画像と試合情報，カメラ情報を表示する．試合情報は，選手とボール位置を色分けして円で描画する．カメラ情報は，カメラの通し番号とともに，カメラ設置位置に楕円を描画する．この時，カメラの光軸方向と楕円の長軸方向を一致させる．また，選択中の視点は，楕円の色を赤色に変更し，そのカメラが撮影しているフィールドの領域をハイライト表示させる．これにより，どのカメラが試合状況を撮影できているのかを直感的に把握できるようにしている．

3.2.2.3　多視点映像の編集作業

「多視点映像の編集結果には視聴者の視聴傾向が現れる」という考えのもと，複数の被験者に第 3.2.2.1 で抽出した各シーンを編集させる．被験者は，赤羽のデータセットが 5 人で，豊田のデータセットが 10 人であり，全部で 15 人である．被験者は，エンターテインメントとして，テレビジョンでサッカーの試合を視聴する程度の一般的な視聴者である．被験者のサッカーのプレー頻度については，最頻者で年 1 回の試合経験程度と，玄人とは言えないレベルである．

具体的な編集作業は，第 3.2.2.2 に示すインタフェースを用いて，多視点映像を視聴し，

表 3.3　編集結果における選択ショットのフレーム長の分布とボール含有率

Game		Akabane	Toyota
Cut	Sum	393	289
Shot	Sum	478	389
	Avg	5.6	3.9
	Std	2.3	1.6
Length [frame]	Avg	132.7	253.5
	Std	94.1	193.8
Ball [%]	Avg	95.0	98.5
	Std	13.7	7.1

個人の判断にしたがい最適だと思う「視点切替えを行うフレーム」と「切替え元の視点」，「切替え先の視点」の 3 つを決定することである．この作業を反復的に実施することで，時間的に連続した 1 本の映像が編集される．

映像編集中は，時間を巻き戻し，映像を反復的に視聴することを許す．さらに，各シーンの編集終了時には，編集結果を再度見返し，編集結果に間違いがないか確認させ，間違いがある場合には修正させる．これらは，リアルタイムに視点を切り替えながら多視点映像を視聴するという，本論文で想定する多視点映像の配信および視聴のイメージとは異なる．しかし，一般的な視聴者が，多視点映像の視聴中に，全ての視点映像を把握しつつ個人の嗜好に沿って最適な視点を選択することは不可能である．そのため，本編集作業は妥当であり，これにより視聴者の嗜好に沿った最適な視点切替えの情報が取得できると考える．

3.2.3　編集結果の分析

第 3.2.3 項では，取得した全 185 本の編集映像 (= 赤羽 + 豊田 = (17 シーン × 5 人) + (10 シーン × 10 人)) について分析する．分析した結果を表 3.3 に示す． 表 3.3 は，上から順に，選択されたカットの総数 (Cut Sum)，分割された選択ショットの総数 (Shot Sum)，シーンごとの選択ショット数の平均 (Shot Avg) と標準偏差 (Shot Std)，選択ショットのフレーム数の平均 (Length Avg) と標準偏差 (Length Std)，選択ショットのボールが映るフレーム数の割合の平均 (Ball Avg) と標準偏差 (Ball Std) をそれぞれ表す．

表 3.3 より，シーンごとの選択ショット数にはばらつきがあり，選択ショットごとのフレーム数にもばらつきがある．また，赤羽と豊田の間にも，選択ショット数や選択ショットのフレーム数に差がある．これらは，カメラ配置やコンテンツの内容の違いにより，被

験者の映像に対する視聴傾向が異なることを表している．一方で，選択ショットのボールが映るフレーム数の割合は，赤羽と豊田で共に 95 % 以上と高い値を示している．これは，カメラ配置やコンテンツの内容の違いにかかわらず，一般的な視聴者は，ボールという特定の被写体に注目して視点を切り替える傾向が強いことを示している．この「ボールに注目して映像を視聴する」という結果は，サッカーのコーチングにおける視線配布について分析した先行研究 [62] における素人の視線配布の傾向と一致している．したがって，エンターテインメントとしてサッカーの試合を多視点映像で視聴する際にも，一般的な視聴者がボールに注目して映像を視聴するという指標は有効であると考えられる．

この分析結果より，以降では，ボールに着目して視点切替えの条件について詳しく分析する．

3.2.4　ボールに着目した視点切替えの分析

第 3.2.4 項では，ボールに着目し，多視点映像の視点切替えについて分析する．映像編集に用いた赤羽と豊田の多視点映像は，図 3.1 と図 3.2 に示すように，フィールドを取り囲むようにカメラを配置している．そのため，地面と水平な方向にのみ視点を切り替えることができる．そこで，ボールの 3 次元的な移動を，フィールド座標系に射影した 2 次元的な移動で近似し，ボールとカメラ間の距離 $d_c(t)$ とボールとカメラの光軸間の角度 $\theta_c(t)$ に着目し，視点切替えを分析する．

$$d_c(t) = \|\mathbf{B}_c(t)\| \tag{3.4}$$

$$\theta_c(t) = \arccos\left(\frac{\mathbf{B}_c(t)\cdot\mathbf{L}_c(n)}{\|\mathbf{B}_c(t)\|}\right) \tag{3.5}$$

$$\mathbf{B}_c(t) = p_b(t) - p_c(n) \tag{3.6}$$

ここで，$\mathbf{B}_c(t)$ は，フィールド座標系におけるカメラ c の位置 $p_c(n)$ とボールの位置 $p_b(t)$ を結ぶベクトルを表す．また，角度 $\theta_c(t)$ は光軸を中心に右回転を正，左回転を負とする．図 3.7 に，ボールとカメラ間の関係を示す．

次に，「選択されたフレーム」と「映像中にボールが映るフレーム」，視点切替え元と視点切替え先の「ボールとカメラ間の距離」と「ボールと光軸間の角度」を表すグラフを図 3.8 に示す． 図 3.8 は，豊田で撮影した多視点映像の 180∼1080 frame（30 秒）にあたるシーンを編集した結果であり，ある被験者の視点切替えを表している．この被験者は，379 frame 目に視点を視点 4(viewpoint # 4) から視点 3(viewpoint #3) へ，405 frame 目に視点 3 から視点 1(viewpoint # 1) へ切り替えている．(1) から (3) は，横軸に時間 $frame$ を，左端の縦軸に距離 (distance)$d_c(t)$ を，右端の縦軸に角度 (angle)$\theta_c(t)$ をとるグラフであり，各線の意味は次のようである．

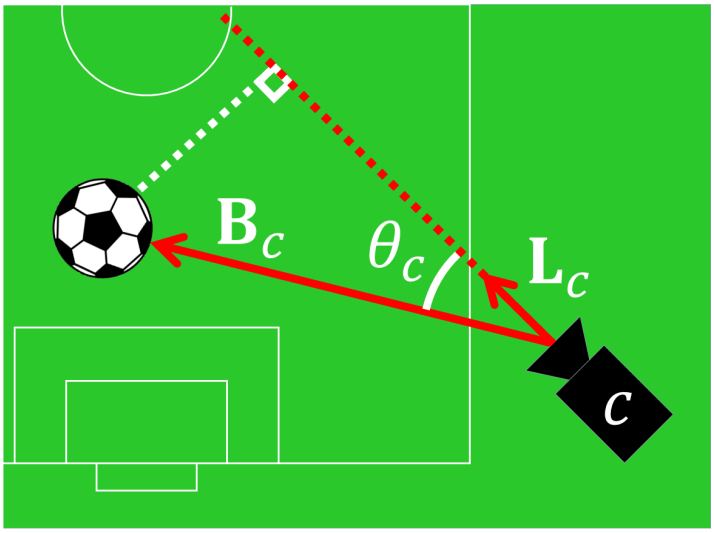


図 3.7　ボールとカメラ間の関係

- 赤背景: 被験者が選択したフレーム
- 青背景: 画像中にボールが映るフレーム
- 黄背景: 被験者が選択した画像中にボールが映るフレーム（赤背景 ∩ 青背景）
- 赤実線: 被験者が視点を切替えたフレーム
- 緑破線: 視点切替え元のカメラとボールとの距離（$d_c(t)$）
- 緑点線: 視点切替え先のカメラとボールとの距離（$d_{\hat{c}}(t)$）
- 緑実線: 距離の差分（$d_c(t) - d_{\hat{c}}(t)$）
- 青破線: 視点切替え元のカメラとボールとの角度（$\theta_c(t)$）
- 青点線: 視点切替え先のカメラとボールとの角度（$\theta_{\hat{c}}(t)$）
- 青実線: 角度の絶対値差分（$|\theta_c(t)| - |\theta_{\hat{c}}(t)|$）

ここで，c は視点切替え元の視点番号を，$\hat{c}$ は視点切替え先の視点番号を表す．そのため，$d_{\hat{c}}(t)$ と $\theta_{\hat{c}}(t)$ は，$d_c(t+1)$ と $\theta_c(t+1)$ と同じ値となる．例えば，図 3.8 の上段の c は視点切替え元の視点 4 を，$\hat{c}$ は視点切替え先の視点 3 を表し，中段の c は視点切替え元の視点 3 を，$\hat{c}$ は視点切替え先の視点 1 を表す．

同様のグラフを，被験者の編集結果ごとに作成し，編集映像の視点切替え条件について分析を行った．その結果，「視聴対象フレーム」と「視点切替フレーム」と呼ぶ二つの傾向を発見することができた．以降では，これらの傾向について詳細を説明する．

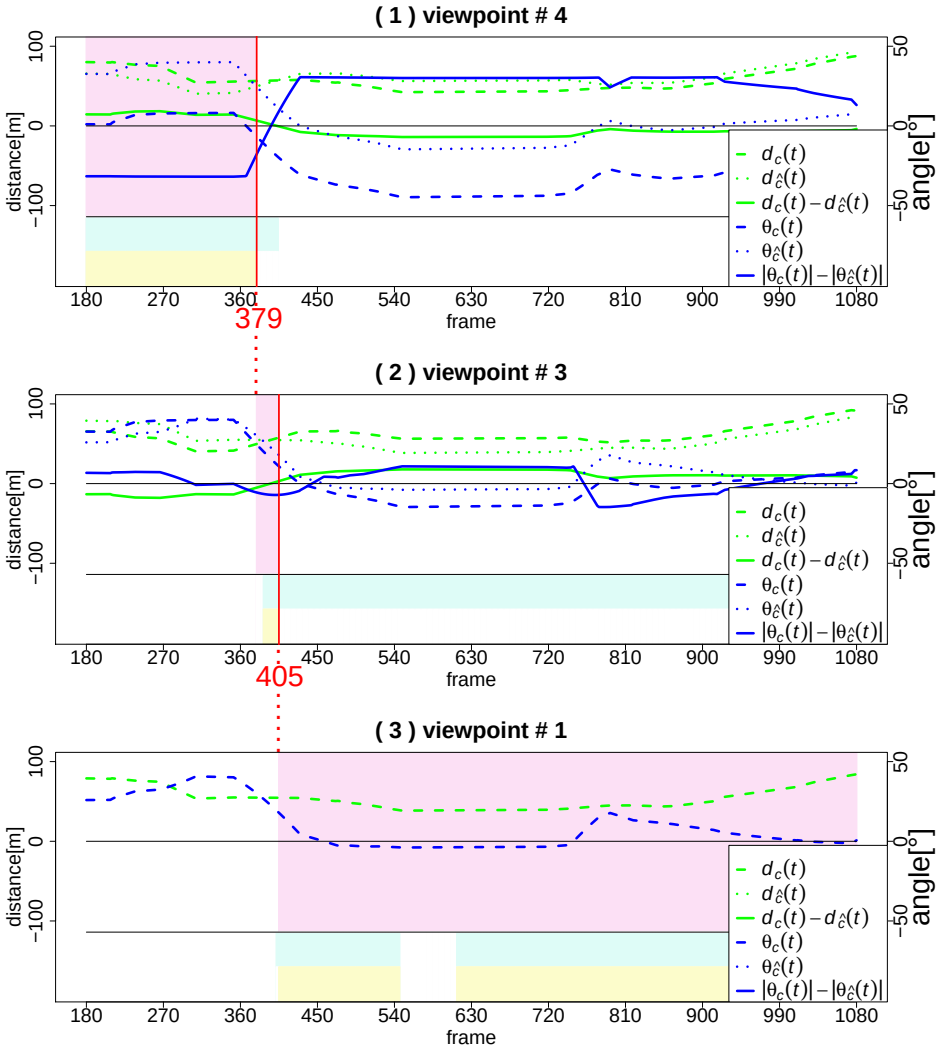


図 3.8　編集結果とボールとカメラ間の関係

3.2.4.1　視聴対象フレーム

視聴対象フレームとは，視聴者が視聴対象とするフレームのことである．表 3.3 より，視聴者はボールが映っているフレームを視聴対象とする傾向がある．一方で，ボールが映っていない残りのフレームについて分析を行うと，ボールがフレームインする直前やフレームアウトした直後のフレームを視聴対象とする傾向にあることが分かった．例として，図 3.8 の視点切替え (視点 4→ 視点 3) におけるフィールド上でのボールの動きを図 3.9 に示す． 図 3.9 は，俯瞰フィールド座標系でのカメラの画角とボールの軌跡を表している．カメラは，カメラの設置位置を始点に，光軸を点線で，画角を V 字の実線で描画する．ボールは，各時刻のボール位置を，選択された視点の色と同じ色で描画する．ボー

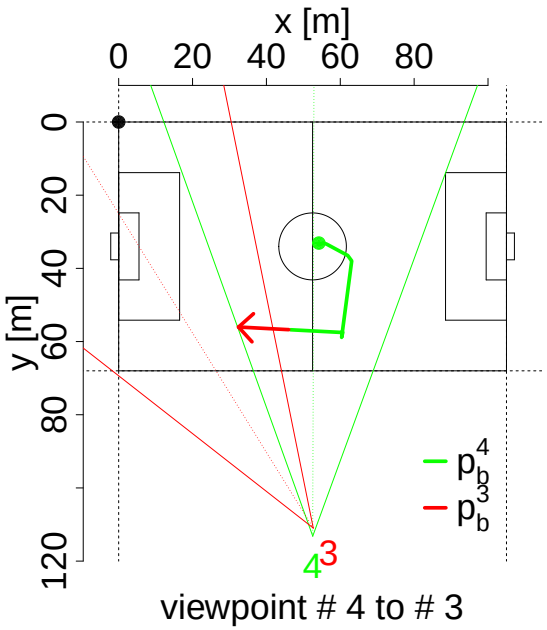


図 3.9　視点切替えとボール位置

ルの位置を時系列として見ると，サッカーフィールド内の太い実線がボールの移動を表しており，開始位置を丸 (●) で，終了位置を矢印の先端 (→) で表現している．図 3.9 を見ると，ボールが視点 3(赤) の撮影範囲内 (V 字領域内) に入る直前で，視点 4(緑) から視点 3(赤) に切替わっていることがわかる．

以上より，一般的な視聴者は，主にボールに着目し，フレームインやフレームアウトなどのボールの動きを念頭に置きながら視点を切替える傾向にあるといえる．つまり，単に画像中におけるボールの有無に着目しているのではなく，ボールとカメラ間の位置関係まで考慮して，視点を切替えていると考えられる．

3.2.4.2　視点切替フレーム

視点切替フレームとは，視聴者が視点切替えを行うフレームのことである．実験者 (27 歳男性) が，図 3.8 に示すようなグラフを目視で確認し，視点切替え時の特徴的な波形を調査することで視点切替え条件を設定した．実験者は，映像編集及びサッカーに対する専門家ではないが，映像編集におけるシーンやショット，カットなどの知識やサッカーにおける基本ルールなどの一般的な知識は持ち合わせていた．次の 4 つが主たる条件として見出された．

(a)　ボールが視点切替え元でフレームアウト／視点切替え先でフレームイン

(b)　視点切替え元か視点切替え先で，ボールが画面中央付近

(c)　視点切替え元と視点切替え先のボールと光軸間絶対値角度が同程度

(d)　視点切替え元と視点切替え先のボールとカメラ間距離が同程度

例えば，図 3.8 における上段の 379 frame は条件 (a) に当たり，視点 3 にボールがフレームインするタイミングで視点を切替えている．また，中段の 405 frame は条件 (d) に当たり，ボールとカメラ間の距離が 2 視点間で等しいタイミング (緑色の実線が 0 付近) で視点を切替えている．

3.2.5　編集映像の分析に基づくカットの抽出方法

第 3.2.4 項の分析結果に基づきボールとカメラ間の関係を利用した 3 ステップからなる新たな映像分割方法を提案する．

3.2.5.1　視聴対象フレームの抽出

一般的な視聴者は，次に示すフレームを視聴対象フレームとしている．

1. ボールが画像に映るフレーム
2. ボールがフレームインする直前のフレーム
3. ボールがフレームアウトした直後のフレーム

1 については，俯瞰フィールド座標系において，カメラの画角（以降，視聴対象領域と呼ぶ）内にボールが位置することを評価すれば判定できる．しかし，ボールが画像に映らない 2 と 3 については，単に同様の方法を適用するだけでは判定できない．そこで，視聴対象領域を拡大することで，これらに対応する．

具体的には，カメラの位置 $p_c(n)$ をカメラ光軸の単位方向ベクトル $\mathbf{L}_c(n)$ と反対方向に r[m] 移動させる．これにより，移動前に比べ，俯瞰フィールド座標系における視聴対象領域が拡大する．そして，拡大後の視聴対象領域内にボール位置 p_b を含むフレーム t を視聴対象フレームとする．これを逆に考えると，カメラに対するボールの相対位置 $\mathbf{B}_c(t)$ をカメラの光軸方向に r[m] 移動させた位置と光軸 $\mathbf{L}_c(n)$ とのなす角 $\theta'_c(t)$ が，式 (3.7) を満たすフレームが視聴対象フレーム $\mathbf{F}_c$ となる．

$$\mathbf{F}_c = \left\{ t_i^c \middle| |\theta'_c(t)| \leq \frac{\alpha_c(n)}{2} \right\} \tag{3.7}$$

$$\theta'_c(t) = \arccos\left(\frac{(\mathbf{B}_c(t) + r\mathbf{L}_c(n)) \cdot \mathbf{L}_c(n)}{\|\mathbf{B}_c(t) + r\mathbf{L}_c(n)\|} \right) \tag{3.8}$$

3.2.5.2 視点切替候補フレームの抽出

視点切替候補フレームは，連続する視聴対象フレーム群から複数抽出される．これは，複数の視点映像にボールが同時に映るため，プレーによっては，選択している視点映像にボールが映り続けていても，他の視点映像の方がボールが見やすくなることがあり，それに応じて視聴者が視点を切替えるのに対応するためである．具体的には， 式 (3.9) から式 (3.15) のいずれかを満たす視聴対象フレームを視点切替候補フレーム $\bar{\mathbf{F}}_c \subseteq \mathbf{F}_c$ とする．

条件 (a) は，式 (3.9) から式 (3.12) で表す．式 (3.9) はボールが画像からフレームアウトするフレーム，式 (3.10) はボールが画像にフレームインするフレーム，式 (3.11) はボールが第 3.2.5.1 で拡大した視聴対象領域から出るフレーム，式 (3.12) はボールが同領域内へ入るフレームをそれぞれ表す．

$$|\theta_c(t)| \leq \frac{\alpha_c(n)}{2} \wedge |\theta_c(t+1)| \geq \frac{\alpha_c(n)}{2} \wedge |\theta_{\hat{c}}(t+1)| \leq \frac{\alpha_{\hat{c}}}{2} \tag{3.9}$$

$$|\theta_c(t)| \leq \frac{\alpha_c(n)}{2} \ \wedge \ |\theta_{\hat{c}}(t)| \geq \frac{\alpha_{\hat{c}}}{2} \ \wedge \ |\theta_{\hat{c}}(t+1)| \leq \frac{\alpha_{\hat{c}}}{2} \tag{3.10}$$

$$|\theta'_c(t)| \leq \frac{\alpha_c(n)}{2} \wedge |\theta'_c(t+1)| \geq \frac{\alpha_c(n)}{2} \wedge |\theta'_{\hat{c}}(t+1)| \leq \frac{\alpha_{\hat{c}}}{2} \tag{3.11}$$

$$|\theta'_c(t)| \leq \frac{\alpha_c(n)}{2} \ \wedge \ |\theta'_{\hat{c}}(t)| \geq \frac{\alpha_{\hat{c}}}{2} \ \wedge \ |\theta'_{\hat{c}}(t+1)| \leq \frac{\alpha_{\hat{c}}}{2} \tag{3.12}$$

条件 (b) から条件 (d) はそれぞれ式 (3.13) から式 (3.15) で表す．被写体が連続的に変化するため条件を満たすフレームが複数存在することがある．そのため， 各式を満たすフレーム t のうち，極値検出用の時間幅 T_e で見たときに極小値となるフレームを視点切替候補フレームとする．

$$|\theta_c(t)| \leq \xi_\theta(d_c(t)) \vee |\theta_{\hat{c}}(t)| \leq \xi_\theta(d_{\hat{c}}(t)) \tag{3.13}$$

$$\mathrm{diff}(\theta_c(t), \theta_{\hat{c}}(t)) \leq \xi_\theta(d_c(t)) \tag{3.14}$$

$$\mathrm{diff}(d_c(t), d_{\hat{c}}(t)) \leq \xi_d(d_c(t)) \tag{3.15}$$

ここで，diff$(\cdot)$ は，式 (3.16) で定義される引数の絶対値差分の絶対値を算出する関数である．$\xi(\cdot)$ は式 (3.17) で定義されるボールとカメラ間の距離 $d(t)$ にしたがい変化するしきい値であり，距離が短いほど小さな値となる．定数 $\tilde{u}$ は，カメラから距離 $\tilde{d}_u$ 離れた位置におけるしきい値 $\tilde{\xi}_u$ と最大しきい値 $\acute{\xi}_u$ を事前に決定することで式 (3.18) より求める．

$$\mathrm{diff}(u_1(t), u_2(t)) = \Big| |u_1(t)| - |u_2(t)| \Big| \tag{3.16}$$

$$\xi_u(d(t)) = \acute{\xi}_u \times \exp\left(-\frac{\tilde{u}}{d(t)}\right) \tag{3.17}$$

$$\tilde{u} = \tilde{d}_u \times (\log\acute{\xi}_u - \log\tilde{\xi}_u) \tag{3.18}$$

3.2.5.3　視点切替候補フレームの統合

視聴者の自由な視点切替えに対応するため，条件 (a) から条件 (d) を用意している．そのため被写体の位置やカメラの配置によって，短時間の間に複数の視点切替候補フレームが検出されることがある．これはフレーム数の少ない部分ショットを多量に生成し，部分ショットの意味的内容の損失や視点推薦における頻繁な視点切替えの原因となる．そこで，近接する視点切替候補フレームを視点ごとに統合する．

