

# 複数固定カメラ映像からのなめらかな回り込み映像の生成

武田利明<sup>†1</sup> 三上弾<sup>†1,†2</sup> 木全英明<sup>†1</sup>

**概要:** 複数固定カメラから注目被写体を中心にクリッピングを行い、被写体を中心に回り込む映像を生成するシステムにおいて被写体検出ノイズによる視聴者に知覚される品質低下を防ぐ手法を提案する。提案手法では、相対的に誤差の影響を小さくすることを考え、拡大縮小効果を加える（提案 I）、検出結果の多項式近似（提案 II）、それらの組み合わせ（提案 III）を検討する。3通りのコンテンツに対して、提案手法および従来手法により回り込む映像を作成した。これを被験者に見せ品質の評価を行った。実験により、提案によりぐると滑らかに回り込む映像が生成できることを確認した。ただし、効果には差があるため今後検証を行っていきたい。

**キーワード:** バレットタイム, 人物検出, 検出誤差

## Bullet time view synthesis from multiple fixed cameras

TOSHIAKI TAKEDA<sup>†1</sup> DAN MIKAMI<sup>†1</sup> HIDEAKI KIMATA<sup>†1</sup>

**Abstract:** This article proposes methods to prevent deterioration of video quality in bullet-time video that goes around a target object from multiple fixed cameras due to the noises included in target detection methods. We propose three methods focuses on reducing noises relatively; first one is the video-resizing, second one is the polynomial approximation of detected position of the target, and the last one is combination of them. We prepared three types of content, the above three proposed methods and conventional method are applied to the three content types, 12 content are generated in total. Ten subjects are watched and evaluated the content. As the result, our proposal improves the quality of content. However, we need further analysis to conclude the effect of the proposed methods.

**Keywords:** bullet-time, human detection, detection noise

### 1. はじめに

近年、バレットタイム撮影、マルチアングル映像撮影、自由視点映像合成など様々な映像撮影および表現技術が普及してきている。Ohtaら、三功らは、それぞれ自由視点映像合成およびその提示に関する研究を行っている[1][2]。自由視点映像合成はカメラが存在しない視点の映像を合成・提示可能にする技術であり非常に魅力的であるものの、今のところ簡易かつ広く利用される状況には至っていない。

一方で、Bullet time 映像表現はそのシンプルな処理のため、映画などに限らず、スポーツ中継などにおいても実利用されるなど様々な取り組みが行われてきている。ここで我々が注目する2つのシステムを紹介する。EyeVision システム[3]、および、ぐるっとビジョンシステム[4]である。これらは、スポーツ中継を対象として、被写体を「ぐるっと」まわって見る視覚効果を提供するものであり、視点位置が概ね固定されるスタジアム観戦とは全く異なる魅力を提供する。これらの研究の鍵となるのはカメラ制御であり、スタジアム外周部に取り付けられた複数のカメラの全てが注

目対象の被写体を向くように制御する技術である。

ところで、近年、カメラの高解像度化、低価格化が急速に進んでいる。その結果、高解像度映像の中から見たい部分を視聴者が選択して視聴するインタラクティブパノラマ技術[5]なども提案されてきている。

このような背景から我々は、複数の高解像度固定カメラによりフィールド全体を捉え、注目被写体のクリッピングにより「ぐるっと」回り込む映像表現を作り出すことを目指した研究を行っている。その実現のためには、各カメラのどの部分を、どのサイズでクリッピングするかを決定する必要がある。ただし、各カメラからクリッピングした画像をつなぎ合わせて映像にする場合、各カメラでのクリップのノイズにより、滑らかにつながることがあり、これが大きな課題となる。

本稿は、クリッピングに基づいた回り込み映像生成において、視聴者に映像が滑らかにつながっているように感じさせるための提案を行うものである。提案手法では、主にスポーツに注目し、回転する中心の被写体が人物であることを仮定する。注目被写体を、人物 2D 姿勢推定手法を用いて、その位置と大きさを検出する。これに基づいて注目被写体をクリッピングする。この際に検出誤差の合成映像への影響を抑える処理を加える。この処理が本稿での主な提案である。これにより回り込む映像がなめらかにつなが

<sup>†1</sup> NTT メディアインテリジェンス研究所  
NTT Media Intelligence Laboratories  
神奈川県横須賀市光の丘 1-1

<sup>†2</sup> NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories  
神奈川県厚木市森の里若宮 3-1

るように感じさせることを期待する。

以降、第2節でシステム全体の流れを述べる。そして第3節で提案のポイントとなるクリッピング時の処理について述べる。第4節で実験とその結果を述べる。

## 2. システム概要と課題

本システムでは複数の固定カメラから注目対象をクリッピングし、それらをつなぎ合わせることで、ぐるっと回り込んだ映像を作成する。その概念を図1に示す。図1に示す通り提案システムでは、複数の固定カメラにより映像が撮影されていることを前提とする。そして、それらのカメラ映像から、注目被写体をクリッピングし、視聴サイズへとリサイズする。それらをつなぎ合わせることにより、ぐるっと回り込む効果の映像を合成する。

ここで、このようなシステムの実現に向けて生じる課題について述べる。注目被写体の画像処理による検出を行うと、検出の誤差により合成される映像がなめらかにつながっているように感じられないという課題が生じる。ただし、近年では OpenPose [6]に代表される精度の良い人物姿勢推定器が提案されていることから、本節ではまずいくつかの題材を対象に、OpenPoseでの人物検出の性能評価を行う。題材は、以降の実験を意識して、スポーツシーンを周囲に配置した複数の固定カメラで撮影したものとする。