はじめに，切替え先の視点が同じ候補フレーム $\bar{\mathbf{F}}_{c\to\hat{c}} \subseteq \bar{\mathbf{F}}_c$ について，隣接する候補フレームの時刻差がしきい値 T_i 以下となる組を，小さいものから順に統合する．統合には条件があり，統合対象となる候補フレーム群の時刻重心 $\bar{t}^c$ が，$\lceil \bar{t}^c \rceil \in \mathbf{F}_c$ となる場合のみ統合を行い，$\bar{t}^c$ を新たな視点切替候補フレームとする．この処理を，統合する候補フレームがなくなるまで繰り返す．その後，全候補フレームで同じ処理を繰り返し，最終的な視点切替フレームを求め，これに基づき映像を分割する．

3.3　映像分割の評価実験

提案手法の有効性を確認するため，映像を分割する基準のフレームである「カット」の抽出精度を評価した．はじめに，第 3.3.1 項で提案手法の全体的な評価として，「視聴対象フレームの抽出」，「カット抽出に用いる特徴量」，「視点切替フレームの抽出」，「視点切替フレームの統合」という 4 種類の要素の組合せについて比較をした．次に，第 3.3.2 項で提案手法の部分的な評価として，多視点映像の編集結果から設定した 4 つのカット抽出条件について比較をした．最後に，第 3.3.3 項で提案手法がカットの抽出条件に用いた編集映像データセット以外のデータセットに対しても精度よくカットを抽出できるか調査を行った．

3.3.1　全体評価実験

第 3.3.1 項では，提案手法が設計通りに機能することを確認するため，一般的な視聴者が視点切替えを行うのと同じタイミングで，カットを抽出できるか評価を行った．

3.3.1.1　全体評価の比較対象

「視聴対象フレームの抽出方法」，「カット抽出に用いる特徴量」，「映像を部分ショットに分割する視点切替フレームの抽出方法」，「視点切替フレームの統合方法」という 4 種類の要素を組合せた 9 種類のカット抽出方法（B-S, B-I-E, B-D-E, B-A-E, B-DA-E, B-DA-R, P-DA-E, P-DA-R, P-DA-R-C）を比較した．

- 視聴対象フレームの抽出方法：**Tar***get frame*
  - B: カメラの撮影領域を視聴対象領域とする手法
  - P: 視聴対象領域を拡大する手法 (提案)
- カット抽出に用いる特徴量：**Fea***ture*
  - I: グレースケール画像のフレーム間差分値
  - D: ボールとカメラ間の距離
  - A: ボールと光軸間の角度
- 映像を部分ショットに分割する視点切替フレームの抽出方法：**Swi***tch frame*
  - S: 時間幅 T_e の等間隔分割手法 (ベースライン)
  - E: 特徴ベクトルの極値検出手法
  - R: ルールベースの手法 (提案：条件 (a) から (d))
- 視点切替フレームの統合手法：**Int***egration*
  - C: 視点切替え可能条件に基づく統合手法 (提案)

B-S は，ボールが画像に映るフレーム区間のみを視聴対象フレーム群とし，視聴対象フレーム群中の視点切替えを表現するため等間隔に区間を分割してカットとするベースライン手法である．B-I-E は，同区間に対し，グレースケール画像のフレーム間差分値を求め，この時系列データの極値をカットとする単純な手法である．B-D-E は，同区間に対し，ボールとカメラ間の距離を求め，この時系列データの極値をカットとする手法である．B-A-E は，同区間に対し，ボールと光軸間の角度を求め，この時系列データの極値をカットとする手法である．B-DA-E は，同区間に対し，ボールとカメラ間の距離及びボールと光軸間の角度を求め，これらの両時系列データの極値をカットとする手法である．B-DA-R は，同区間に対し，ボールとカメラ間の距離及びボールと光軸間の角度を求め，これらの両時系列データから，提案するルールに基づきカットを抽出する手法である．P-DA-E は，視聴対象領域を拡大した後のフレーム区間に対し，ボールとカメラ間の距離及びボールと光軸間の角度を求め，これらの両時系列データの極値をカットとする手法である．P-DA-R は，同拡大領域の区間に対し，ボールとカメラ間の距離及びボールと光軸間の角度を求め，これらの両時系列データから，提案するルールに基づきカットを抽出する手法である．P-DA-R-C は，P-DA-R に視点切替フレームの統合処理を加えた全組合せの提案手法である．

3.3.1.2 全体評価の評価方法

第 3.2.2 項に示す多視点映像の編集結果を正解とし，第 3.3.1.1 に示す各手法で抽出した視点切替フレームとの一致率を評価した．

正解データには，多視点映像データセットを編集した際の視点切替フレームを用

いた．多視点映像データセットは，20 視点分の映像からなる 2 つの多視点映像（赤羽：28,000 $frame$ = 約 15 分 30 秒，豊田：21,894 $frame$ = 約 12 分 16 秒）を用いた．両データセットから抽出したシーンの総数は，27 個であった．また，各シーンを編集した編集映像の総数は，185 本であった．さらに，編集映像の視点切替フレームの総数は，赤羽が 393 frame，豊田が 289 frame であった．以降，このフレームを正解視点切替フレームと呼ぶ．

一致率の評価には，式 (3.19) に示す再現率（recall）を用いた．

$$\text{再現率} = \frac{\text{検出成功視点切替フレーム数}}{\text{全正解視点切替フレーム数}} \tag{3.19}$$

再現率とは，正解視点切替フレームを検出できる割合である．検出成功視点切替フレームとは，各手法で抽出した視点切替フレームのうち，正解視点切替フレームとの時刻誤差が，しきい値以内となるフレームの事である．以降，このしきい値を検出成功許容しきい値と呼ぶ．再現率を高めるためには，視点切替フレームを多量に抽出し，検出成功視点切替フレームの数を増やせばよい．しかし，視点切替フレームの数を増やすと，有効グラフの規模が大きくなり，推薦視点系列を探索するための計算量が急激に増加してしまう．

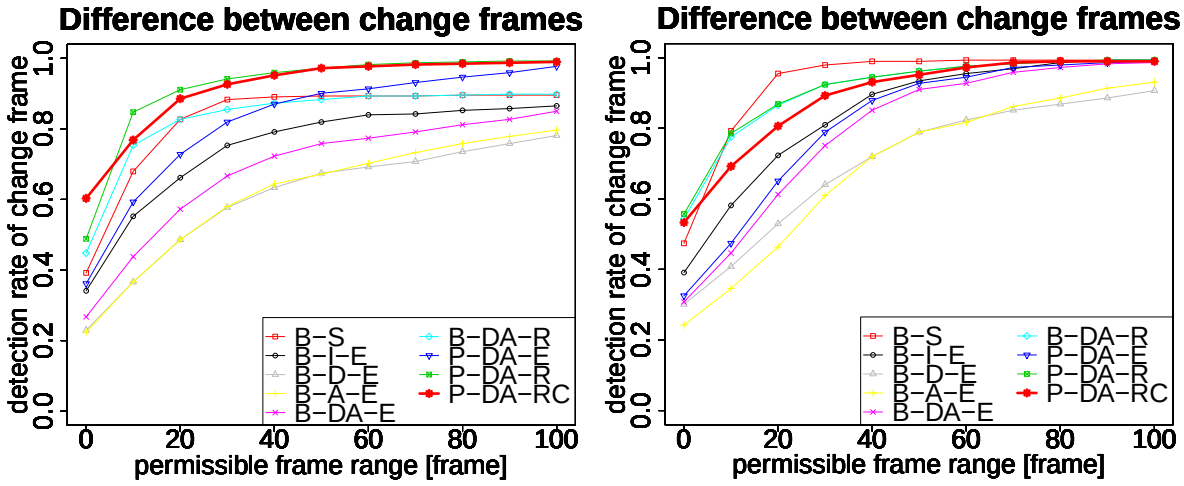
そこで，各手法により抽出したカット数とリンク数についても評価した．カット数とは，各手法により映像から抽出した視点切替フレームの総数である．リンク数とは，抽出した視点切替フレームから切替えることができる切替え先の視点の総数である．切替え元の視点情報のみを利用してカット抽出を行う手法 S や手法 E では，切替え先の画像中にボールが映っていればリンクを結ぶこととした．一方，提案手法 R では，切替え元の視点と切替え先の視点の両方を考慮して視点切替フレームを算出しているため，条件に合致した切替え先の視点のみリンクを結ぶとした．

3.3.1.3　全体評価の実験設定

計算に必要なパラメータを次に示す．第 3.2.1.2 より，俯瞰フィールド座標系におけるカメラの画角 $\alpha_c(n)$ は，カメラごとに異なる値が算出された．これは，フィールド座標系に対するカメラの角度が全て異なるためである．画角の平均と標準偏差は，赤羽が $41.4 \pm 0.6\,^\circ$，豊田が $39.5 \pm 0.7\,^\circ$ であった．

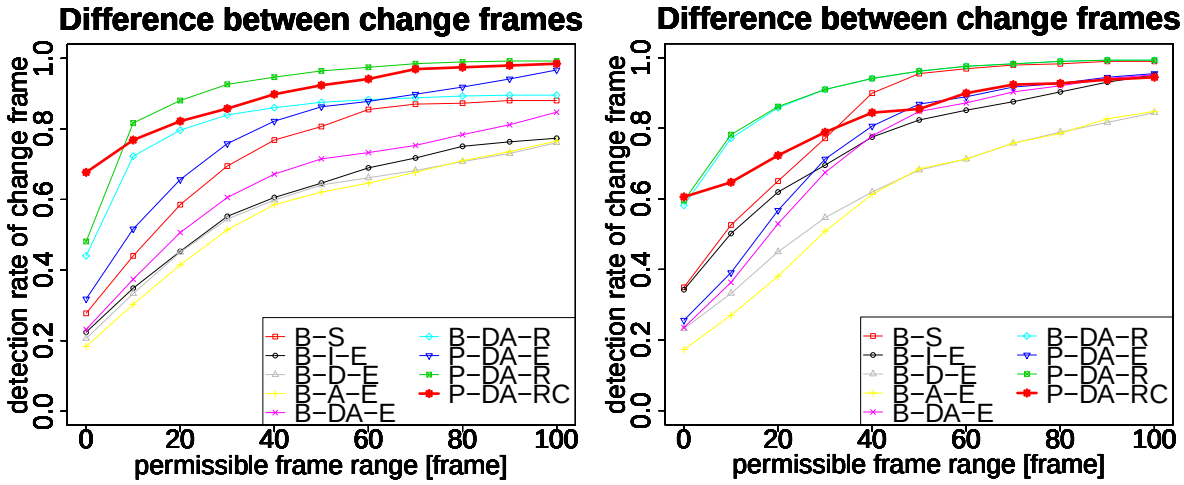
次に，視聴対象フレームの抽出に必要なカメラ位置の移動量 r は，各カメラの画角にしたがいカメラの光軸に対し垂直方向に両端各 5 m ずつ視野を広げるものとした．具体的な r の平均と標準偏差は，赤羽が 5.7±0.1 m，豊田が 6.1±0.2 m であった．

最後に，極値検出に必要なしきい値は，最大しきい値 $\xi_{max} = 2.0$ m で，距離 20 m のときに 1.0 m の誤差を許容するとして，定数 $\bar{u} = 13.9$ とした．また，極値検出の時間幅 T_e と視点切替候補フレームの統合時間幅 T_i は同じ値とし，変数として変化させた．



(a) Akabane　(b) Toyota

$T_e = T_i = 30$ frame



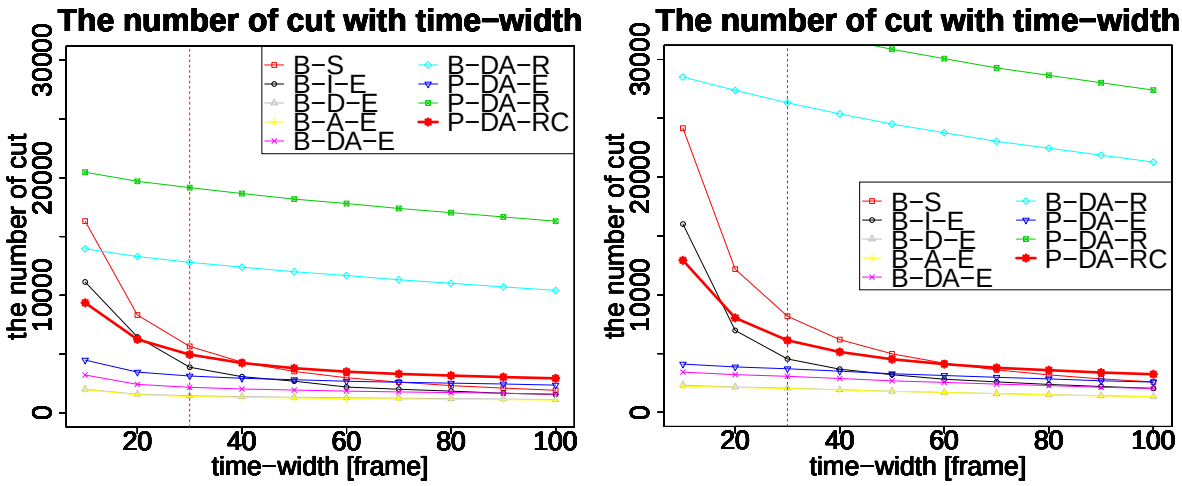
(a) Akabane　(b) Toyota

$T_e = T_i = 90$ frame

図 3.10　検出成功許容しきい値に対する再現率の変化

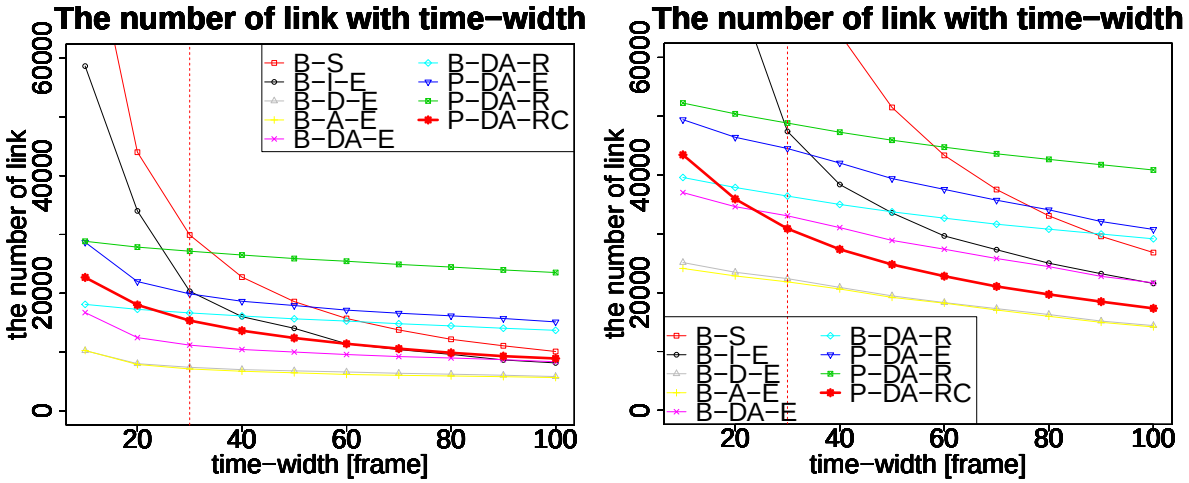
3.3.1.4　全体評価の結果

抽出した視点切替フレームを検出成功と判定する基準である「検出成功許容しきい値」を，変化させた際の再現率の変化を図 3.10 に示す．これは，極値検出の時間幅 T_e と視点切替候補フレームの統合時間幅 T_i が，30 frame(1 秒) と 90 frame(3 秒) の時の結果である．横軸にしきい値を，縦軸に再現率をとる．



(a) Akabane　　(b) Toyota

図 3.11　両時間幅 $T_e = T_i$ に対するカット数の変化



(a) Akabane　　(b) Toyota

図 3.12　両時間幅 $T_e = T_i$ に対するリンク数の変化

また，両時間幅を変化させた際のカット数のグラフを図 3.11 に，リンク数のグラフを図 3.12 に示す．横軸が時間幅であり，縦軸がカット数とリンク数である．赤色の縦線は，時間幅 30 frame を示す． これらの結果を表 3.4 にまとめる．

さらに，多視点映像の編集結果と比較手法による多視点映像の各分割結果を比較した例を図 3.13 に示す．図 3.13 は，赤羽のシーン 1 を被験者 1 が編集した結果である．各グラフは，縦軸に視点番号を，横軸にフレーム数をとる．視点は，被験者 1 が選択した視点 2，視点 3，視点 12，視点 13，視点 14，視点 15 を全 20 視点から抜粋して描画している．フレームは，シーン 1 と定義した 2490 フレームから 3300 フレームまでの 810 フレーム

表 3.4　再現率とカット数とリンク数

$(a) T_i = T_e = 30$ frame

Method	Akabane			Toyota		
Tar-Fea-Swi-Int	Recall	Cut	Link	Recall	Cut	Link
B - - S -	0.883	5,661	29,914	0.979	8,173	84,506
B - I - E -	0.753	3,888	20,346	0.810	4,541	47,419
B - D - E -	0.578	1,464	7,418	0.640	2,061	22,393
B - A - E -	0.580	1,398	7,135	0.609	2,023	21,841
B -DA- E -	0.667	2,180	11,188	0.751	3,066	33,072
B -DA- R -	0.855	12,808	16,665	0.924	26,317	36,431
P -DA- E -	0.819	3,142	19,903	0.789	3,707	44,514
P -DA- R -	0.941	19,158	27,160	0.924	32,763	48,842
P -DA- R - C	0.926	4,965	15,360	0.893	6,124	30,884

$(b) T_i = T_e = 90$ frame

Method	Akabane			Toyota		
Tar-Fea-Swi-Int	Recall	Cut	Link	Recall	Cut	Link
B - - S -	0.880	2,112	11,075	0.990	2,848	29,581
B - I - E -	0.763	1,675	8,669	0.931	2,199	23,228
B - D - E -	0.730	1,185	6,067	0.817	1,412	15,176
B - A - E -	0.735	1,130	5,821	0.827	1,396	14,963
B -DA- E -	0.812	1,677	8,718	0.941	2,132	22,780
B -DA- R -	0.896	10,718	14,058	0.993	21,868	29,978
P -DA- E -	0.941	2,464	15,716	0.945	2,684	32,101
P -DA- R -	0.992	16,665	23,971	0.993	28,026	41,765
P -DA- R - C	0.980	3,051	9,310	0.938	3,387	18,477

(27 秒) 分を描画している．各視点は，視聴対象フレーム（灰色領域）と視点切替フレーム（青色縦線），被験者の編集フレーム（赤色領域），検出成功許容しきい値（赤色縦線）をそれぞれ描画している．したがって，被験者がシーン 1 を，図 3.13 の左から右に向かって，視点 2→ 視点 3→ 視点 13→ 視点 12→ 視点 14→ 視点 13→ 視点 14→ 視点 15 の順で視点を切替えるように編集したことを表している．本結果は，検出成功許容しきい値=30 フレーム（1 秒）の結果であり，検出成功許容しきい値（赤色縦線）は，切替え元の視点に

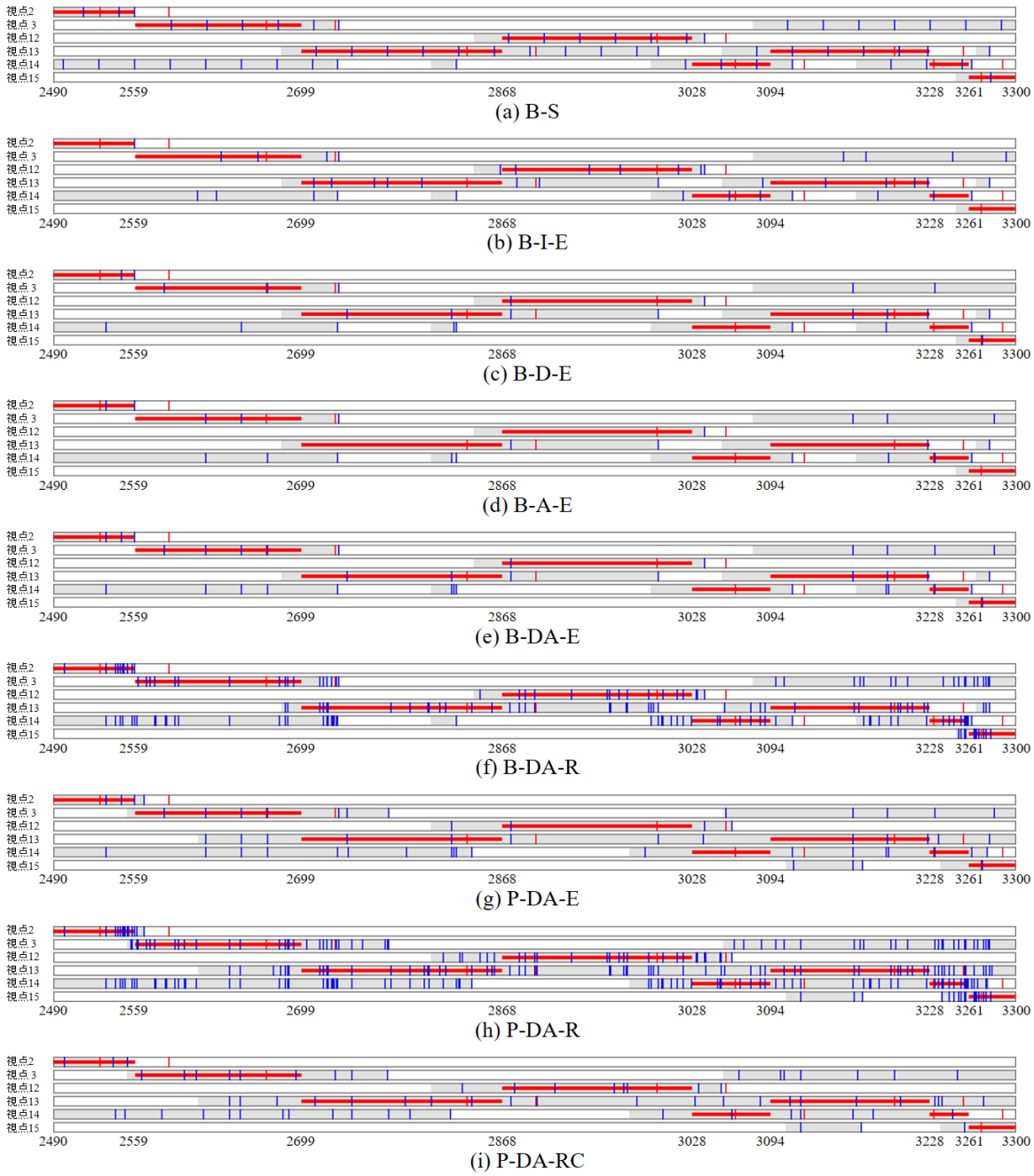


図 3.13　多視点映像の編集結果と分割結果の比較

おける視点切替えの前後 30 フレームの位置に描画している．この赤色の縦線の間に，各手法により検出した視点切替えフレーム（青色縦線）が含まれれば追跡成功とする．

3.3.1.5　全体評価の考察

提案手法の全組合せである P-DA-R-C（太赤）は，表 3.4 より，許容しきい値 30 frame において，赤羽で 0.926，豊田で 0.893 の再現率を示した．また，図 3.10 より，しきい値を変化させた時の全体的な変化に着目すると，赤羽で 2 番目，豊田で 4 番目に高い再現率であった．さらに，表 3.4 より，時間幅 30 frame において赤羽で 4,965 カットの 15,360 リンク，豊田で 6,124 カットの 30,884 リンクであった．一方，図 3.11 と図 3.12 を見ると，図 3.10 で P-DA-R-C よりも高い再現率を示した P-DA-R（緑）や B-DA-R（水），B-S（細赤）の方が，提案手法よりもカット数とリンク数が大幅に多いことを確認できる．

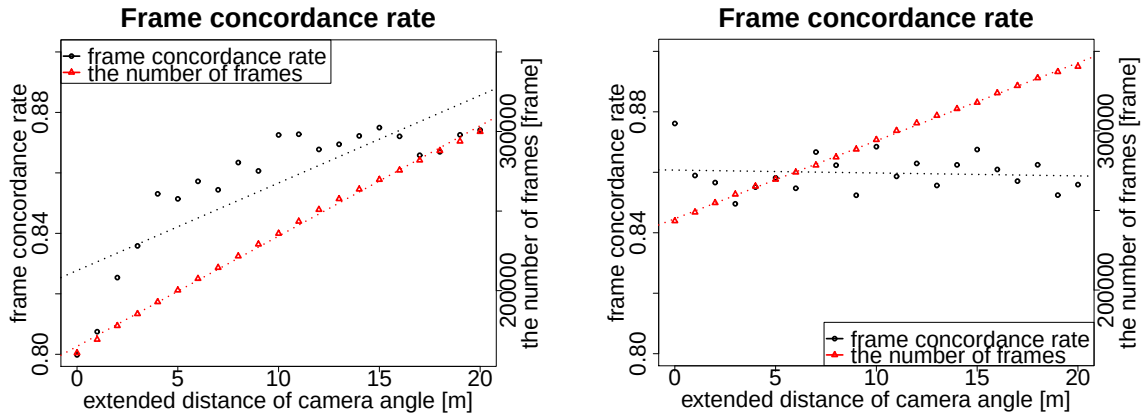
また，図 3.10 と表 3.4 より，再現率が収束したと思われる許容しきい値 90 frame では，赤羽において，P-DA-R-C と B-S(ベースライン) との再現率の差が 0.100 であった．この差は，視聴対象フレームの拡大手法 P や特徴量 DA，ルールベースの視点切替フレームの抽出手法 R に起因していると考えられる．また，同許容しきい値におけるカット数やリンク数も，P-DA-R-C の方が大幅に少なく，提案手法は効果的に機能しているといえる．

また，図 3.13(a) より，B-S が 30 フレーム間隔で映像を分割する手法であり，検出成功許容しきい値内に必ず視点切替フレームを含むことを確認できる．したがって，表 3.4 が示すように高い再現率となる．しかし，視点切替え先の視点を考慮していないので，映像を分割した数だけリンク数が増加してしまう．次に図 3.13(b) から (e) を見ると，他の手法よりも検出した視点切替フレームの数が少ないことがわかる．そのため，他の被験者の視点切替えに対応することができず，表 3.4 が示すように低い再現率となったと考えられる．図 3.13(f) と (h) は，提案するルールベースの手法であり，多くの視点切替フレームが検出されたことを確認できる．このため表 3.4 が示すように再現率は高くなるが，これに応じてカット数とリンク数も増加したと考えられる．図 3.13(i) の P-DA-R-C は，全ての提案手法を組み合わせた手法である．(h) の P-DA-R と比較すると，検出した視点切替フレームの数が減っていることが確認できる．また，視聴対象フレーム全体にわたり，視点切替え先を考慮した視点切替フレームが検出されていることも確認できる．したがって，他の被験者の視点切替えにも対応できるため，少ない視点切替フレーム数で，高い再現率を示すことができると考えられる．

したがって，提案手法の全組合せである P-DA-R-C は，カット抽出の条件作成に用いた多視点映像データセットに対して，他の手法よりもカット数やリンク数の増加を抑えつつ高い再現率を出せる実用的な手法であるといえる．以下に提案手法の詳細な分析について記す．

(i) 視聴対象領域拡大の有無 (B · P) の有効性に関する考察

図 3.10 より，P-DA-E（青）は B-DA-E（桃）よりも高い再現率を示し，P-DA-R（緑）



(a) Akabane　　(b) Toyota

図 3.14　視聴対象領域の拡大に伴う，視聴対象フレーム数の増加とフレーム一致率

は B-DA-R（水）以上の再現率を示している．これより，提案手法であるボールとカメラ間の関係を考慮して視聴対象領域を拡大する手法 P は有効であるといえる．また，赤羽と豊田の結果を比較すると，提案手法 P による再現率の改善量には差があることがわかる．そこで，提案手法 P における，視聴対象領域の拡大量とカットの抽出精度について考察を行う．

正解許容しきい値 30 frame において，各正解視点切替フレームに最も近い視点切替フレームを求め，そのフレームで視点切替えを行うとして，推薦視点系列を生成した．そして，全 185 本分の編集視点系列を正解とし，この編集視点系列と推薦視点系列との一致率をフレーム単位で求めた．ただし，正解許容しきい値内に視点切替フレームが存在しない場合は，その正解視点切替フレームの前後に位置する部分ショットの全てのフレームを推薦失敗とした．

図 3.14 に，視聴対象領域の拡大量（横軸）と視聴対象フレーム数の増加量（縦軸右端，赤三角），推薦視点系列と編集視点系列の一致率（縦軸左端，黒丸）を示す．視聴対象領域の拡大量は，カメラの光軸方向にカメラ位置を移動させた際の，光軸と垂直方向へ視聴対象領域が拡大した量を表す．図中の点線は，各点群に対する回帰直線である．

図 3.14 より，赤羽と豊田の両方において，視聴対象領域の拡大に伴い，視聴対象フレーム数 (赤三角) が線形に増加したことを確認できる．また，フレーム一致率 (黒丸) は，豊田ではほとんど変化しなかったのに対し，赤羽では，0 m∼4 m で急激に増加し，4 m∼10 m で緩やかに増加し，10 m 以降でほとんど変化しないという変則的な結果になった．これらより，視聴対象領域の拡大量に関係して，提案手法 P が有効に作用する場合とそうでない場合が存在し，フレーム一致率の改善には限界があるとわかる．

初めに提案手法 P の効果の有無については，赤羽と豊田でカメラ配置が異なる点に起因

すると考えられる (図 3.1，図 3.2)．固定カメラでの広視域角多視点映像の撮影では，カメラの画角が固定であるため，被写体を近くから撮影すると，撮影可能領域が狭くなってしまう．これに伴い，一般的に切替え元の視点と切替え先の視点の視聴対象領域の重複は小さくなる．すると視聴対象領域が重複しない部分が増え，両カメラの映像に被写体が映らないという状況が発生する．視聴者はこのようなときにボールとカメラ間の関係を考慮して視点切替えを行っており，赤羽においては，被写体を近くから撮影しているため，提案手法が効果的に作用したと考えられる．一方，豊田は遠方からの撮影であり，もとより撮影可能領域の重複が十分にあるため，提案手法の効果が弱かったと考えられる．次に，提案手法 P の限界については，視聴対象領域を拡大していくと，切替え元の視点と切替え先の視点の視聴対象領域が重複するようになるため，効力を示さなくなったと考えられる．

以上より，提案手法 P は，赤羽のようにカメラ間の視聴対象領域の重なりが小さい場合に，効果的な方法であるといえる．また，カメラの視聴対象領域の重なり量は，カメラの設置状況（位置，角度，台数）やカメラの性能（画角，ズーム）によって変化するものであり，任意の組に対し最適な拡大量が存在すると考えらえるが，これは今後の課題とする．

(ii) 特徴量 (I · D · A) の考察

図 3.10 より，B-I-E（黒）と B-D-E（灰），B-A-E（黄），B-DA-E（桃）を比較すると，画像特徴量 I が最も高い再現率を示し，距離と角度を併用する DA，距離のみ利用する D，角度のみ利用する A の順となった．一方，図 3.11 と図 3.12 を見ると，B-A-E（黄）のカット数とリンク数が最も少なく，B-D-E（灰），B-DA-E（桃），B-I-E（黒）の順となった．以上より，再現率とカット数及びリンク数はトレードオフの関係にあることが分かる．また，B-DA-E（桃）は B-D-E（灰）や B-A-E（黄）よりも高い再現率を示しており，距離 D と角度 A は異なる特徴を持っており，これらを併用することは効果的であるといえる．

(iii) 視点切替フレームの検出手法 (S · E · R) の考察

図 3.10 より，P-DA-R（緑）が P-DA-E（青）よりも高い再現率を示し，B-DA-R（水）が B-DA-E（桃）よりも高い再現率を示していることを確認できる．さらに，B-DA-R（水）と B-S（細赤）を比較すると，極値検出の時間幅 $T_e = 30\ frame$ では，B-S（細赤）の方が高い再現率を示し，極値検出の時間幅 $T_e = 90\ frame$ では，B-DA-R（水）の方が高い再現率を示している．これは，ベースラインの B-S（水）が映像を等間隔に分割する手法であり，$T_e = 30\ frame$ という細かい幅で分割する方が $T_e = 90\ frame$ という粗い幅で分割するよりも，正解視点切替フレームの近傍に視点切替フレームを抽出する可能性が高くなるためであると考えられる．また，図 3.12 を見ると，極値検出の時間幅 $T_e = 30\ frame$ では，B-DA-R（水）の方が B-S（細赤）よりもリンク数が少ない．これは，B-S（細赤）が，映像分割に用いた全ての視点切替フレームをリンクで結ぶのに対し，提案手法である B-DA-R（水）は，切替え元の視点と切替え先の視点の両方を考慮してリ

ンクを結ぶため，リンク数を抑えられるからである．以上より，提案手法であるルールベースの視点切替フレームの検出手法 R は，リンク数を抑えながら高い再現率を示しており，有効な手法であるといえる．

(iv) 視点切替フレームの統合手法 (C) の考察

図 3.10 より，P-DA-R-C（太赤）は P-DA-R（緑）よりも再現率が低いが，図 3.11 と図 3.12 より，カット数とリンク数を大幅に削減できていることを確認できる．特に，リンク数に着目すると，P-DA-R-C（太赤）の方が P-DA-E（青）よりも低い値を示している．極値検出手法 E は，切替え元の視点の情報のみを考慮して視点切替フレームを抽出する．これに対し，提案手法 R は，切替え元の視点と切替え先の視点の両視点の情報を考慮しており，考慮対象が倍になったことで，抽出する視点切替フレームの数が増える．しかし，視点切替フレームの統合手法 C を用いることで，時間的に近接して抽出された視点切替フレームをまとめることができ，無駄なリンクを削減することができる．以上より，視点切替候補フレームの統合手法 C は，効果的に機能しているといえる．

3.3.2　部分評価実験

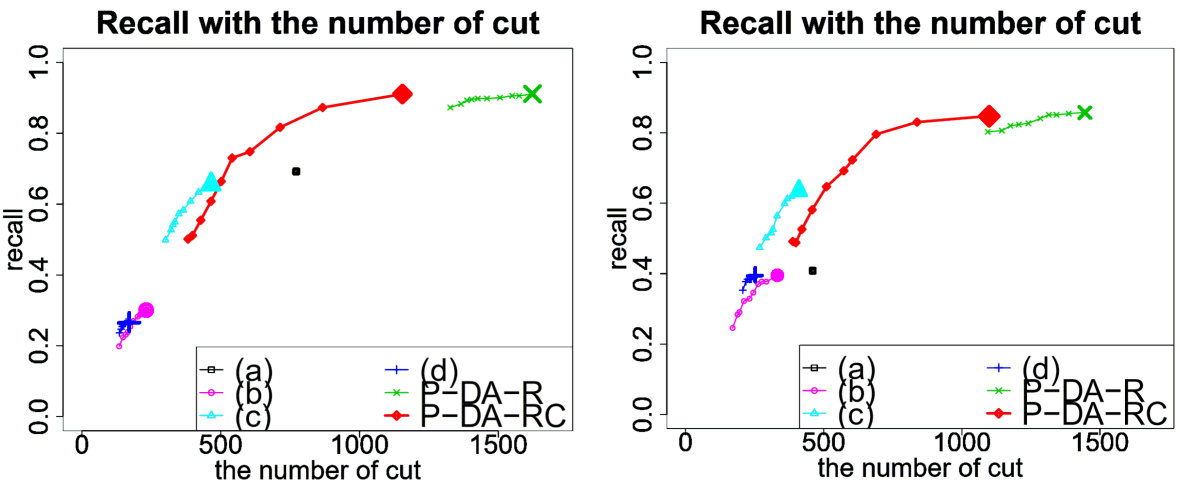
第 3.3.2 項では，提案した 4 つのカット抽出条件 (第 3.2.4.2(a)∼(d)) で，カットを適切に抽出できることを確認すると共に，これらを全て組合せた提案手法が各条件単体よりも優れていることを示す．

3.3.2.1　部分評価の実験方法

第 3.2.2 項に示す多視点映像の編集結果を正解とし，第 3.2.4.2(a)∼(d) に示す 4 つのカット抽出条件を用いて抽出した視点切替フレームとの一致率を評価した．

評価方法は，第 3.3.1.2 と同様の手法を用いた．正解データには，2 つの多視点映像データセット（赤羽と豊田）から抽出した 27 シーンを編集した 185 本の編集結果における視点切替フレームを用いた．一致率の評価には，再現率 (= 検出成功視点切替フレーム数/全正解視点切替フレーム数) と，抽出したカット数を用いた．これは，カット数と再現率の間には，カット数の増加に伴い再現率が上昇するという直接的な関係があるのに対し，カット数の増加に伴い増加する可能性のあるリンク数と再現率の間には直接的な関係がないからである． 検出成功視点切替フレームとは，各条件で抽出した視点切替フレームのうち，正解視点切替フレームとの時刻誤差が，30 frame(1 秒) 以内となるフレームの事である．比較対象には，視聴対象領域を拡大した状態における条件 (a)∼(d) と全条件の組合せである P-DA-R，さらに視点切替フレームの統合処理も行う P-DA-R-C とした．

実験設定は，第 3.3.1.3 と同様の設定を用いた．評価用パラメータとして，極値検出の時間幅 T_e と視点切替候補フレームの統合時間幅 T_i は同じ値とし，10∼100 frame の間を



(a) Akabane　(b) Toyota

図 3.15　4 つのカット抽出条件におけるカット数と再現率の変化

10 frame 刻みで変化させた．

3.3.2.2　部分評価の結果

4 つのカット抽出条件に対し，評価用パラメータである極値検出の時間幅 T_e と視点切替候補フレームの統合時間幅 T_i を変化させた際の再現率の変化を図 3.15 に示す．図 3.15 は，横軸に抽出したカットの数を，縦軸に再現率を示す．$T_e = T_i = 10\ frame$ のときの結果を大きな記号で描画し，それ以降の結果を線でつないでいる．

3.3.2.3　部分評価の考察

図 3.15 より，カットの抽出条件 (a)∼(d) で，カットの抽出が可能であることを確認できた．また，全ての条件を組合せた P-DA-R 及びこれに視点切替フレームの統合処理を加えた P-DA-R-C は，4 つの抽出条件よりも高い再現率を示した．したがって，4 つの抽出条件は，異なる特徴を表現したものであり，それぞれ単体で用いるよりも，複合的に用いる方が効果的であると言える．

次に，抽出条件 (a)∼(d) について詳細に分析する．4 つの条件は，多視点映像撮影時のカメラ配置と被写体位置との関係を表しており，特にカメラ位置とボール位置の相対関係を考慮している．各条件の例を図 3.16 に示す．図 3.16 は，視点 c_1 から視点 c_2 への視点切替えを表している．上段に俯瞰フィールドとカメラとボールを描画する．下段に視点 c_1 の画像と視点 c_2 の画像のイメージを描画する．各視点を同じ色で表現し，ボールは白のボール位置から緑のボール位置へ移動するものとする．

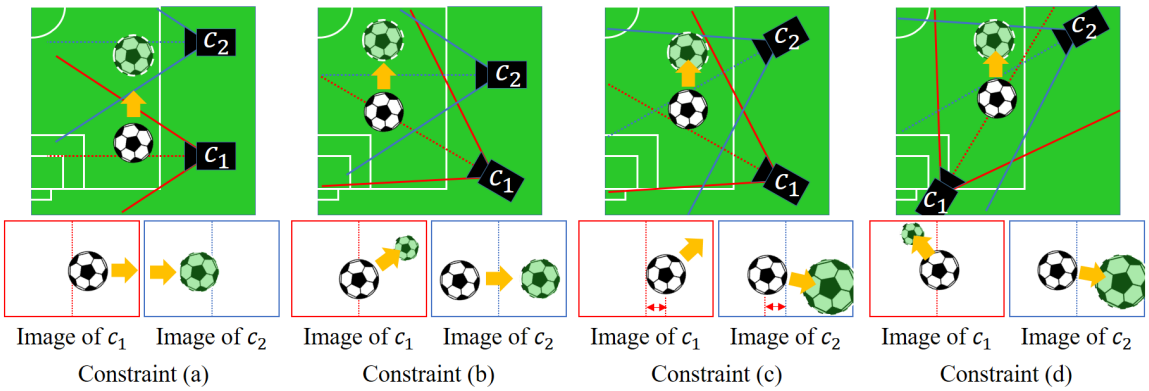


図 3.16　4 つのカット抽出条件の例

(i) カット抽出条件 (a) の考察

カット抽出条件 (a) は，ボールのフレームイン・フレームアウトを考慮した条件である．図 3.16(a) に，条件 (a) の視点切替えの例を示す．図は，ボールが画像に映る視点 c_1 から，ボールのフレームアウトを起点に，新たにボールが画像に映るようになった視点 c_2 に切り替える様子を表している．条件 (a) は，図 3.15 より，評価パラメータ ($T_e = T_i$) に関係なく，再現率とカット数において，赤羽では共に高い値を示し，豊田では共に低い値を示した．これは，条件 (a) が，カメラの視野の広さに影響を受ける条件であるためである．例えば，赤羽は近景からの撮影であり，1 台のカメラで撮影できる領域が限定的である．一方，豊田は遠景からの撮影であり，1 台のカメラでフィールド全体を撮影できる．そのため，赤羽ではボールの移動に対して，フレームインとフレームアウトが頻繁に発生し，カット数が上昇したと考えられる．また，被験者もフレームインとフレームアウトを中心に映像を編集する傾向が高くなるため，再現率が高くなったと考えられる．

(ii) カット抽出条件 (b) の考察

カット抽出条件 (b) は，切替え元の視点あるいは切替え先の視点において，ボールが画像中央にあることを考慮した条件である．図 3.16(b) に，条件 (b) の視点切替えの例を示す．図は，ボールが画像中央に映る視点 c_1 から，ボールが画像右端に移動するのを起点に，ボールが画像左端から画像右端に移動する視点 c_2 に切り替える様子を表している．条件 (b) は，図 3.15 より，赤羽と豊田の両者で，再現率とカット数の変化が小さい分布となった．これは，条件 (b) が，映像コンテンツに依存する条件であり，他の条件が画像全体で発生し得る可能性があるのに対し，画像中央にボールが存在するという条件は発生頻度が低いからである．そのため，全

ての視点切替フレームにおける本条件に合致するカット数が少なく，再現率が低くなったと考えられる．

(iii) カット抽出条件 (c) の考察

カット抽出条件 (c) は，切替え元の視点と切替え先の視点において，ボールと光軸間の絶対値角度が同程度であることを考慮した条件である．図 3.16(c) に，条件 (c) の視点切替えの例を示す．図は，視点 c_1 における画像中心からボールまでの距離と，視点 c_2 における画像中心からボールまでの距離が同程度であるときに，視点 c_1 から視点 c_2 に視点を切替える様子を表している．条件 (c) は，図 3.15 より，赤羽で 0.499 以上，豊田で 0.474 以上と共に高い再現率を示した．これは，条件 (c) が隣接する視点間での視点切替えにおいて発生しやすい条件だからである．特にサッカーにおいては，サイドラインをドリブルで駆け上がる，パスで前線にボールを送るなどのシーンが多く，条件 (c) に合致しやすい．また，隣接視点間での視点切替えは，視点切替え前後の視点映像において，映像内容として重複する部分が多く，視聴者にとって視認性が高いため好まれやすいという点も理由として考えられる．

(iv) カット抽出条件 (d) の考察

カット抽出条件 (d) は，切替え元の視点と切替え先の視点において，ボールとカメラ間の距離が同程度であることを考慮した条件である．図 3.16(d) に，条件 (d) の視点切替えの例を示す．図は，視点 c_1 と視点 c_2 がフィールドを挟んで反対の位置にあり，視点 c_1 におけるボールとカメラ間の距離と，視点 c_2 におけるボールとカメラ間の距離が同程度であるときに，視点 c_1 から視点 c_2 に視点を切替える様子を表している．条件 (d) は，図 3.15 より，条件 (b) と同様に赤羽と豊田の両者で，再現率及びカット数が共に低い分布を示した．これは，2 視点からボールまでの距離が同程度になることが少なかったことが原因であると考えられる．

3.3.3 信頼性評価実験

第 3.3.3 項では，カットの抽出条件を作成した編集映像データセット以外の編集映像データセットに対し，カットの抽出精度を評価することで，提案手法の信頼性評価を行った．

3.3.3.1 信頼性評価の実験方法

第 3.2.2 項で述べた多視点映像の編集において，豊田の多視点映像を編集した 10 人の被験者に，第 3.2.2 項と同様の方法で赤羽の多視点映像を編集してもらった．カメラ台数は，

20 視点中の 14 台 (図 3.1 の 1∼10 と 17∼20) とした．被験者の負担を考慮して，フィールド上の各所にボールが存在するように赤羽の多視点映像から 11 シーンを選択し，合計 10,111 frame(約 337 秒) の時区間を編集対象とした．これにより，全 110 本 (=11 シーン×10 人) の編集映像を収集した．編集されたカット数は合計 420 個であり，選択ショットの数は合計 530 個，全選択ショットのフレーム数の平均が 191.6 frame(約 6.4 秒)，標準偏差 141.4 frame であった．評価方法は，第 3.3.1.2 と同様の手法として，再現率 (= 検出成功視点切替フレーム数/全正解視点切替フレーム数) と抽出したカット数で評価した．これは，カット数と再現率の間には，カット数の増加に伴い再現率が上昇するという直接的な関係があるのに対し，カット数の増加に伴い増加する可能性のあるリンク数と再現率の間には直接的な関係がないからである． 420 個の視点切替フレーム (カット) を正解とし，抽出した視点切替フレームのうち正解フレームに最も近いフレームを求め，正解フレームとのフレーム時刻誤差が 30 frame(1 秒) 以内の場合を検出成功とした．比較対象は第 3.3.1.1 と同じとし，実験設定は第 3.3.1.3 と同じとした．このとき，評価用パラメータとして，極値検出の時間幅 T_e と視点切替候補フレームの統合時間幅 T_i は同じ値とし，10 から 100 frame の間を 10 frame 刻みで変化させた．

3.3.3.2　信頼性評価の結果

提案手法の信頼性評価として，横軸に抽出したカットの数を縦軸に再現率をとるグラフを図 3.17 に示す．$T_e = T_i = 10$ のときの結果を太線の記号で描画し，それ以降の結果を線でつなぎ描画している．

3.3.3.3　信頼性評価の考察

図 3.17 より，提案手法 P-DA-R 及び $T_e = T_i = 30$ 以下の P-DA-R-C が他の手法よりも高い再現率を示した．また，視点切替フレームの統合処理を行った P-DA-R-C の方が，P-DA-R よりも抽出したカット数が少ないことが確認できた．

以上より，提案手法 P-DA-R-C は，被験者やカメラ台数の異なる他の編集映像データセットに対しても，カットの抽出精度において高い再現率を示しており，有効な手法であるといえる．また，110 本の編集映像に対する全選択ショットのフレーム数の平均は 191.6 frame(約 6.4 秒) であり，極値検出の時間幅 T_e と視点切替候補フレームの統合時間幅 T_i が 30 frame(1 秒) 以下において，提案手法が最も良い精度を示している．したがって，1 秒よりも細かい間隔でカットの抽出を行う際に，提案手法が最もよい結果を示すといえる．

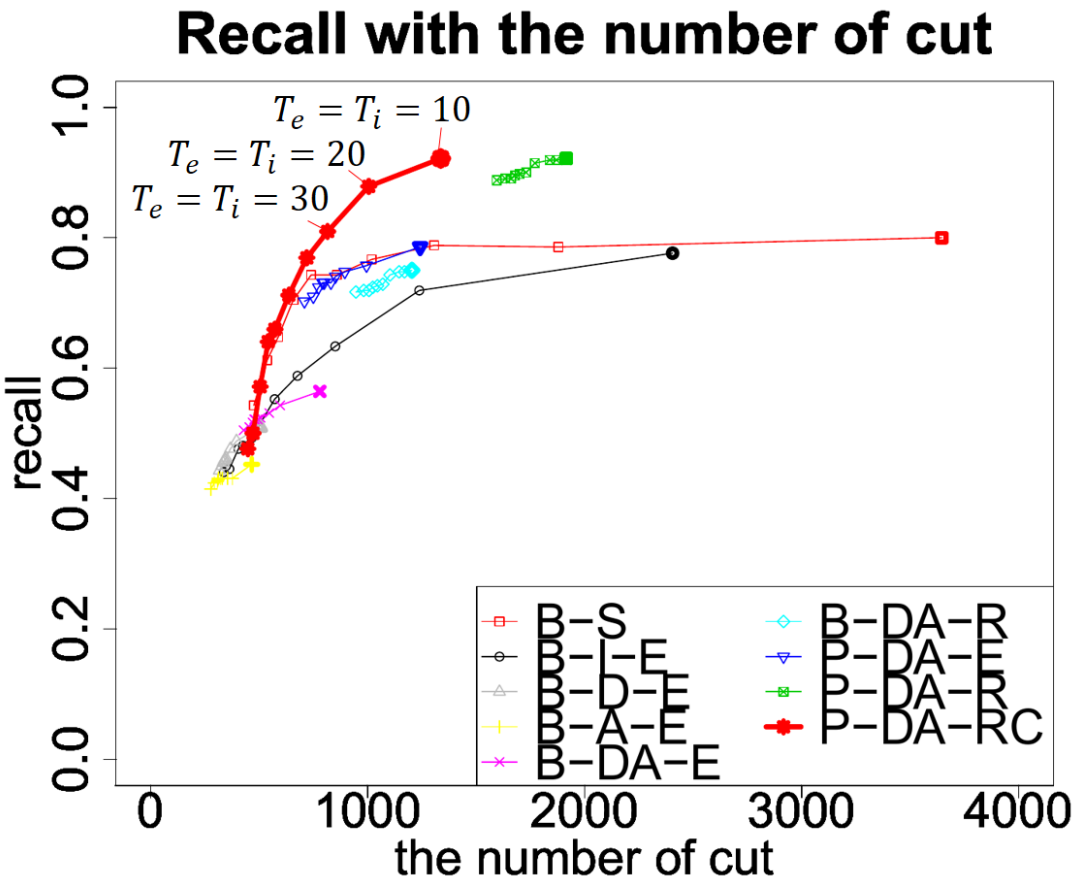


図 3.17　新たな編集映像データセットに対するカット抽出精度の再現率とカット数の変化

3.4　第 3 章のまとめ

第 3 章では，サッカーの試合を固定カメラで撮影した広視域角多視点映像の分割方法を提案した．具体的には，以下のようである．

- **一般的な視聴者が視点切替えを行うフレームで映像を分割できること．**

  一般的な視聴者の視聴傾向を調査するため，複数の一般的な視聴者の多視点映像編集結果を分析した．その結果，一般的な視聴者はボールに着目してサッカーの試合を視聴することが分かった．

  次に，同編集結果に対し，俯瞰フィールド座標系におけるボールとカメラとの幾何関係から視点切替えを行うフレームの決定条件を分析した．分析の結果，ボールがフレームインする直前やフレームアウトした直後のフレームも視聴対象としており，視聴者は，画像からだけでは判断できない情報も考慮しながら視点切替えを行っていることを確認した．

- **全ての視聴者の視点切替えに対応できるように不足なく映像を分割すること．**

  上述した俯瞰フィールド上におけるボールとカメラとの幾何関係に基づき，「視聴対象領域の拡大を利用した視聴対象フレームの抽出方法」及び「ボールとカメラ間の関係を利用した視点切替フレームの抽出方法」を提案した．

  サッカーの試合を撮影した実映像を用いて，カット抽出における再現率とカット数，リンク数について評価した．その結果，提案手法 (P-DA-R) は全比較手法に対し，最も高い再現率を示した．検出成功許容しきい値 30frame において，赤羽で 0.941，豊田で 0.924 であった．但し，本手法は全比較手法に対して，カット数が最多となり（赤羽：19,158，豊田：32,763），リンク数が二番目に多い数（赤羽：27,160，豊田：48,842）となった．したがって，本手法 (P-DA-R) は，視点切替えフレームを過剰検出していることが分かった．

- **実時間での映像生成を目指し有向グラフの規模を小さくするため，過剰に映像を分割しないこと．**

  上述した手法 (P-DA-R) に対し，視点切替え可能条件に基づき視点切替え候補フレームを統合し削減する手法 (P-DA-R-C) を提案した．

  同様にサッカーの試合を撮影した実映像を用いる評価実験を行ったところ，P-DA-R の手法よりも再現率が赤羽で 0.926，豊田で 0.893 に低下したものの，カット数とリンク数を大幅に削減することができた．したがって，本手法 (P-DA-R-C) が手法 (P-DA-R) と同程度の高い再現率を示し，視点切替え候補フレームの過剰抽出を抑制できることを確認した．

以上より，視点切替えフレームの抽出はトレードオフの関係にあるが，提案手法は無駄な映像分割を抑えつつ最適なタイミングで映像分割できることを確認した．これにより，多視点映像の視聴支援として，推薦視点系列を生成するために必要な有向グラフを作成することができるようになった．

第 4 章

粗密探索に基づくサッカー選手の特徴点追跡

4.1　はじめに

4.1.1　背景

我々の身の回りには，多くのカメラが存在しており，日々膨大な量の映像を撮影している．例えば，デジタルカメラやスマートフォンが普及したことで，個人が日常生活の出来事を容易に記録できるようになった．また，工場だけでなく商業施設や交通機関など様々な場所に監視カメラが設置されるようになり，我々の行動を記録し続けている．

撮影された映像には，多くの有益な情報が含まれている．例えば人物が被写体の場合を考えると，画像から抽出できる情報には，画像中での人物位置や大きさ，表情や服装などのテクスチャ，身体動作など様々なものが考えられる．さらに，時間変化を考慮すると，映像から人物の移動軌跡や行動などが情報として抽出できる．

本論文で対象とするスポーツ映像においても，多くの視聴者は選手に注目して映像を視聴する傾向にある．そのため，映像中の選手の位置や移動軌跡，選手の行動といった情報は，多視点映像の視聴支援を行う上でとても有益な情報となる．例えば，特定の選手が画像に映っているという情報が得られれば，その選手が映る視点映像だけをつなぎ合わせることで，その選手を追従するような視聴が可能となる．また，選手の行動を時系列的に解析し，選手の活動が顕著な部分ショットを抽出することができれば，そのショットを映像の見どころとして提示することも可能である．

本論文では，広視域角多視点映像の視聴支援として，映像に対する視聴者の嗜好に沿った視点系列を提示することで視点選択の負担を軽減し，視聴者が映像視聴に集中できるようにしたい．そこで，第 3 章の方法で分割した広視域角多視点映像に対し，分割した映像

をノード，分割したフレームをリンクとして重み付き有向グラフを作成する．そして，本グラフの最適経路探索問題として視聴者の嗜好との一致率が高い視点系列を算出する．この時，リンクのコストは，視点の切替えやすさを表し，視点切替え元の視点と視点切替え先の視点における両部分ショットの内容に基づき設定する．コストの大きさは，部分ショットの内容が視聴者の嗜好と一致しており，視点切替えによる視聴映像の変化が視聴者の負担にならない場合に，小さな値とする．

サッカー映像を視聴する一般的な視聴者は，第 3.2.3 項に示すようにボールに着目して視聴対象フレームを選択する．そして，サッカー選手に着目して同時刻における複数の部分ショットの中から視聴対象とする部分ショットを選択する．この時，視聴者は，画像上におけるサッカー選手の位置や大きさ，移動軌跡，身体動作，他選手との関係性，プレー内容など様々な情報を考慮して部分ショットを選択する．これらの情報のうち，画像上での選手の位置や大きさ，移動軌跡，他選手との関係性などの情報は，第 3.2.1 項で述べた俯瞰フィールド座標系における選手の位置情報を利用することで得ることができる．例えば，フィールド座標系の選手位置を画像に投影することで，画像上での選手位置が分かる．また，カメラ位置と選手位置との関係から，画像上での選手の大きさを算出することができる．さらに，画像上での選手位置を時系列に並べることで選手の移動軌跡を求めることができ，他選手の位置との距離を比較することで関係性が分かる．ボールと選手の関係についても，俯瞰フィールド座標系におけるボール位置と選手位置との距離を利用することでボール保持者を推定することができ，ボールの移動軌跡と移動速度からドリブルやパスなどのプレー内容を判別することができる．

しかし，選手の身体動作については，画像のテクスチャ情報を利用しなければ取得することができない．ここで言う選手の身体動作とは，キックやトラップ，フェイント，ヘディングのような選手の移動によらない動作の事である．本論文におけるリンクのコストの設定では，部分ショットにおけるサッカー選手の身体動作に対する時系列的な変化量を指標とする．この指標は，選手の身体動作が大きい場合に小さな値とし，身体動作が小さい場合に大きな値とする．つまり，大きな身体動作として選手がフェイントなどを行っている部分ショットへ視点を切替えやすくし，選手が単に移動しているような部分ショットへは視点を切替えにくくする．

映像から選手の身体動作の変化量を抽出するためには，連続する 2 フレーム間における選手領域のテクスチャ変化を数値化する必要がある．しかし，単純に選手領域の差分値を算出するだけでは，選手が動作したことしかわからない．つまり，選手が同じように手や足を動かしたのであれば，選手が走っているのか，歩いているのかを区別することができないということである．そこで，画像における選手の手や足などの局所領域から特徴点を検出し，連続する 2 フレーム間においてこの特徴点を追跡することで，特徴点の移動量を

選手の身体動作の変化量として利用する．特徴点とは，ある画素を中心とした画像の局所領域から得られる色や輝度勾配などの特徴量が，周辺画素に比べて特徴的な画素のことである．特徴点の追跡とは，連続するフレームに対し，特徴量が類似する特徴点同士を対応付けることである．

4.1.2 目的と研究課題

第 4 章では，リンクのコストの設定に必要なサッカー選手の身体動作の変化量を算出するために，スポーツ選手の特徴点を追跡したい．

本目的の達成には，スポーツ選手の緩急のある動作による特徴点の急激な速度変化や，スポーツ選手の 3 次元的な姿勢変化による特徴点の特徴量の変化に対応しながら特徴点を追跡する必要がある．

特徴点追跡に関する先行研究には，パターンベクトル間距離尺度や残差逐次検定（SSDA）法, 相互相関法などによる特徴点マッチング [40], Mean-Shift 探索 [42], Kalman–Filter[43, 44] や Particle–Filter[45] による特徴点の移動予測，Kanade-Lucas-Tomasi（KLT）法 [26] など様々な手法が存在する．

特徴点マッチングは，2 枚の画像から検出した特徴点同士を，特徴量の類似度に基づき対応付ける手法である．特徴点の移動量に関係なく特徴点を探索することができる．しかし，類似した特徴量を持つ特徴点が複数存在する場合や画像変化に伴い特徴量が急激に変化する場合には，探索精度が低下する．Mean-Shift 探索は，特徴ベクトル間距離に基づく極値探索手法であり，狭域に対し密な特徴点探索が可能である．しかし，特徴点が大きく移動すると，追跡対象の特徴点が探索領域外に移動してしまい，Mean-Shift 探索が局所解に陥り特徴点探索に失敗する．Kalman–Filter や Particle–Filter は，特徴点の移動を予測することで特徴点を追跡する手法である．予測モデルに合致した動きをする場合は良いが，特徴点が緩急ある動きをするなど予測に反した動きをした場合に追跡に失敗する．KLT 法は，微小時間において局所領域における各画素が同一の動きをすると仮定し，移動量を算出するために目的関数を最小化する手法である．高速に計算できるという利点を持つが，照明変化などにより局所領域における輝度値が変化すると特徴点を追跡できなくなる．

以上より，広視域角多視点映像におけるスポーツ選手の追跡は，先行研究で単純に解決することができず，自動視点推薦システムを実現する上で解決しなければならない特注の課題であると考え，次の研究課題を設定する．

- **連続する 2 フレーム間で特徴点の位置が大きく変化しても追跡できること．**
- **連続する 2 フレーム間で特徴点の特徴量が変化しても追跡できること．**

4.1.3　基本戦略

第 4.1.2 項に示す課題に対し，連続する 2 フレーム間において，特徴点を粗密に探索（Coarse-to-Fine Search）するという基本戦略を立てる．これは，$t-1$ フレームの特徴点と特徴量の類似度が最も高い t フレームの特徴点を探索するにあたり，広域を粗に探索することで特徴点の急激な速度変化に対処し，探索結果を初期値として狭域を密に探索することで特徴点の特徴量の変化に対処するものである．

具体的には，$t-2$ フレームの特徴点に対応する $t-1$ フレームの特徴点が探索できていると仮定して，$t-1$ フレームの特徴点に対応する t フレームの特徴点を探索することを考える．ここで，特徴点を，画像位置 $I(x,y)$ とこの位置を中心としスケール s によって決まる局所領域の特徴ベクトル V の組合せと定義する．

初めに，スポーツ選手の緩急ある動作による特徴点の急激な速度変化に対応するため，追跡対象である $t-1$ フレームの特徴点（以降，追跡点と呼ぶ）と t フレームの特徴点に対し，特徴点マッチングを行う．これにより，広域に対し粗な特徴点の探索を行うことができる．t フレームの特徴点は，次の 3 つからなる．1 つ目は，t フレームから検出した特徴点である（以降，検出点と呼ぶ）．2 つ目は，t フレームに対し，追跡点と同じ位置に設定した特徴点である（以降，投影点と呼ぶ）．3 つ目は，t フレームに対し，$t-2$ フレームと $t-1$ フレームの追跡点の位置から予測した位置に設定した特徴点である（以降，予測点と呼ぶ）．これらに対し，追跡点と予測点との距離に基づき，特徴点マッチングを行う探索範囲を決定し，この範囲内にある特徴点（検出点 + 投影点 + 予測点）とのマッチングを行う．マッチングのとれた特徴点を，対応点と呼ぶ．

次に，スポーツ選手の 3 次元的な姿勢変化による特徴点の特徴量変化に対応するため，対応点を初期値として，画像空間 I に対し Mean-Shift 探索を行う．これにより，狭域に対し密な特徴点の探索を行うことができる．

4.1.4　第 4 章の構成

本章の構成は次のようである．第 4.2 節で特徴点マッチングと Mean-Shift 探索を併用する特徴点追跡の手法について提案し，第 4.3 節でシミュレーション映像と実際のサッカー映像に対する特徴点追跡の評価実験について述べ，第 4.4 節でまとめる．

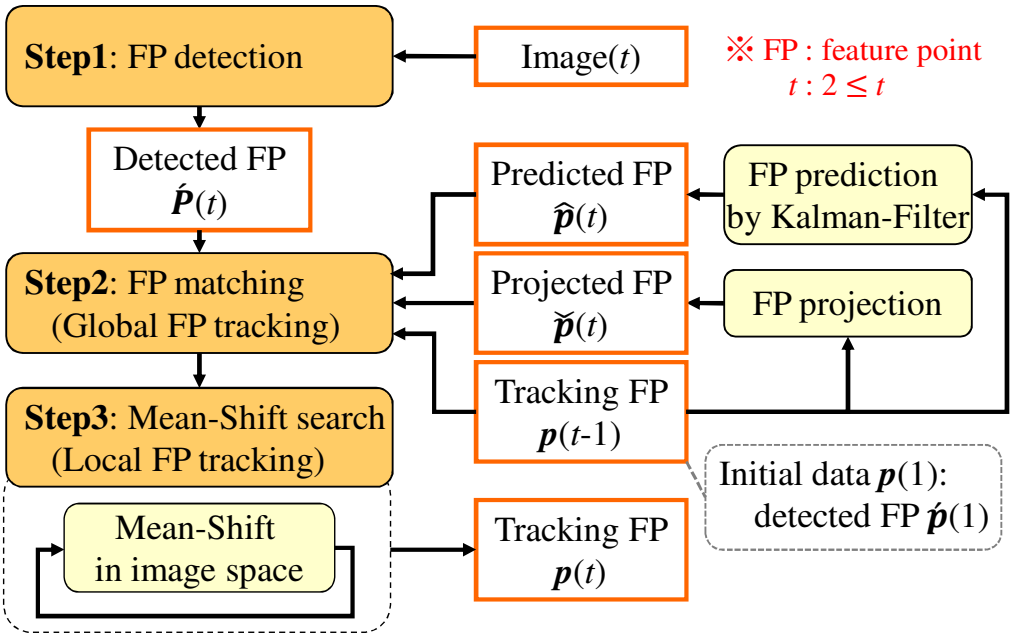


図 4.1　特徴点追跡処理の流れ

4.2　特徴点マッチングと Mean-Shift 探索による特徴点の粗密探索

第 4.2 節では，スポーツ選手の特徴点を追跡する方法について記す．提案手法の処理の流れを図 4.1 に示す． 図 4.1 は，t フレーム目の特徴点を追跡する処理の流れを表している．本手法は 3 つのステップからなる．第 4.2.1 項で「Step 1 : 特徴点検出」について，第 4.2.2 項で，「Step 2 : Kalman-Filter 予測を用いた特徴点マッチングに基づく粗な広域特徴点探索」について，第 4.2.3 項で，「Step 3 : Mean-Shift 探索に基づく密な狭域特徴点探索」についてそれぞれ記す．第 4.2.4 項では，提案手法の計算量について言及する．

手法の説明にあたり，t フレームから抽出した特徴点群を $\boldsymbol{P}(t)$，各特徴点を $\boldsymbol{p}_i(t)$ とする．

$$\boldsymbol{P}(t) = \{\boldsymbol{p}_1(t), \cdots, \boldsymbol{p}_N(t)\} \tag{4.1}$$

$$\boldsymbol{p}_i(t) = \{\mathbf{x}_i(t), s_i(t), \mathbf{v}_i(t)\} \qquad (i = 1, \cdots, N) \tag{4.2}$$

$$\mathbf{x}_i(t) = \{x_i(t), y_i(t)\} \in \mathbf{R}^2 \tag{4.3}$$

$$\begin{aligned}\mathbf{v}_i(t) &= \mathrm{SURF}(\mathbf{x}_i(t), s_i(t)) \\ &= \{v_{i,1}, \cdots, v_{i,V}\} \in \mathbf{R}^V \qquad (i = 1, \cdots, N)\end{aligned} \tag{4.4}$$

ここで，特徴点 $\boldsymbol{p}_i(t)$ は，画像空間における位置 $\mathbf{x}_i(t)$ とスケール空間におけるスケール

表 4.1　各特徴点の名称と表記方法

Point type	Notation	Num	Origin
Tracking	$\boldsymbol{p}(t)$	N	updated each time
Predicted	$\hat{\boldsymbol{p}}(t)$	N	each for a tracking point
Projected	$\check{\boldsymbol{p}}(t)$	N	each for a tracking point
Detected	$\acute{\boldsymbol{p}}(t)$	M	detected from input image

値 $s_i(t) \in \mathbf{R}$ という 2 つの空間変数を持ち，この空間変数により決まる局所領域から算出した特徴ベクトル $\mathbf{v}_i(t)$ を持つ．特徴ベクトル $\mathbf{v}_i(t)$ は，V 次元の SURF 特徴量を算出する関数 $\mathrm{SURF}(\mathbf{x}_i(t), s_i(t))$ より求める．

以下の説明では，$2 \leq t$ とし，$t-1$ フレームの特徴点追跡結果（以降，追跡点と呼ぶ）が得られているとする．また，初期値 $\boldsymbol{p}(1)$ は，Step1 にて初期フレームの画像から検出した特徴点 $\acute{\boldsymbol{p}}(1)$ である．説明で使用する特徴点（追跡点と予測点と投影点と検出点）を表 4.1 にまとめる．各特徴点の詳細は，後述する．表 4.1 の特徴点数 N と M は，入力画像に依存する値であり大小関係はない．

4.2.1　特徴点検出

Step1 では，SURF 検出器 [39] を用いて，入力画像から特徴点を検出する．この特徴点を検出点と呼ぶ．検出点とは，V 次元の特徴ベクトルを持ち，周囲の画素に比べて特徴量が特徴的な画素のことである．これを Step2 の特徴点マッチングで利用する．検出点群を $\acute{\boldsymbol{P}}(t)$，各検出点を $\acute{\boldsymbol{p}}_j(t)$，その点における特徴ベクトルを $\acute{\mathbf{v}}_j(t)$ と表記する．また，検出点の数を M 点とする．

4.2.2　Kalman-Filter 予測を用いた特徴点マッチングに基づく粗な広域特徴点探索

Step2 では，スポーツ選手の緩急ある動作に対する特徴点の急激な速度変化に対応するため，特徴点マッチングにより $t-1$ フレームの追跡点 $\boldsymbol{p}_i(t-1)$ と t フレームの特徴点を対応付ける．そして，この特徴点を Mean-Shift 探索の初期値とする．提案手法では，特徴点マッチングの計算量の増加と誤対応の発生を軽減するため，Kalman-Filter[43, 44] を用いて t フレームにおける追跡点の移動予測を行い，この予測位置と $t-1$ フレームの追跡点の位置を中心に探索範囲を設定することで，特徴点マッチングの候補となる検出点の絞込みを行う．

4.2.2.1 Kalman-Filter による追跡点の移動予測

スポーツ選手の加速による特徴点の速度変化に対応するため，Kalman-Filter を用いて，t フレームにおける追跡点の位置を予測する．

Kalman-Filter の式を，式 (4.5) に示す．

$$\hat{\mathbf{Q}}_i(t) = \mathbf{F}\mathbf{Q}_i(t-1) + \begin{pmatrix} \mathbf{0}_{4\times 1} \\ w_x \\ w_y \end{pmatrix} \tag{4.5}$$

ここで，追跡点は等加速度運動し，Kalman-Filter の状態モデルと観測モデルは線形であると仮定する．状態モデル $\mathbf{Q}_i$ は，追跡点の画像上における位置 $\mathbf{x}_i(t) = (x_i(t), y_i(t))$，速度 $(\dot{x}_i(t), \dot{y}_i(t))$ 及び加速度 $(\ddot{x}_i(t), \ddot{y}_i(t))$ を持つ．なお，w_x と w_y は，期待値が 0 で標準偏差がそれぞれ σ_x と σ_y のガウスノイズである．また，状態遷移行列 $\mathbf{F}$ を式 (4.6) と表記する．

$$\mathbf{F} = \begin{pmatrix} \mathbf{I}_{2\times 2} & \delta t\mathbf{I}_{2\times 2} & \frac{1}{2}\delta t^2\mathbf{I}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \delta t\mathbf{I}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} \end{pmatrix} \tag{4.6}$$

式 (4.6) の δt は単位フレームの時間長を表し，$\mathbf{I}_{2\times 2}$ は 2 行 2 列の単位行列を，$\mathbf{0}_{2\times 2}$ は 2 行 2 列の零行列を表す．

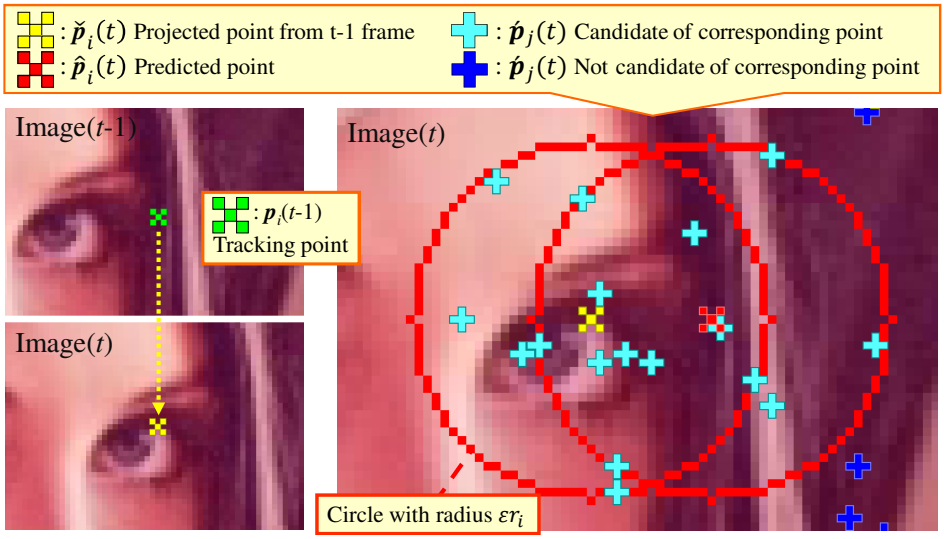
Kalman-Filter により予測した t フレームにおける追跡点の位置 $\hat{\mathbf{x}}_i(t)$ と，$t-1$ フレームの追跡点 $\boldsymbol{p}_i(t-1)$ のスケール値 $s_i(t-1)$ を持ち，これら空間変数に従い算出した特徴ベクトル $\hat{\mathbf{v}}_i(t)$ を持つ t フレームの特徴点を予測点 $\hat{\boldsymbol{p}}_i(t) = \{\hat{\mathbf{x}}_i(t), s_i(t-1), \hat{\mathbf{v}}_i(t)\}$ とする．また，予測点群を $\hat{\boldsymbol{P}}(t)$ と表記する．$t-1$ フレームの各追跡点に対し，t フレームの予測点は 1 つである．

4.2.2.2 $t-1$ フレームから t フレームへの追跡点の投影

スポーツ選手の減速による特徴点の速度変化に対応するため，$t-1$ フレームの追跡点の位置を，t フレームにおける追跡点の位置と予測する．

$t-1$ フレームにおける追跡点 $\boldsymbol{p}_i(t-1)$ の位置 $\mathbf{x}_i(t-1)$ と，スケール値 $s_i(t-1)$ を持ち，これら空間変数に従い算出した特徴ベクトル $\check{\mathbf{v}}_i(t)$ を持つ t フレームの特徴点を投影点 $\check{\boldsymbol{p}}_i(t) = \{\mathbf{x}_i(t-1), s_i(t-1), \check{\mathbf{v}}_i(t)\}$ とする．また，投影点群を $\check{\boldsymbol{P}}(t)$ と表記する．$t-1$ フレームの各追跡点に対し，t フレームの投影点は 1 つである．

図 4.2(a) に，投影点の説明を示す．上段に，時刻 $t-1$ フレーム中の追跡点 $\boldsymbol{p}_i(t-1)$（緑色の $\times$ 印）を示す．下段に，追跡点の位置を継承し，時刻 t フレームの画像に描画した投影点 $\check{\boldsymbol{p}}_i(t)$（黄色の $\times$ 印）を示す．



(a) Projected point.　　(b) Candidate points selection.

図 4.2　特徴点マッチング候補検出点の絞込み

4.2.2.3　特徴点マッチングの候補となる特徴点の絞込み

特徴点マッチングの計算量の増加と誤対応の発生を軽減するため，投影点と予測点に基づいて特徴点マッチングの候補となる特徴点の選定を行う．

画像上での特徴点間ユークリッド距離が，式 (4.7) を満たす特徴点を特徴点マッチングの候補とする．

$$\acute{r}_j \leq \varepsilon r_i \tag{4.7}$$

$$r_i = ||\check{\mathbf{x}}_i(t) - \hat{\mathbf{x}}_i(t)|| \tag{4.8}$$

$$\acute{r}_j = \min(||\acute{\mathbf{x}}_j(t) - \check{\mathbf{x}}_i(t)||, ||\acute{\mathbf{x}}_j(t) - \hat{\mathbf{x}}_i(t)||) \tag{4.9}$$

ここで，$\varepsilon \in \mathbf{R}$ は，候補を絞り込む範囲を決めるための係数である．距離 r_i は，t フレームにおける投影点の位置 $\check{\mathbf{x}}_i(t)$ と予測点の位置 $\hat{\mathbf{x}}_i(t)$ との距離である．距離 $\acute{r}_j$ は，t フレームにおける各検出点の位置 $\acute{\mathbf{x}}_j(t)$ と，投影点の位置 $\check{\mathbf{x}}_i(t)$ または予測点の位置 $\hat{\mathbf{x}}_i(t)$ との距離のうち小さい方の値である．

図 4.2(b) に，特徴点マッチングの候補となる特徴点を絞込む様子を示す．t フレームの投影点 $\check{\boldsymbol{p}}_i(t)$（黄色の × 印）と予測点 $\hat{\boldsymbol{p}}_i(t)$（赤色の × 印）を中心に半径 εr_i の赤色の円内に存在する特徴点が特徴点マッチングの候補である．このとき，式 (4.7) を満たす検出点（水色の十字）と投影点（黄色の × 印）及び予測点（赤色の × 印）が候補となる．以降の処理では，円内に存在する全ての特徴点と $t-1$ フレームの追跡点との特徴点マッチングを行う．

表 4.2　特徴点マッチングによる Mean-Shift 探索の初期値

Result of feature point matching	Detected point	Predicted point	Projected point
Location	$\acute{\mathbf{x}}_j(t)$	$\hat{\mathbf{x}}_i(t)$	$\check{\mathbf{x}}_i(t)$
Scale value	$s_i(t-1)$		
Feature vector	$\mathbf{v}_i(t-1)$		

4.2.2.4　特徴点マッチングと Mean-Shift 探索の初期値

特徴点マッチングにより，$t-1$ フレームの追跡点 $\boldsymbol{p}_i(t-1)$ と t フレームの特徴点の対応付けを行う．

特徴点マッチングは，式 (4.10) より求めた各特徴ベクトル間ユークリッド距離のうち最小の値を示す特徴点を対応点とする．

$$dist(\mathbf{v}_1, \mathbf{v}_2) = \sqrt{\sum_{k=1}^{V}(v_{1,k} - v_{2,k})^2} \tag{4.10}$$

次に，特徴点マッチングの結果に基づき Mean-Shift 探索の初期値を決定する．初期値は，対応点の位置 $\mathbf{x}$ と $t-1$ フレームの追跡点 $\boldsymbol{p}_i(t-1)$ のスケール値 $s_i(t-1)$，特徴ベクトル $\mathbf{v}_i(t-1)$ を持つ特徴点とする．表 4.2 に，特徴点マッチングの結果に応じた Mean-Shift 探索の初期値をまとめる．

図 4.3 に，$t-1$ フレームの追跡点が持つ特徴ベクトル $\mathbf{v}_i(t-1)$ と t フレームの各画素における特徴ベクトルとの類似度をグレースケールで示す．類似度の高い画素を白色で，低い画素を黒色で表現する．この例では，画像中央に位置する正解である特徴点 $\boldsymbol{p}_i(t)$（緑色の $\times$ 印）を Mean-Shift 探索で探索したいとする．しかし，従来手法のように，単に投影点 $\check{\boldsymbol{p}}_i(t)$（黄色の $\times$ 印）の位置から Mean-Shift 探索を開始すると，探索領域（桃色の正方形）内にある類似度の高い画素を探索してしまい局所解に陥ってしまう．これに対して，検出点 $\acute{\boldsymbol{p}}_j(t)$（水色の十字）や予測点 $\hat{\boldsymbol{p}}_i(t)$（赤色の $\times$ 印）から Mean-Shift 探索を開始することで，正しい特徴点を探索できるとわかる．

4.2.3　Mean-Shift 探索に基づく密な狭域特徴点探索

Step3 では，画像空間に対して Mean-Shift 探索を適用する．以下では，広視域角多視点映像において，スケール空間に対して Mean-Shift 探索を適用しない理由と，画像空間に対する Mean-Shift 探索 [24] の方法について述べる．

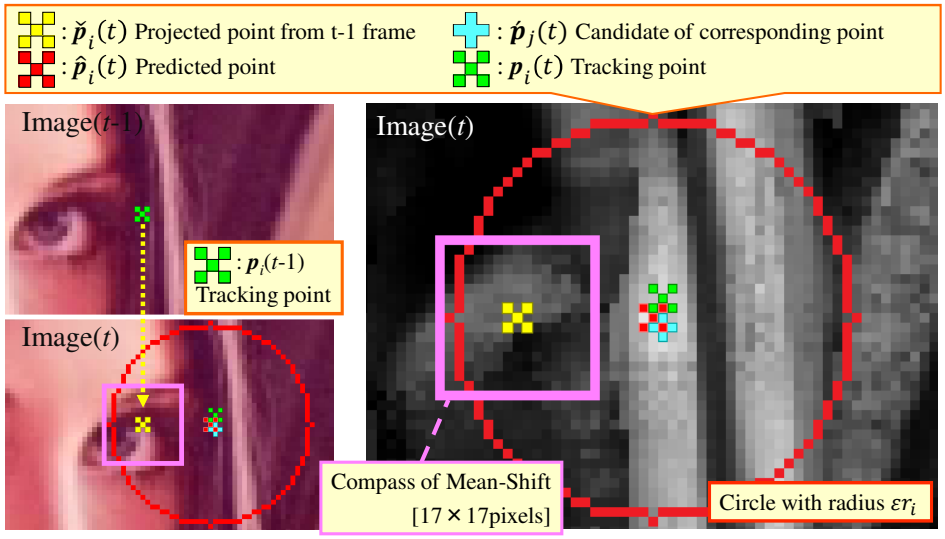


図 4.3　追跡点と各画素との特徴ベクトル間類似度分布

4.2.3.1　広視域角多視点映像におけるスケール空間の Mean-Shift 探索

一般的なスポーツ映像は，2 つに分類される．

1. 試合全体を把握しやすくするためフィールド全体を撮影した映像
2. 特定の選手を見やすくするためにスポッティングして撮影した映像

前者は，サッカーのような大きなフィールドで行うスポーツ映像に多く，撮影する映像に対し選手の大きさは小さくなる．後者では，被写体を同じ大きさで画面の特定位置に固定するように撮影する．これは，視聴者が被写体へ注意を向けやすいように，撮影者が故意に撮影するものである．したがって，一般的なスポーツ映像では，選手の大きさの変化が小さいと考えられる．

本論文では，サッカーの試合を固定カメラで撮影した広視域角多視点映像を対象とする．そのため前者のフィールド全体を撮影した映像に該当し，画像中での選手のスケール変化は小さいという仮定を置くことができる．したがって，スケール空間への Mean-Shift 探索は行わず，画像空間に対してのみ Mean-Shift 探索を適用するものとする．

4.2.3.2　画像空間の Mean-Shift 探索

Mean-Shift 探索では，各追跡点が持つ V 次元の特徴ベクトルとスケール値を参照データ $\tilde{\mathbf{v}}$，$\tilde{s}$ として利用する．参照データに基づき，画像空間における追跡点の変化量 $\Delta\mathbf{x} = (\Delta x, \Delta y)$ を求め，この値を用いて追跡点の位置 $\mathbf{x}$ を更新する．

初めに，追跡点の位置 $\tilde{\mathbf{x}}$ を中心とした周辺画素 $\mathbf{x}_l (l = 0, \cdots, N_{loc})$ の各特徴ベクトル

$\mathbf{v}_l$ を追跡点の参照用スケール $\tilde{s}$ に基づいて計算する．

次に，式 (4.10) を用いて，追跡点の参照用ベクトル $\tilde{\mathbf{v}}$ と求めた各特徴ベクトル $\mathbf{v}_l$ との特徴ベクトル間ユークリッド距離を式 (4.11) より求め，式 (4.12) の重みを算出する．

$$d(\mathbf{x}_l, \tilde{s}) = dist(\mathrm{SURF}(\mathbf{x}_l, \tilde{s}), \tilde{\mathbf{v}}) = dist(\mathbf{v}_l, \tilde{\mathbf{v}}) \tag{4.11}$$

$$w(\mathbf{x}, s) = \exp\left(-\frac{d(\mathbf{x}, s)^2}{2\sigma_d^2}\right) \tag{4.12}$$

ここで，σ_d は，特徴ベクトル間の類似度を重みに変換するための変数であり，$w(x, s)$ が 1 以下になるように特徴量の種類に応じて決定する値である．追跡点中心に向かって高い重みをかけて探索するために，求めた重み $w(\mathbf{x}_l, \tilde{s})$ と式 (4.13) のガウスカーネル関数を用いて，式 (4.14) の画像空間における追跡点の移動量を決定する．

$$K_{loc}(\mathbf{x}, \sigma_{loc}) = \exp\left(-\frac{x^2 + y^2}{2\sigma_{loc}^2}\right) \tag{4.13}$$

$$\Delta\mathbf{x} = \frac{\sum_{l=0}^{N_{loc}} K_{loc}(\mathbf{x}_l - \tilde{\mathbf{x}}, \sigma_{loc}) w(\mathbf{x}_l, \tilde{s})(\mathbf{x}_l - \tilde{\mathbf{x}})}{\sum_{l=0}^{N_{loc}} K_{loc}(\mathbf{x}_l - \tilde{\mathbf{x}}, \sigma_{loc}) w(\mathbf{x}_l, \tilde{s})} \tag{4.14}$$

ここで，σ_{loc} は，特徴点の移動距離を重みに変換するための変数である．K_{loc} が 1 以下になるように追跡対象の移動量に応じて決定する．

最後に，追跡点の位置を $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$ に応じて平行移動させる．上記の処理を，$||\Delta\mathbf{x}|| < \varepsilon_{loc}$ の収束条件を満たすまで反復する．

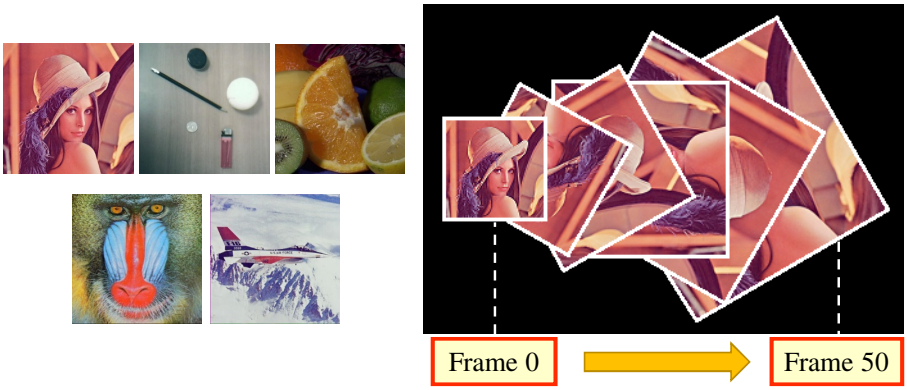
4.2.4 提案手法の計算量

提案手法は，特徴点マッチングと Mean-Shift 探索の組合せである．さらに，画像全体に特徴点マッチングを適用すると誤対応と計算量が増加するので，Kalman-Filter による特徴点の移動予測を利用して探索範囲を設定し，マッチング候補を限定している．

提案手法の定性的な計算量は，以下の 3 点となる．

1. Mean-Shift 探索のみ行う手法に比べ，特徴点マッチング分の計算量が増加する．
2. 画像全体に特徴点マッチングを行う手法に比べ，マッチング範囲を Kalman-Filter による予測位置を用いて限定しているため計算量が減少する．
3. 対応点から Mean-Shift 探索を開始するため，正解の画素に近い特徴点が対応点に選ばれると，Mean-Shift 探索の探索距離が短くなるため計算量は減少する．

しかし，これらは，特徴点マッチングの候補点の数，追跡対象の大きさ，Mean-Shift 探索の探索窓の大きさ，探索開始位置から終了位置までの距離など様々な要因が複雑に関係しており，映像や被写体によっても異なる．そのため，実際の計算量は条件よって大きく



(a) Five tracking target images.　(b) Moving example of G-v sequence.

図 4.4　追跡対象画像

変わると考えられる．また，本手法で特徴点の追跡精度を上げようとすると，計算量が増加するため，両者はトレードオフの関係にある．

4.3　特徴点追跡の評価実験

提案手法の有効性を確認するため，特徴点の追跡精度を評価した．はじめに第 4.3.1 項で，シミュレーション映像に対する特徴点の追跡精度について評価を行った．次に第 4.3.2 項で，実際のサッカーの試合を撮影した映像に対する特徴点の追跡精度について評価を行った．

4.3.1　シミュレーション映像を用いた特徴点の追跡実験

第 4.3.1 項では，提案手法が設計通りに機能することを確認するため，追跡対象である被写体の移動変化をシミュレートした画像シーケンスを作成し，特徴点の追跡精度を評価した．本実験は，都築ら [24] の評価手法を参考にして実施した．

4.3.1.1　画像合成による実験用画像シーケンスの作成

追跡する被写体として，追跡対象画像を用意した．追跡対象画像には，図 4.4(a) に示す 5 枚（320×320 画素）を使用した． この追跡対象画像を黒地の背景画像に重合せ，連続する 50 フレームの画像シーケンスを作成した．黒地の背景を利用したのは，背景の影響をなくし，特徴点追跡手法の効果を明確に評価するためである．

この追跡対象画像を移動変化させることで，特徴点の移動変化を表現した．本実験では，10 種類の移動変化を模擬し，各移動変化に対してフレーム間の変化量を 5 パターン

表 4.3 追跡対象画像の変化パターンと各フレーム間の変化量

Change pattern			Amount of change at each frame				
			i	ii	iii	iv	v
A	Parallel	[pixel]	3	6	9	12	15
B	Rotation	[°]	3	6	9	12	15
C	Scale-up	[pixel]	1.5	3	4.5	6	7.5
D	Parallel	[pixel]	3	6	9	12	15
	Rotation	[°]	3	6	9	12	15
E	Parallel	[pixel]	3	6	9	12	15
	Scale-up	[pixel]	1.5	3	4.5	6	7.5
F	Rotation	[°]	3	6	9	12	15
	Scale-up	[pixel]	3	6	9	12	15
G	Parallel	[pixel]	3	6	9	12	15
	Rotation	[°]	3	6	9	12	15
	Scale-up	[pixel]	1.5	3	4.5	6	7.5
H	Speed change of tracking target by sin function						
I	Perspective	[pixel]	3	6	9	12	15
J	Perspective	[pixel]	3	6	9	12	15
	Parallel	[pixel]	3	6	9	12	15

用意した．したがって，画像シーケンスは全部で 250 種類（移動変化 × フレーム間変化量 × 追跡対象画像 =10 種類 ×5 パターン ×5 枚）である．この詳細を表 4.3 に示す．例えば，移動変化 G の変化量パターン v は，連続フレーム間で追跡対象画像が 15 画素平行移動し，15°回転し， 7.5 画素拡大する画像を 50 枚つなげた画像シーケンスを表す．このシーケンス例を，図 4.4(b) に示す．これは Lena の画像が 0 フレーム目に左端の位置におり，フレームが経過するにつれて右端方向へ時計回りに回転拡大しながら平行移動していく様子を示している．

4.3.1.2 特徴点追跡精度の評価方法

追跡精度の評価には，式 (4.15) の特徴点の追跡成功率を用いた．

$$\text{追跡成功率} = \frac{\text{許容範囲以下の誤差で追跡できた特徴点の数}}{\text{特徴点の総数}} \tag{4.15}$$

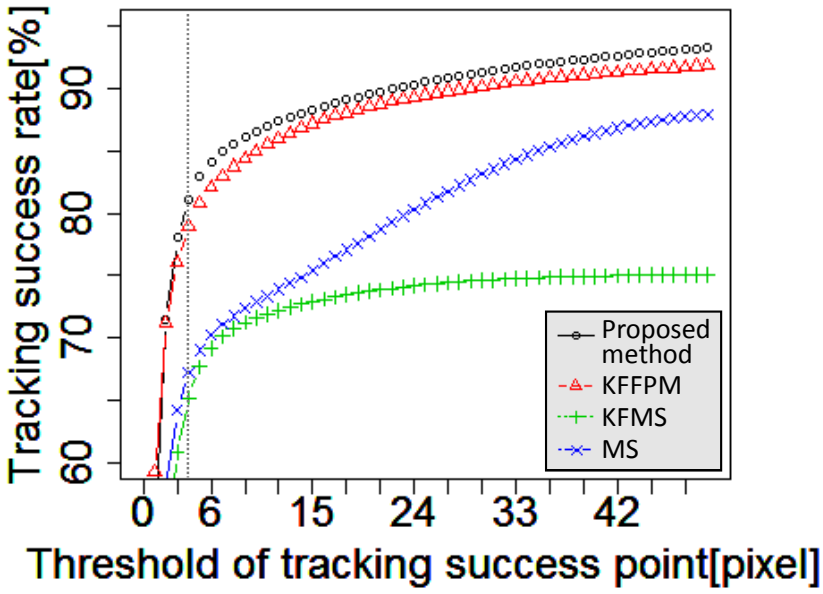


図 4.5　全画像シーケンスの各手法別特徴点追跡成功率

特徴点の正解位置は，追跡対象画像を $t-1$ フレームから t フレームへ変換する行列を用いて，$t-1$ フレームの特徴点を変換することで求めた．比較対象には，Kalman-Filter による予測を用いた特徴点マッチング（KFFPM）[40]，Kalman-Filter による予測を用いた Mean-Shift 探索（KFMS）[33]，Mean-Shift 探索（MS）[24] を採用した．これらに対し提案手法は，Kalman-Filter による予測を用いた特徴点マッチングと Mean-Shift 探索を組合せた手法である．また，提案手法及び比較対象において特徴点マッチングの ε を 1.2，Mean-Shift 探索の探索幅を 17×17 画素とした．使用する SURF 特徴量の次元数は 128 次元とした．

4.3.1.3　シミュレーション実験の結果と考察

図 4.5 に 250 種類の全画像シーケンスを追跡した結果を示す．これは，全連続フレーム間の追跡点（計 4,893,815 点）の追跡成功率である． このグラフは，横軸に追跡成功許容範囲を，縦軸に追跡成功率を表している．図 4.5 より，許容範囲を大きくすると追跡成功率が高くなり，追跡成功許容範囲 2 画素以上で提案手法が最も高い追跡成功率を示すことが分かる．

次に，追跡対象画像の移動変化パターン毎に，特徴点の正解移動量と追跡成功率について分析する．特徴点の正解移動量とは，$t-1$ フレームの特徴点位置と t フレームの正解特徴点位置との画像ユークリッド距離のことである．本分析では，提案手法と KFFPM との追跡成功率の差が最大となる許容範囲 4 画素に着目する．

図 4.6 に，追跡対象画像の移動変化パターン毎の特徴点の正解移動量と追跡成功率の関

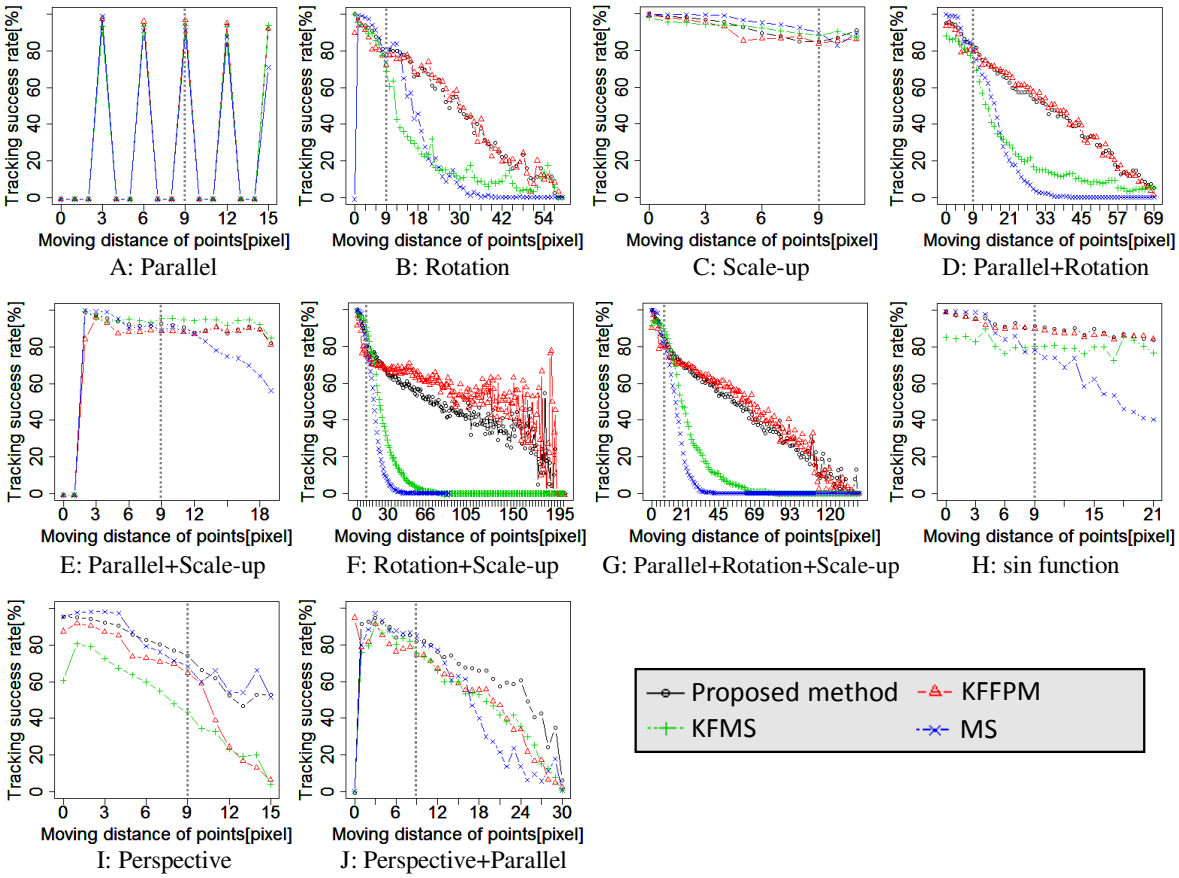


図 4.6　追跡成功許容範囲 4 画素における追跡対象画像の変化パターン毎の追跡成功率

係を示す．各グラフは，横軸に特徴点の正解移動量を，縦軸に追跡成功率を表す．追跡成功率は，正解移動量を四捨五入し，画素単位で分類してから計算した．

図 4.6 のパターン A のグラフは，平行移動の追跡結果である．追跡対象画像の正解移動量が 3 画素から 15 画素まで 3 画素ずつ変化する 5 パターンしかないため，それ以外の移動量の結果は 0 の値を示している．また，変化パターン毎に横軸のスケールが異なるのも同様の理由により，各画像シーケンスに依存するためである．さらに，先行研究より，Mean-Shift 探索は狭域に対し密な探索を行うため精度の高い追跡が行えると考えられる．そこで，探索幅の半分 9 画素までを特徴点の移動量が小さい狭域の場合，それ以上を特徴点の移動量が大きい広域の場合として結果を分析する．図 4.6 中の点線は，この境界である 9 画素を表す．

提案手法と KFFPM は，図 4.6 のパターン A∼H において，特徴点の移動量が大きな場合に特徴点を追跡できている．これは，Kalman-Filter による予測位置を用いて広域から範囲を限定し，特徴点マッチングを行ったためである．これらのパターンは追跡対象画像自体の変化がなく，抽出される特徴量も変化しないため特徴点マッチングが効果的に機能

したと考えられる．これに対し，MS は，特徴点が探索範囲外に移動したことにより，MS が局所解に陥り追跡に失敗した．また，KFMS は，追跡対象画像の回転変化に伴う特徴点の移動に対して，Kalman-Filter の予測が外れ，この値を MS の初期位置としたため追跡に失敗した．

さらに，各特徴点追跡手法の追跡成功率の代表値（最大値，最小値，平均）を追跡対象画像の変化パターン毎に表 4.4 と表 4.5 にまとめる．このとき，代表値に対し上位 2 位までを太字で印字し，更に最上位に*を記す．

パターン I とパターン J は，透視投影変換により追跡対象画像自体が変形する画像シーケンスである．そのため，画像から得られる SURF 特徴量も変化する．これらのパターンにおいて，表 4.5 の左列に示す特徴点の移動量が小さい場合では，提案手法が MS とともに高い追跡成功率を示している．これは，MS が狭域に対して密な特徴点の探索を行うことにより，追跡点と類似した特徴量を持つ画素を探索できたためであると考えられる．表 4.4 と表 4.5 より，同様の理由でほかのパターンにおいても特徴点の移動量が小さい場合に，提案手法と MS が高い追跡成功率を示していることを確認できる．また，追跡対象画像が透視投影変換により変形しながら平行移動するパターン J では，図 4.6 より，特徴点の移動量が 12 画素を過ぎたあたりから提案手法が最も高い追跡成功率を示している．これは，MS が特徴点の大きな移動に対して局所解に陥り探索に失敗したのに対し，提案手法は Kalman-Filter 予測を用いた特徴点マッチングにより MS の初期位置を変更したため追跡できたと考えられる．特徴量の変化に弱い KFFPM と比較しても高い追跡成功率を挙げていることが分かる．

以上より，提案手法は他の比較手法に比べ，特徴点の移動量が小さい場合も大きい場合も，また特徴量が変化する場合も特徴点を追跡することができるといえる．

4.3.2　サッカー映像を用いた特徴点の追跡実験

第 4.3.2 項では，提案手法がスポーツを撮影した実映像に対して機能することを確認するため，サッカー選手を撮影した映像に対し特徴点の追跡精度を評価した．

4.3.2.1　サッカー映像データセット

サッカーの試合を撮影した映像（1920×544, インタレース除去済み, 30fps）を用意した．本映像は，固定カメラで撮影した広視域角多視点映像のある視点における撮影映像である．遠方から広いフィールド上のサッカー選手を撮影しているため，画像上での選手のスケール変化は少ない．具体的には，白色のユニフォームを着たサッカー選手（以降，白色選手と表記）が映っている．白色選手は，画面中央に画面に対して背を向けた状態で映っており，この位置から右方向に走って移動する．その間，他選手とのオクルージョン

はなく，急加速や急停止などの速度変化とそれに伴う 3 次元的な姿勢変化を行いながら 2 回の切り返しを行う．

4.3.2.2 評価方法

白色選手の特徴点を追跡し，追跡精度を評価した．

はじめに，映像の各フレームから白色選手の全身を包含する矩形を手動で抽出した．この矩形を正解矩形と呼び，この矩形内に入っている特徴点を追跡成功とした．特徴点を追跡した全フレームに対する矩形の平均サイズは 62.9×67.1 画素 (最大サイズ 78×80 画素，最小サイズ 53×56 画素)，矩形重心の平均移動量は 2.7 画素 (最大移動量 8 画素，最小移動量 0 画素) であった．したがって，本サ像における白色選手の移動量は，第 4.3.1 項のシミュレーション実験で求めたパラメータで対応できる範囲であると言える．次に，初期フレームから検出した特徴点のうち正解矩形内に存在する 12 点を追跡する特徴点とした．

比較手法には，提案手法及び Kalman-Filter による予測を用いた特徴点マッチング（KFFPM）[40]，Kalman-Filter による予測を用いた Mean-Shift 探索 (KFMS)[33]，Mean-Shift 探索（MS）[24] を用いた．

追跡に必要なパラメータは，第 4.3.1.2 のシミュレーション実験で求めた値を用いた．提案手法及び比較対象における特徴点マッチングのパラメータを ε =1.2，Mean-Shift 探索の探索幅を 17 × 17 画素とした．

4.3.2.3 サッカー選手の追跡結果

白色選手の特徴点を追跡した様子を図 4.7 に示す．図 4.7 は，20 フレーム目から 140 フレーム目まで，20 フレーム毎にフレームを抽出したものである．各画像は，白色選手が移動した領域全体を包含するように切り出した画像 (312×141 画素) である．左から (a) 提案手法，(b)KFFPM，(c)KFMS，(d)MS による特徴点の追跡結果である．特徴点の位置を点で，過去 20 フレーム分の特徴点の軌跡を線で表す．

図 4.7 より，提案手法と KFMS，MS が白色選手を追跡できていることがわかる．Frame40 付近では横向きだった選手の体が左に 90 度回転し背中が映るようになる．この 3 次元的な姿勢の変化に対して，提案手法は全身の特徴点を追跡できているが，MS は臀部周辺に特徴点が集まってしまった．また，Frame140 付近では選手が急停止し切り返し動作を行っている．この速度が急に変化する動きに対して，提案手法は選手の身体を追跡しているが，KFMS は Kalman-Filter が予測に失敗したために Frame120 で選手の頭を追跡していた点が Frame140 では選手の腕を追跡している．

次に，白色選手の特徴点の追跡結果に対し，追跡に成功した特徴点数の変移を図 4.8 に示す．図 4.8 は，横軸にフレーム番号を，縦軸に追跡に成功した特徴点数を表す．

図 4.8 より，提案手法は，全フレームに渡り全ての特徴点を追跡することができた．こ

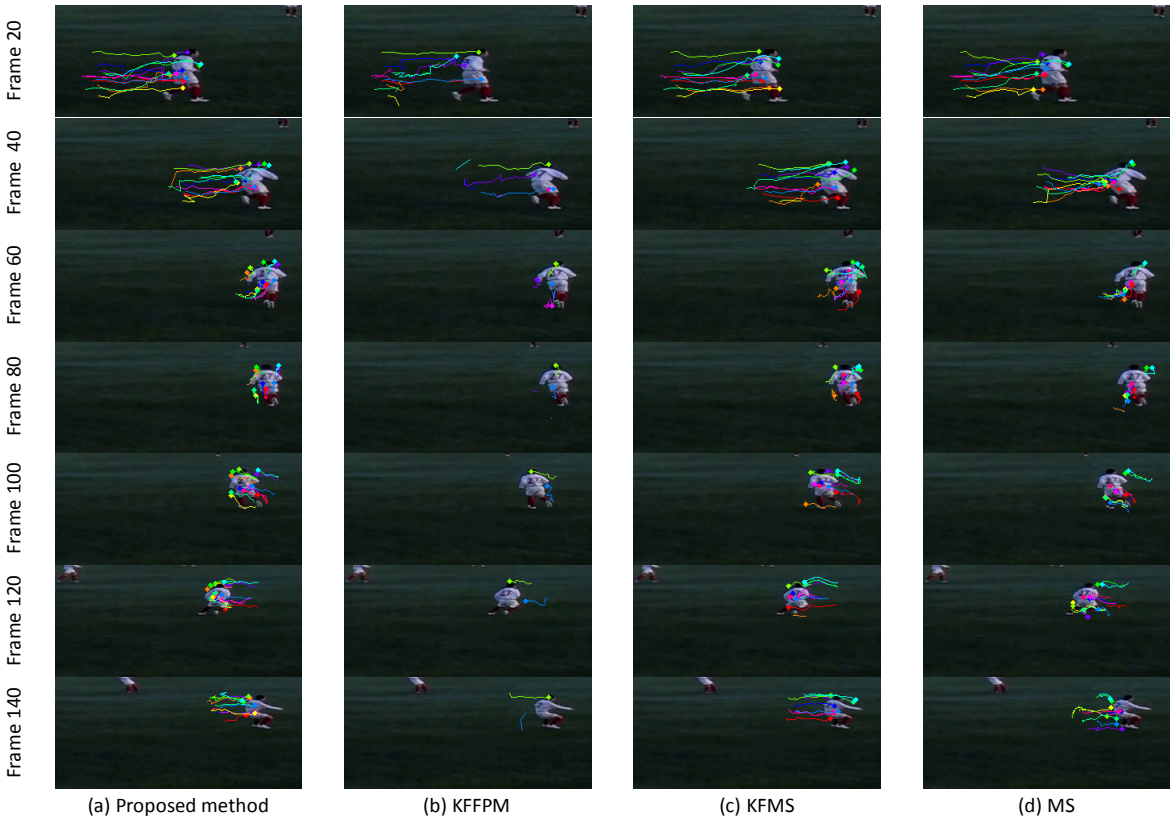


図 4.7　20 フレーム置きのサッカー映像の追跡結果

れに対し，KFMS は，24 フレーム目と 107 フレーム目で選手の足の追跡に失敗した．これは，Kalman-Filter が予測に失敗したためである．MS は，73 フレーム目で足の大きな動きに，121 フレームで手の大きな動きに Mean-Shift 探索が対応できず追跡に失敗した．KFFPM は，早い段階で特徴点の追跡に失敗した．これは，比較的特徴量の変化が少ないフレームであるにもかかわらず，Mean-Shift 探索を行わないために特徴量変化に対応できなかったためと考えられる．

提案手法と KFMS，MS の 3 手法を比較すると，最終的に追跡できた特徴点数の差は 2 つであった．これは，本映像において白色選手の動きは複雑であるが，サイズ変化は小さく背景の影響もないため，姿勢変化に伴う特徴量変化が比較的少なく，Mean-Shift 探索で対応できたためと考えられる．しかし，急激な速度変化や方向変化を行う腕や足といった部位の追跡に対して，KFMS や MS は対応できていない．

以上より，提案手法は，サッカー選手を撮影した実映像に対して，サッカー選手の急激な速度変化や姿勢変化に伴う特徴量の変化が存在する場合においても，特徴点を長期にわたり追跡することができるといえる．

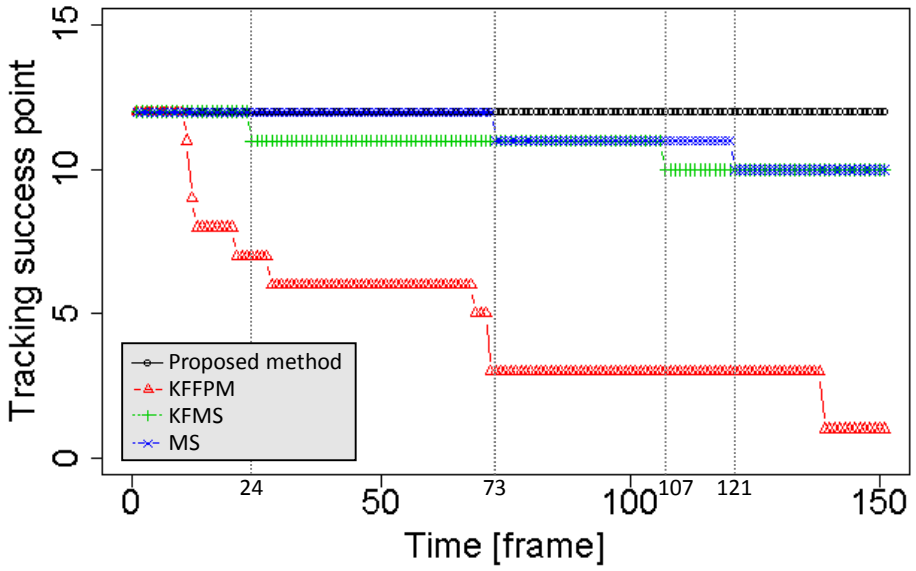


図 4.8　サッカー映像の追跡成功特徴点数の変移

4.4　第 4 章のまとめ

第 4 章では，サッカーの試合を固定カメラで撮影した広視域角多視点映像を対象に，映像から抽出されるサッカー選手の特徴点を，精度よく追跡する手法を提案した．スポーツ選手は，プレーにおいて，急激な速度変化や 3 次元的な姿勢変化を行う特徴があり，これにより特徴点の速度変化や特徴点の特徴量変化が発生する．この課題に対し，特徴点マッチングによる粗な広域特徴点探索と Mean-Shift 探索による密な狭域特徴点探索を組合せる手法を提案した．

提案手法の有用性を示すために，シミュレーション映像とサッカー映像に対する特徴点の追跡実験を行った．シミュレーション映像を用いた実験では，特徴点の追跡に成功したと判断する許容範囲を 2 画素以上にした場合において，提案手法が最も高い追跡成功率を示した．また，サッカー映像を用いた実験では，提案手法が長時間に渡り最も多くの特徴点を追跡することができた．

具体的には，次のようである．

- **連続する 2 フレーム間で特徴点の位置が大きく変化しても追跡できること．**

  連続する 2 フレーム間の特徴点に対して，任意の 2 点で特徴量の類似度を比較し，類似度の高い特徴点の組を対応付ける特徴点マッチングを用いることで，特徴点の大きな移動に対応できるようにした．これによりシミュレーション映像を用いた評価実験では，平行移動，回転移動，拡大縮小移動，移動速度変化およびこれら

を組合せた特徴点の移動に対して，特徴点が大きく移動（10 画素以上移動）する場合に比較手法（KFFPM，KFMS，MS）と同等以上の追跡精度を示すことを確認した．

- **連続する 2 フレーム間で特徴点の特徴量が変化しても追跡できること．**

  連続する 2 フレーム間に対して，前フレームの特徴点の類似度を基準に，後フレームから類似する特徴量を持つ画素を密に探索する Mean-Shift 探索を用いることで，特徴点の特徴量の変化に対応できるようにした．これによりシミュレーション映像を用いた評価実験では，透視投影変換による特徴点の特徴量変化に対して，比較手法である MS と同等以上の追跡精度を示すことを確認した．

さらに，シミュレーション映像を用いた評価実験において，特徴点の移動距離が大きくかつ特徴点の特徴量も変化する場合において，全比較手法（KFFPM, KFMS, MS）よりも高精度に特徴点を追跡できることを確認した．これに加えて，サッカー選手を撮影した実映像を用いた評価実験においても，提案手法が，1 人のサッカー選手から抽出した 15 個の全特徴点を 150 フレームに渡り追跡できることを確認した．これは，全比較手法（KFFPM, KFMS, MS）の中で最高の追跡精度であった．

以上より，スポーツ選手の特徴点の追跡において，提案手法が精度よく選手の特徴点を追跡できることを確認した．これにより，多視点映像の視聴支援として，推薦視点系列を生成するために必要な有向グラフのリンクの重みを，画像上での選手の動きに着目して設定することができるようになった．

表 4.4　追跡成功率の代表値

Move pattern	Method	∼9 pixel			10∼ pixel		
		Max	Min	Avg	Max	Min	Avg
A	Proposed	94.9	**92.3**	**93.2**	**92.3**	**90.5**	**91.4**
	KFFPM	**96.5**	***92.7**	***95.0**	***95.2**	***91.8**	***93.5**
	KFMS	89.6	88.8	89.2	89.3	88.0	88.6
	MS	***96.8**	88.9	91.6	85.1	68.7	76.9
B	Proposed	**96.3**	***73.3**	**85.4**	**77.3**	2.0	**40.0**
	KFFPM	95.1	69.1	80.1	76.1	***2.8**	***40.1**
	KFMS	92.9	63.4	82.1	58.9	**2.4**	15.1
	MS	***99.3**	**69.2**	***87.5**	***79.9**	0.1	20.7
C	Proposed	**98.0**	80.1	**88.5**	***88.9**	***82.8**	***85.8**
	KFFPM	97.2	80.0	86.6	83.7	5.0	57.1
	KFMS	91.6	***82.7**	87.9	**84.4**	2.8	57.0
	MS	***99.4**	**81.9**	***92.4**	82.2	**75.0**	**78.6**
D	Proposed	**95.2**	***78.7**	**86.0**	***78.1**	***4.2**	**38.3**
	KFFPM	94.3	74.7	81.3	**76.7**	**1.5**	***38.8**
	KFMS	83.3	69.2	78.3	64.8	0.5	15.2
	MS	***98.7**	**75.1**	***87.0**	69.9	0.0	17.6
E	Proposed	98.4	***88.3**	***91.4**	**89.4**	***74.1**	**85.1**
	KFFPM	92.9	82.1	86.4	88.8	**73.9**	85.0
	KFMS	**98.9**	**88.2**	**91.3**	***91.8**	73.5	***87.3**
	MS	***99.8**	85.5	91.2	88.4	47.6	70.6
F	Proposed	**97.1**	**79.1**	87.3	76.3	**4.0**	**43.0**
	KFFPM	96.1	67.9	79.1	**77.8**	***10.0**	***54.1**
	KFMS	95.9	***84.3**	***90.9**	***82.3**	0.0	19.2
	MS	***99.3**	75.9	**87.9**	70.3	0.1	15.7
G	Proposed	**96.7**	**80.3**	**87.3**	**79.1**	**1.2**	**37.7**
	KFFPM	96.2	73.7	81.4	77.6	***1.4**	***40.4**
	KFMS	90.5	***82.9**	***87.8**	***81.2**	0.1	18.2
	MS	***99.2**	76.6	87.1	72.7	0.0	16.0

表 4.5　追跡成功率の代表値（続）

Move pattern	Method	∼9 pixel			10∼ pixel		
		Max	Min	Avg	Max	Min	Avg
H	Proposed	96.0	**87.6**	***91.0**	***88.5**	**80.8**	**84.9**
	KFFPM	**96.6**	***88.6**	**90.8**	**87.7**	***83.1**	***85.3**
	KFMS	83.7	69.3	75.6	77.8	64.9	72.0
	MS	***99.1**	73.3	86.5	71.8	39.1	55.4
I	Proposed	**94.2**	***69.0**	***81.2**	***61.2**	**38.7**	**49.5**
	KFFPM	90.9	**62.1**	74.3	57.0	5.6	24.5
	KFMS	75.1	38.3	57.2	29.9	3.5	19.5
	MS	***98.1**	60.3	**78.9**	**59.8**	***46.4**	***51.7**
J	Proposed	**92.0**	**78.1**	**85.1**	***78.2**	***3.1**	***54.3**
	KFFPM	88.2	71.7	77.0	72.1	**2.2**	**40.9**
	KFMS	86.6	69.0	76.5	66.4	0.3	37.6
	MS	***96.0**	***79.9**	***85.2**	**75.8**	0.5	32.9

第 5 章

結論

第 5 章では，第 5.1 節で本論文のまとめを，第 5.2 節で今後の研究課題をそれぞれ述べる．

5.1　まとめ

本論文では，複数の視点を有する映像のうち，サッカーの試合を固定カメラで撮影した広視域角多視点映像を対象に，視聴者が不便なく自らの意思に基づき視点を切替えながら映像を視聴することができる人とシステムの協調視聴方式を想定して，その実現に向けた基礎研究を行っている．協調視聴方式は，視聴者の「手動視点選択機能」とシステムの「自動視点推薦機能」からなっており，システムが視聴者の視点切替えから映像に対する嗜好の変化を読み取り，それに応じた視点系列を推薦することを特徴とする．これにより視聴者は，常に自身の望む視点からの映像を視聴することが可能となる．

本論文では，協調視聴方式について第 1 章で述べ，この内，自動視点推薦システムの実現に向けた 2 つの研究を第 3 章と第 4 章で述べた．1 つ目 (第 3 章) は，被写体とカメラとの幾何関係に基づき多視点映像から一般的な視聴者の視点切替えを過不足なく抽出し，これに基づき映像を分割することを特徴とする映像分割方法である．2 つ目 (第 4 章) は，画像空間を粗密に探索することを特徴とするスポーツ選手の特徴点追跡方法である．本論文の貢献は，現在注目が集まっている複数の視点を有する映像の視聴方法について，実利用の観点から人とシステムとの協調視聴方式を想定し，その実現に向けた基礎研究を通して新たな知見を提供した点にある．以降では，各章のまとめを記す．

第 1 章では，序論として本論文の背景から問題，目的，アプローチ，課題について述べた．情報技術の進歩により，「高解像度映像の撮影」や「大容量データの高速な処理や通

信」が可能となったことで，複数の視点を有する映像のエンターテインメント利用に注目が集まっている．複数の視点を有する映像とは，被写体を様々な位置や角度から同時に撮影した映像群のことである．全視点の映像を視聴者に提供することで，視聴者が視点と視線を切替えながら映像を視聴できるようになる．これにより，視聴者は被写体を自由な位置と角度から観察することができるようになる．しかし，視点切替えは，7 次元もの変数（時間，視点，視線）を適宜選択し続ける負荷の高い操作であるため，視点切替え自体が視聴者の負担となってしまう．

先行研究では，視点切替えの「簡便化」と「自動化」という 2 つの方向性について議論している．しかし，実際に複数の視点を有する映像を視聴する状況を考えると，視点切替えの利点を活かしつつ欠点を改善するためには，これら 2 つの方向性を同時に満たす人とシステムの協調視聴方式が必要である．

協調視聴方式とは，複数の視点を有する映像の視聴を，視聴者と視点推薦システムが互いの行動を理解しつつ視聴すべき視点を選択する協調作業であるとみなしたものである．本方式は，手動視点選択部と自動視点推薦部にて決定された視点映像を映像表示部で表示する構成となっている．特に本論文では，自動視点推薦部の実現に着目しており，システムが視聴者の視点切替えから映像に対する嗜好の変化を読み取り，これに応じて新たな視点系列を推薦することができる仕組みを自動視点推薦部に導入した．これにより視聴者は，常に自身が望む視点からの映像を最小限の視点切替えで視聴可能となる． 具体的には，視点系列の生成を，多視点映像の視点切替えを表現した重み付き有向グラフの最適経路探索問題として解く．この時，視聴者の視点切替えに応じてグラフの重みを変更することで，生成する視点系列を変更する．

これに対し，自動視点推薦システムを実現するという基本戦略を示すと共に，自動視点推薦機能の実現に必要な「広視域角多視点映像の分割手法」と「サッカー選手の特徴点追跡手法」という 2 つの研究課題を示した．

第 2 章では，複数の視点を持つ映像の社会的な普及という観点から，「撮影」，「伝送」，「編集」，「提示」の 4 つの関係性について述べ，本論文で対象とする「編集」と「提示」について先行研究をまとめ，本論文の位置づけを示した． これら 4 要素は，明確に独立しておらず，相互に関係しあっている．

「編集」では，複数の視点映像を一本の映像にまとめる編集技術に関して，映像から画像特徴量を抽出する手法の一つである「特徴点追跡手法」と，映像を意味のある最小単位に分割する「映像分割手法」および分割した映像群から視聴すべき視点映像を選択する「視点推薦手法」について先行研究をまとめた．

特徴点追跡の先行研究は，物体追跡手法の一種として議論されており，画像の局所領域から得られる特徴量に基づいて 2 フレーム間で類似する特徴点を対応づけることで特徴点

を追跡する技術とまとめることができる．特徴点の追跡手法として，画像上での特徴点の移動距離によらず特徴点の対応付けを行うことが可能な「特徴点マッチング」の手法と，特徴量の変化に対し頑健に特徴点の対応付けを行うことが可能な「Mean-Shift 探索」による手法をまとめた．

映像分割の先行研究は，複数の映像から一本の映像を編集するための技術の一部として議論されている．映像および映像に付随した様々な情報を元に，映像から最も視聴すべき部分を適切に切り出す技術としてまとめることができる．また，誰向けの映像を生成するためにどんな基準にしたがい映像を分割するのかという観点で先行研究を分類した．これより提案手法が，ある複数の視点を有する映像を，複数の視聴者が個人の興味に基づき視点を切替えながら視聴するために，複数人の視点切替えに対応できるように映像を分割するという点で新しいということを示した．

視点推薦の先行研究は，履歴に基づく手法とルールに基づく手法，機械学習に基づく手法の 3 つに分類される．履歴に基づく手法は，多数の視聴者が視聴した視点を他の視聴者に推薦する手法である．ルールに基づく手法は，映像文法などによる映像の評価値が高い視点を推薦する手法である．機械学習に基づく手法は，両者を組み合わせた手法であり，視聴者の視聴傾向を学習することにより，視聴者に応じた視点推薦が可能となる．これに対し，本手法が視聴者の視点切替えに応じて推薦する視点系列を変更する手法であることから，従来手法とは異なり視点切替えという少量の学習データから学習可能な機械学習に基づく手法に分類されることを示した．

「提示」では，複数の視点を有する映像を視聴するインタフェースに関し，視点選択の複雑さを解消することを目的として，視点と視線の操作性向上と被写体の視認性向上という観点から先行研究をまとめた．

以上を踏まえ，本論文の位置づけとして，フィールドを取り囲むように複数台の固定カメラとレンジセンサを配置し，これを用いてサッカーの試合を撮影した広視域角多視点映像を対象に，映像から抽出した選手の身体動作の変化量やレンジデータから抽出したフィールド上での選手位置などの情報を元に，視聴者の嗜好の変化に応じた視点を推薦しつつ，専用の UI を用いて視聴者が不便なく視点や視線を切替えながら本映像を視聴する人とシステムの協調視聴方式を前提として，本方式の構成要素である自動視点推薦機能を実現するための 2 つの基礎研究であることを示した．

第 3 章では，被写体とカメラとの幾何関係に基づき多視点映像から一般的な視聴者の視点切替えを過不足なく抽出し，これに基づき映像を分割することを特徴とする映像分割方法について研究成果を述べた．

広視域角多視点映像の視聴における視点切替えを表現した重み付き有向グラフの作成において，映像分割は，グラフのノードおよびリンクの作成に必要な重要な技術である．し

かし，多視点映像の視聴を対象として，複数の視聴者が視点を切替えるタイミングで映像を分割する手法は存在しない．

そこで，多視点映像に対する視聴者の視点切替えを分析し，そこからルールを導出することで，多視点映像から全ての視聴者の視点切替えを過不足なく自動で抽出する手法を開発した．具体的には，多視点映像の編集結果を収集し，視聴対象となる被写体に着目して分析を行った．その結果，一般的な視聴者は，ボールに着目して映像を視聴し，被写体であるボールとカメラとの幾何関係を考慮して視点を切替えていることが分かった．したがって，この規則を満たす幾何特徴を用いた映像分割アルゴリズムを開発した．広視域角多視点映像の編集結果を用いた評価実験では，提案手法の方が従来手法よりも過不足なく視聴者の視点切替えと同じタイミングで映像を分割できることを確認した．

第 4 章では，画像空間を粗密に探索することを特徴とするスポーツ選手の特徴点追跡方法について研究成果を述べた．

広視域角多視点映像の視聴における視点切替えを表現した重み付き有向グラフの作成において，スポーツ選手の特徴点追跡は，グラフのリンクのコストを設定するために用いる選手の身体動作の変化量の計算に必要な技術である．画像上でのサッカー選手の位置や大きさなどの画像特徴量は，俯瞰フィールド座標系での選手位置とカメラ幾何が分かれば算出することができる．しかし，選手の身体動作など，細かな画像特徴量は画像解析からしか得られない．そこで，画像から抽出したサッカー選手の特徴点の追跡結果を選手の身体動作の変化量とする．

特徴点追跡は，画像処理の基本技術であり，様々な手法が開発されている．しかし，サッカー選手は姿勢変化と速度変化が顕著なため，サッカー選手から抽出される特徴点は特徴量の変化と位置の変化が同時に発生してしまい，従来手法では追跡が困難である．そこで，特徴点の位置変化に対応するため大まかに特徴点を探索してから，特徴点の特徴量変化に対応するため詳細に特徴点を探索するという 2 段階の方法を開発した．具体的には，特徴点マッチングにより画像空間の広域を粗に探索し，その後，探索した特徴点を開始位置として Mean-Shift 探索により狭域を密に探索する．特徴点の移動量，移動速度，特徴量を変化させたシミュレーション実験において，提案手法は従来手法よりも特徴点を長時間正確に追跡できることを確認した．また，実際のサッカー映像を用いた選手の追跡においても，従来手法よりも長時間選手の特徴点を追跡できた．

5.2　今後の研究課題

本論文では，複数の視点を有する映像を視聴者が視点切替えしつつ不便なく視聴することができる人とシステムの協調視聴方式を前提に，この実現に必要な自動視点推薦機能の

開発の一環として「広視域角多視点映像の分割手法」と「サッカー選手の特徴点追跡手法」という 2 つの研究成果を述べた．第 5.2 節では，協調視聴方式の実現に向けた本論文および周辺技術の今後の展望として，研究課題をまとめる．

5.2.1　主要技術の展望

- **協調視聴方式**

  協調視聴方式は，複数の視点を有する映像を視点切替えしながら不便なく視聴する手法である．

  本論文では，協調視聴方式を前提としており，協調視聴方式自体の評価を実施していない．したがって，視聴者が視点切替えしながら不便なく映像を視聴可能かという観点から協調視聴方式の効果を評価する必要がある．
- **自動視点推薦部**

  自動視点推薦部は，視聴者の視点切替えから映像に対する嗜好の変化を読み取り，それに応じた視点系列を推薦するように設計している．これを実現しているのは，重み付き有向グラフのリンクの重み関数とその更新方法である．したがって，この重み関数と更新方法についてより深い議論を行うために，この観点から視点系列の生成結果を評価する必要がある．
- **広視域角多視点映像の分割方法**
  - **自由視点映像への拡張**

    本論文では，複数の視点を有する映像として，多視点映像を対象としている．そのため，視点と視線を表す 6 次元空間中に点在する視点と視線の組に対し，被写体とカメラとの 3 次元的な幾何関係を利用することで，映像分割を行っている．これに対し，複数の視点を有する映像の一つである自由視点映像を分割する方法への拡張が必要である．
  - **狭視域角映像への拡張**

    本論文では，複数の視点を有する映像として，広視域角映像を対象としている．そのため，被写空間の規模が大きく，視点間での視野の重複が少ないため，特定の被写体が映る視点と映らない視点が存在する．これに対し，被写体とカメラとの 3 次元的な幾何関係を利用することで，映像分割を行っている．

    一方，狭視域角映像は，被写空間が小さく，視点間での視野の重複が多いため，全ての視点映像に特定の被写体が映る．そのため，被写体の向きなども考慮した新たな映像分割方法が必要となる．
- **サッカー選手の特徴点追跡方法**
  - **環境条件に対する頑健性の強化**

本論文では，屋外で行われるサッカーの試合を被写体として選択し，冬場の比較的晴天の日に撮影した映像を利用している．そのため，映像の品質は高く，雲などによる映像中の短期的な輝度変化は比較的少ない．

一方で，実際のスポーツ中継を考えると，雨や雪などの悪天候の日や，夏場の日差しの強い日などでも試合が行われる．悪天候の場合は，視界が悪く，画質が低下すると考えられ，日差しが強い場合は，日向と日陰の輝度差が大きく異なると考えられる．この様な環境下でも頑健に動作する特徴点追跡手法の開発が必要である．

– 選手間のオクルージョンの対策

本論文では，サッカーの試合を被写体として選択している．サッカーは，広いフィールドに対し，被写体が 25 人（選手（11 人 ×2 チーム）+ 審判（3 人））と少数であるため，被写体間のオクルージョンが発生しにくい．また，サッカーの映像をスタジアムの 2 階席などから角度をつけて撮影していることもオクルージョンが発生しにくい要因の一つである．

しかし，スポーツの中にはレスリングなどの選手間の接触が多いものがある．また，実際のサッカー中継では，フィールドにカメラを設置して試合を撮影することもある．したがって，オクルージョンに頑健な特徴点追跡手法が必要である．

5.2.2 周辺技術の進展

- 手動視点選択部

本論文では，前提とする協調視聴方式の構成要素のうち自動視点推薦機能に着目しており，手動視点選択機能については先行研究を紹介し，そこから想定し得る視点切替えコントローラの構成として，直感的な視点切替えが可能な身体動作を用いたジェスチャ型コントローラと被写体の全体像を把握しやすい俯瞰視点型コントローラの併用を例として挙げた．

しかし，これらのコントローラが，本当に複数の視点を有する映像の視聴に適しているかは未確認である．したがって，これらのコントローラが，本映像の視聴において，効果的であることを検証する必要がある．

- 被写体や撮影環境に応じた最適なカメラ配置の自動算出

スタジアム運営において，スタジアムの多目的利用は収益性の面で効果的である．そのため，コンテンツに応じて，測定機器の数や位置，向きを適切に変更して撮影する必要がある．したがって，被写体や撮影環境の特性を考慮して，撮影機器

の最適な配置を自動で計算する方法が必要である．

- **撮影機器間の自動校正・自動同期**

  複数の視点を有する映像は，複数台の撮影機器を利用して測定される．そのため，撮影後の 2D 映像解析において有益な情報となる機器間の幾何関係や，視聴に影響を与える時間や色味の調整を行わなければならない．したがって，カメラ校正や時間同期，画像の色合わせなどを，簡便に行う方法が必要である．

  例えば，スタジアムにおけるカメラ校正については，フィールド管理が徹底されているためフィールド内に関係者以外が立ち入ることができない場合がある．そのような場合も考慮して，広域に分散配置された複数台のカメラを同時に校正する方法が必要である．また，カメラとレンジセンサなどの異種機器間の校正方法も必要である．

  次に，時刻同期については，従来同軸ケーブルなどを用いて，撮影機器を一つの制御システムに接続し，撮影開始命令を一斉に送ることで，同期の取れた撮影を行っている．しかし，カメラ台数の増加やコンテンツに応じた最適なカメラ配置を行うためには，コードレスで運用容易な撮影環境の構築が必要であると考えられる．

  最後に，画像の色合わせについては，各カメラの特性を理解し，あるカメラの色味に他のカメラの色味を統一する必要がある．ただし，撮影時刻や時期，天候に応じて，日照条件が変化し異なるため，コンテンツを撮影するたびに，あるいは撮影中においても自動的に実施可能な簡易な方法が必要である．

- **映像コンテンツの高品質化に向けた映像解析**

  複数の視点を有する映像は，視聴者が視点を切替えられる一方で，視点切替えに意識を奪われ，映像内容の理解に支障をきたす恐れがある．そのため，視聴者が映像内容を理解するのを助けるような情報を映像と共に提供することが，サービスとしての映像の価値を高めることにつながる．したがって，画像処理技術を用いて，映像からこのような付加情報を自動的に抽出する方法が必要である．

  サッカーを例に考えると，画像上での選手位置や大きさなどの一次情報や，一次情報を解析することで得られる選手の運動量やパスの回数などのプレーに関する二次情報，更に二次情報の解析から得られる攻守の切り替えや試合の戦略などの試合に関する三次情報など様々な情報を抽出する必要がある．

- **ビジネスモデルに応じたシステムの拡張**

  サッカー中継や野球中継は，リアルタイムに映像を伝送する必要がある．そのため，従来の中継では，会場に中継車や編集ブースを用意し，複数人で映像を確認・編集することで視点を切替えながら映像を編集しつつ配信を行っている．したがって，本システムをスポーツ中継で利用する場合は，実時間で処理を完了する必要が

ある．

一方で逆に考えると，実時間で処理が完了しない場合は，処理にかかる時間に応じたビジネス展開を検討すればよい．例えば，数秒で処理が完了する場合は，リアルタイム映像配信サービスとして利用することができる．数時間で処理が完了する場合は，試合終了後のハイライト映像の提示やニュース番組などで利用するコンテンツとなる．数日で処理が完了する場合は，インターネットを利用した映像配信サービスの一環として，高付加価値映像コンテンツとして有料配信するなどのビジネスが考えられる．各ビジネスに応じたシステム要件を再度検討し，必要な方向性を見極めて機能拡張していく必要がある．

謝辞

本論文をまとめるにあたり，名古屋大学 大学院情報学研究科 間瀬 健二 教授，石川 佳治 教授, 安田 孝美 教授，平山 高嗣 特任准教授には丁寧な指導と励ましを受け賜りました．また，私の在学中の御教授も併せて深く感謝するとともに厚く御礼申し上げます．

本論文の研究は，名古屋大学 大学院情報科学研究科 社会システム情報学専攻において，多数の方のご指導，ご協力により行われました．特に本論文の全てに関し，開始から今日にいたるまで親身にご指導頂きました名古屋大学 大学院情報学研究科 間瀬 健二 教授には心から感謝いたします．また，平山 高嗣 特任准教授によるご指導により，本論文のうちスポーツ選手の特徴点追跡に関する研究を大きく推進させることができたことに心から感謝の意を表します．また，榎堀 優 助教は，最も年齢が近い研究者ということもあり，日頃より多くのご指導を頂き，今後の研究者人生のお手本を示していただきましたことに感謝の意を表します．静岡大学 森田 純哉 准教授（現 間瀬研究室 協力教員）には，本論文のうち協調視聴方式のインタフェースに関して，認知情報学的観点から多くのご指導をいただきました．心より感謝の意を表します．三菱電機株式会社 丸谷 宜史 博士（元 間瀬研究室 協力教員）には，学部時代に研究の仕方を基礎からご指導頂き，研究の面白さを教えて頂きました．これが研究者人生の始まりと感じており，感謝の意を表します．立命館大学 情報理工学部 加藤 ジェーン 教授（元 間瀬研究室 准教授）には，多視点映像処理の実利用の観点から，保育園で実証実験された知見などを通して，ご指導を頂きましたことに感謝の意を表します．

研究生活を送るにあたり，産学技術総合研究所 知能システム研究部門 特別研究員 原 健翔 博士（間瀬研究室後期課程修了）と東京大学 工学部電子情報工学科 相澤・山崎研究室 特任研究員 Wang Xueting 博士（間瀬研究室後期課程修了）とは，学生という立場から多くの議論を重ね，研究を進めるにあたり様々な意見をいただきましたことをここに感謝します．また，川村 泰世 秘書には，生活面および精神面において，多くのサポートをしていただきましたことをここに感謝します．

最後に，両親には，大学進学から博士課程満期退学まで不自由なく学業に専念させていただいたことを感謝します．本論文の執筆と仕事の両立にあたり，妻 莉沙には，生活面で

多分に支えてもらいました．息子 春希からは，無邪気な笑顔を通して，沢山の元気をもらいました．日々の愛情と協力に心から感謝します．

参考文献

[1] PERFORM, “DAZN 革新的ライブ・オンデマンドスポーツサービス,” http://www.performgroup.jp/brands/dazn/.

[2] ワイズ・スポーツ株式会社, “事業内容,” https://live-sports.yahoo.co.jp/special/other/company/info/business.

[3] スカパー JSAT 株式会社, “事業内容,” https://www.sptvjsat.com/business/channel/.

[4] 株式会社 WOWOW, “事業内容,” https://corporate.wowow.co.jp/company/business/.

[5] 馬場口 登, “メディア理解による映像メディアの構造化,” 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 99, No. 183, pp. 39–46, 1999.

[6] Naho Inamoto and Hideo Saito, “Immersive Observation of Virtualized Soccer Match at Real Stadium Model,” ISMAR ’03 Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 188–197, 2003.

[7] Zhenli Zhou, Li Zhuo, Jing Zhang, and Xiaoguang Li, “A User–driven Interactive 3D Video Streaming Transmission System with Low Network Bandwidth Requirements,” Proceedings of 2012 14th International Conference on Advanced Communication Technology, pp. 19–22, 2012.

[8] 渡邊 哲哉, 北原 格, 亀田 能成, 大田 友一, “正確で直感的なカメラ操作を可能とする両手を用いた自由視点映像撮影インタフェース,” 電子情報通信学会論文誌 D, Vol. J95-D, No. 3, pp. 687–696, 2012.

[9] Tokai Shogo, “Pegged to Point Browsing: An Approach to Browse Multi-view Video with View-Switching, and its Applications,” ICPR 2008 workshop on Sensing Web, pp. 41–46, 2008.

[10] Takafumi Marutani, Kenji Mase, Toshiaki Fujii, and Tetsuya Kawamoto, “Multi-view Video Contents Viewing System by Synchronized Multi–view Streaming Architecture,” MM ’12 Proceedings of the 20th ACM international conference on Multimedia, 10.1145/2393347.2396440, pp. 1277–1278, 2012.

[11] Satoshi Gondo, Tomoo Tnoue, Kasumi Tarukawa, and Ken–ichi Okada, “Soccer Tac-

tics Analysis Supporting System Displaying the Player's Actions in Virtual Space," Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design, 10.1109/CSCWD.2014.6846909, pp. 581–586, 2014.

[12] Kenji Mase, Kosuke Niwa, and Takafumi Marutani, "Socially Assisted Multi–View Video Viewer," ICMI '11 Proceedings of the 13th international conference on multimodal interfaces, 10.1145/2070481.2070541, pp. 319–322, 2011.

[13] Changsong Shen, Chris Zhang, and Sidney Fels, "A Multi–Camera Surveillance System that Estimates Quality–of–View Measurement," 2007 IEEE International Conference on Image Processing, 10.1109/ICIP.2007.4379279, pp. 193–196, 2007.

[14] Zhenchen Wang and Stefan Poslad, "Personalising Live Sports Video Zooming," 2013 IEEE International Conference on Systems, Man, and Cybernetics, 10.1109/SMC.2013.578, pp. 3390–3395, 2013.

[15] Zhenchen Wang, "Personalizing Live Zooming using the ePlayer," Regular paper of multimedia system, 10.1007/s00530-013-0347-8, Vol. 20, No. 6, pp. 721–733, 2013.

[16] Yuki Muramatsu, Takatsugu Hirayama, and Kenji Mase, "Video Generation Method Based on User's Tendency of Viewpoint Selection for Multi–View Video Contents," AH '14 Proceedings of the 5th Augmented Human International Conference, 10.1145/2582051.2582052, 2014.

[17] Xueting Wang, Yuki Muramatu, Takatsugu Hirayama, and Kenji Mase, "Context–Dependent Viewpoint Sequence Recommendation System for Multi–View Video," 2014 IEEE International Symposium on Multimedia, 10.1109/ISM.2014.44, pp. 195–202, 2014.

[18] Xueting Wang, Yu Enokibori, Takatsugu Hirayama, Kensho Hara, and Kenji Mase, "User Group based Viewpoint Recommendation using User Attributes for Multiview Videos," MUSA2 '17 Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, 10.1145/3132515.3132523, pp. 3–9, 2017.

[19] 北原 格, 橋本 浩一郎, 亀田 能成, 大田 友一, "サッカーの自由視点映像提示における気の利いた視点選択手法," 日本バーチャルリアリティ学会論文誌, 10.18974/tvrsj.12.2_171, Vol. 12, No. 2, pp. 171–179, 2007.

[20] Akihiro Maehigashi, Kazuhisa Miwa, Hitoshi Terai, Kazuaki Kojima, and Junya Morita, "Experimental Investigation of Calibration and Resolution in Human–Automation System Interaction," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 10.1587/transfun.E96.A.1625, Vol. E96.A, No.7, pp. 1625–1636, 2013.

[21] Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object Tracking: A survey," Journal of

ACM Computing Surveys (CSUR), 10.1145/1177352.1177355, Vol. 38, No. 4, pp. 1–45, 2006.

[22] R. Venkatesh Babu, Patrick Perez, and Patrick Bouthemy, “Robust Tracking with Motion Estimation and Local Kernel-based Color Modeling,” Image and Vision Computing, 10.1016/j.imavis.2006.07.016, Vol. 25, No. 8, pp. 1205–1216, 2007.

[23] Hao Ji, Fei Su, and Geng Du, “Multiple Faces Tracking based on Joint Kernel Density Estimation and Robust Feature Descriptors,” Proceedings of IEEE International Conference on Network Infrastructure and Digital Content, pp. 680–685, 2009.

[24] 都築 勇司, 藤吉 弘亘, 金出 武雄, “SIFT 特徴量に基づく Mean–Shift 探索による特徴点追跡,” 情報処理学会論文誌 コンピュータビジョンとイメージメディア (CVIM), Vol. 49, No. SIG6(CVIM20), pp. 35–45, 2008.

[25] Huiyu Zhou, Yuan Yuan, and Chunmei Shi, “Object Tracking using SIFT Features and Mean Shift,” Computer Vision and Image Understanding, 10.1016/j.cviu.2008.08.006, Vol. 113, No. 3, pp. 345–352, 2009.

[26] Carlo Tomasi and Takeo Kanade, “Shape and Motion from Image Streams under Orthography: a Factorization Method,” International Journal of Computer Vision, 10.1007/BF00129684, Vol. 9, No. 2, pp. 137–154, 1992.

[27] H. Wang, A. Klaser, C. Schmid, and C.L. Liu, “Dense Trajectories and Motion Boundary Descriptors for Action Recognition,” International Journal of Computer Vision, Vol. 103, No. 1, pp. 60-79, 2013)

[28] Philippe Loic Marie Bouttefroy, Abdesselam Bouzerdoum, Son Lam Phung, Azeddine Beghdadi, “Vehicle Tracking by non–Drifting Mean–shift using Projective Kalman Filter,” 2008 11th International IEEE Conference on Intelligent Transportation Systems, 10.1109/ITSC.2008.4732659, pp. 61–66, 2008.

[29] Shangbo Zhou, Peng Hu, Kun Li, and Liu Yujiong, “A New Target Tracking Scheme Based on Improved Mean Shift and Adaptive Kalman Filter,” International Journal of Advancements in Computing Technology, 10.4156/ijact.vol4.issue2.35, Vol. 4, No. 2, pp. 291–301, 2012.

[30] R.T. Collins, “Mean–shift Blob Tracking through Scale Space,” Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 10.1109/CVPR.2003.1211475, Vol. 2, No. II, pp. 234–240, 2003.

[31] Zhiwei Zhu, Qiang Ji, K. Fujiwara, and Kuangchih Lee, “Combining Kalman Filtering and Mean Shift for Real Time Eye Tracking under Active IR Illumination,” Object recognition supported by user interaction for service robots, 10.1109/ICPR.2002.1047460, Vol. 4, pp. 318–321, 2002.

[32] Zhenhai Wang and Kicheon Hong, "A New Approach for Adaptive Background Object Tracking Based on Kalman Filter and Mean Shift," RACS '13 Proceedings of the 2013 Research in Adaptive and Convergent Systems, 10.1145/2513228.2513262, pp. 134–139, 2013.

[33] D. Comaniciu and V. Ramesh, "Mean Shift and Optimal Prediction for Efficient Object Tracking," Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101), 10.1109/ICIP.2000.899297, Vol. 3, pp. 70–73, 2000.

[34] Caifeng Shan, Tieniu Tan, and Yucheng Wei, "Real–time Hand Tracking using a Mean Shift Embedded Particle Filter," Pattern Recognition, 10.1016/j.patcog.2006.12.012, Vol. 40, No. 7, pp. 1958–1970, 2007.

[35] K. Deguchi, O. Kawanaka, and T. Okatani, "Object Tracking by the Mean–shift of Regional Color Distribution Combined with the Particle–filter Algorithms," Proceedings of the 17th International Conference on Pattern Recognition, 10.1109/ICPR.2004.1334577, Vol. 3, pp. 506–509, 2004.

[36] D. Comaniciu, V. Ramesh, and P. Meer, "Real–Time Tracking of non–Rigid Objects using Mean Shift," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662), 10.1109/CVPR.2000.854761, Vol. 2, pp. 142–149, 2000.

[37] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel–based Object Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, 10.1109/TPAMI.2003.1195991, Vol. 25, No. 5, pp. 564–577, 2003.

[38] D.G. Lowe, "Object Recognition from Local Scale–Invariant Features," Proceedings of the Seventh IEEE International Conference on Computer Vision, 10.1109/ICCV.1999.790410, Vol. 2, pp. 1150–1157, 1990.

[39] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "SURF: Speeded Up Robust Features," Proceedings of the 9th European Conference on Computer Vision, 10.1007/11744023_32, Vol. 3951, pp. 404–417, 2006.

[40] Barbara Zitova and Jan Flusser, "Image Registration Methods: a Survey," Image and Vision Computing, 10.1016/S0262–8856(03)00137–9, Vol. 21, No. 11, pp. 977–1000, 2003.

[41] M. Brown and D.G. Lowe, "Recognising Panoramas," Proceedings Ninth IEEE International Conference on Computer Vision, 10.1109/ICCV.2003.1238630, pp. 1218–1227, 2003.

[42] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," Proceedings of the Seventh IEEE International Conference on Computer Vision,

10.1109/ICCV.1999.790416, Vol. 2, pp. 1197–1203, 1999.

[43] R.E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” Journals of Fluids Engineering, 10.1115/1.3662552, Vol. 82, No. 1, pp. 35–45, 1960.

[44] Greg Welch and Gary Bishop, “An Introduction to the Kalman Filter,” Technical Report TR95–041, University of North Caroline, Chapel Hill, NC, 1995.

[45] Michael Isard and Andrew Blake, “CONDENSATION–Conditional Density Propagation for Visual Tracking,” International Journal of Computer Vision, Vol. 29, No. 1, pp. 5–28, 1998.

[46] Amjad Rehman and Tanzila Saba, “Feature Extraction for Soccer Video Semantic Analysis: Current Achievements and Remaining Issues,” Transactions on Artificial Intelligence Review, 10.1007/s10462–012–9319–1, Vol. 41, No. 3, pp. 451–461, 2014.

[47] Nguyen Huu Bach, Koichi Shinoda, and Sadaoki Furui, “Robust Scene Extraction Using Multi-Stream HMMs for Baseball Broadcast,” IEICE TRANSACTIONS on Information and Systems, 10.1093/ietisy/e89–d.9.2553, Vol. E89–D, No. 9, pp. 2553–2561, 200).

[48] 山田 一郎, 佐野 雅規, 住吉 英樹, 柴田 正啓, 八木 伸行, “アナウンサーと解説者のコメントを利用したサッカー番組セグメントメタデータ自動生成,” 電子情報通信学会論文誌 D, Vol. J89–D, No. 10, pp. 2328–2337, 2006.

[49] 新田 直子, 馬場口 登, “放送型スポーツ映像の意味内容獲得のためのストーリー分割法,” 電子情報通信学会論文誌 D, Vol. J86–D2, No. 8, pp. 1222–1233, 2003.

[50] Naoko Nitta, Yoshimasa Takahashi, and Noboru Babaguchi, “Automatic Personalized Video Abstraction for Sports Videos using Metadata,” Transactions on Multimedia Tools and Applications, 10.1007/s11042–008–0217–0, Vol. 41, No. 1, pp. 1–25, 2009.

[51] 石川 友哉, Yu Wang, 加藤 ジェーン, “監視カメラ映像を用いた幼稚園児の 1 日ダイジェスト自動生成,” 電気学会論文誌 C（電子 · 情報 · システム部門誌）, 10.1541/ieejeiss.131.385, Vol. 131, No. 2, pp. 385–392, 2011.

[52] 丸谷 宜史, 杉本 吉隆, 角所 孝, 美濃 導彦, “講師行動の統計的性質に基づいた講義撮影のための講義状況の認識,” 電子情報通信学会論文誌 D, Vol. J90–D, No. 10, pp. 2775–2786, 2007.

[53] 熊野 雅仁, 有木 康雄, 春藤 憲司, 塚田 清志, “映像文法に基づいた映像編集支援システムのための使用可能なショット区間の自動抽出,” 映像情報メディア学会誌, 10.3169/itej.57.829, Vol. 57, No. 7, pp. 829–839, 2003.

[54] Yasuyuki Sumi, Sadanori Ito, Tetsuya Matsuguchi, Sidney Fels, Shoichiro Iwasawa, Kenji Mase, Kiyoshi Kogure, and Norihiro Hagita, “Collaborative Capturing, Interpreting, and Sharing of Experiences,” Article in Personal and Ubiquitous Computing, 10.1007/s00779–006–0088–1, Vol. 11, No. 4, pp. 265–271, 2007.

[55] Takeo Kanade and P.J. Narayanan, “Virtualized Reality: Perspectives on 4D Digitization of Dynamic Events,” IEEE Computer Graphics and Applications, 10.1109/MCG.2007.72, Vol. 24, No. 3, pp. 32–40, 2007.

[56] 浮田 宗伯, “能動視覚エージェント群の密な情報交換による多数対象の実時間協調追跡,” 電子情報通信学会論文誌 D, Vol. J88–D1, No. 9, pp. 1438–1447, 2005.

[57] Thaddeus Beier and Shawn Neely, “Feature–based Image Metamorphosis,” SIGGRAPH ’92 Proceedings of the 19th annual conference on Computer graphics and interactive techniques, 10.1145/142920.134003, Vol. 26, No. 2, pp. 35–42, 1992.

[58] Steven M. Seitz and Charles R. Dyer, “Toward Image-based Scene Representation using View Morphing,” Proceedings of 13th International Conference on Pattern Recognition, 10.1109/ICPR.1996.545996, pp. 1–16, 1996.

[59] Jiangjian Xiao and Mubarak Shah, “Tri-View Morphing,” Computer Vision and Image Understanding, 10.1016/j.cviu.2004.03.014, Vol. 96, No. 3, pp. 345–366, 2004.

[60] Takashi Matsuyama, Shohei Nobuhara, Takeshi Takai, and Tony Tung, “3D Video and Its Applications,” book, Springer Publishing Company, 2012.

[61] 谷本 正幸, “自由視点テレビ FTV,” 電子情報通信学会論文誌 A, Vol. J89–A, No. 11, pp. 866–872, 2006.

[62] Atsushi Iwatsuki, Takatsugu Hirayama, and Kenji Mase, “Analysis of Soccer Coach’s Eye Gaze Behavior,” Published in 2013 2nd IAPR Asian Conference on Pattern Recognition, 10.1109/ACPR.2013.185, pp. 793–797, 2013.

研究業績一覧

主論文に関する研究業績

学術雑誌論文

[1] 冨安 史陽, 平山 高嗣, 間瀬 健二, “Kalman–Filter 予測を用いた特徴点マッチングと Mean–Shift を組み合わせた粗密探索に基づく特徴点追跡,” 画像電子学会論文誌, 10.11371/iieej.43.318, Vol. 43, No. 3, pp. 318–329, 2014.

[2] 冨安 史陽, Wang Xueting, 間瀬 健二, “ボールとカメラ間の位置関係を用いた広視域角多視点映像のためのカット抽出方法,” 画像電子学会論文誌, 10.11371/iieej.45.305, Vol. 45, No. 3, pp. 305–317, 2016.

国際学会論文

[3] Fumiharu Tomiyasu, Takatsugu Hirayama, and Kenji Mase, “Wide–range Feature Point Tracking with Corresponding Point Search and Accurate Feature Point Tracking with Mean-Shift,” 2013 2nd IAPR Asian Conference on Pattern Recognition, 10.1109/ACPR.2013.166, pp. 907–911, 2013.

[4] Fumiharu Tomiyasu and Kenji Mase, “Human–Machine Cooperative Viewing System for Wide–angle Multi–view Videos,” IUI Companion’ 15 Proceedings of the 20th International Conference on Intelligent User Interfaces Companion, 10.1145/2732158.2732171, pp. 85–88, 2015.

学会での口頭発表等

[5] 冨安 史陽, 平山 高嗣, 間瀬 健二, “対応点探索と Mean–Shift 探索の逐次処理による特徴点追跡,” 情報処理学会 画像の認識・理解シンポジウム (MIRU2013), SS6–7, 2 頁, 2013.

[6] 冨安 史陽, 平山 高嗣, 間瀬 健二, “Kalman–filter 予測を用いた特徴点マッチングと Mean–Shift 探索の統合による広域特徴点追跡,” 電子情報通信学会技術研究報告 (パターン認識・メディア処理), Vol. 113, No. 197, pp. 77–84, 2013.

[7] 冨安 史陽, 村松 祐希, 飯田 涼太郎, Wang Xueting, 米澤 朋子, 平山 高嗣, 間瀬 健二, “被写体追従視聴のための視点推薦型多視点映像視聴システム,” 情報処理学会 インタラクション 2014, A4–7, pp. 290–295, 2014.

[8] 冨安 史陽, 間瀬 健二, “広視域角多視点映像における視聴対象の移動予測を考慮した意味的に連続なフレーム群の抽出,” 電子情報通信学会技術研究報告（マルチメディア・仮想環境基礎）, Vol. 115, No. 495, pp. 175–180, 2016.

その他の研究業績

国際学会論文

[9] Fumiharu Tomiyasu, Takafumi Marutani, Shoji Kajita, and Kenji Mase, “A parallelepiped calibration method between high–resolution color camera and low–resolution depth camera using cubic calibration box,” 3rd International Conference on 3D Systems and Applications, S1–6, pp. 107–110, 2011.

[10] Yuma Kabeya, Fumiharu Tomiyasu, and Kenji Mase, “Semi–automatic Multiple Player Tracking of Soccer Games using Laser Range Finders,” AH’16 Proceedings of the 7th Augmented Human International Conference 2016, Poster session (Best Poster Award), 10.1145/2875194.2875222, 2 pages, 2016.

[11] Fumiharu Tomiyasu, Asako Yumoto, Yasuhiro Aoki, Yasuhiko Nakano, and Eishi Morimatsu, “Early Driver Drowsiness Detection Using Gaze Features in Combination with Driving Features,” 24th ITS World Congress, Technical/Scientific Sessions 29, 7 pages, 2017.

学会での口頭発表等

[12] 冨安 史陽, 丸谷 宜史, 藤井 俊彰, 梶田 将司, 間瀬 健二, “立方体モデルのフィッティングを利用したカラーカメラとデプスカメラ間のキャリブレーション,” 情報処理学会 画像の認識・理解シンポジウム (MIRU2011) 論文集, IS4-12, 8 頁, 2011.

[13] 汪 雪 åľů, 冨安 史陽, 平山 高嗣, 間瀬 健二, “レスリング競技映像における選手領域の自動追跡,” 平成 24 年度電子情報通信学会東海支部卒業研究発表会, OB1, 1 頁, 2013.

[14] 壁谷 勇磨, 冨安 史陽, 間瀬 健二, “レーザスキャナを用いたサッカー選手の半自動追

跡,” 情報処理学会 インタラクション 2015, B02, pp. 446–451, 2015.

[15] 方 超偉, 冨安 史陽, 間瀬 健二, “シーン中の人物に乗り移る 3D 自由視聴インタフェースの構築と評価,” 電子情報通信学会技術研究報告 (マルチメディア・仮想環境基礎), Vol. 114, No. 486, pp. 27–32, 2015.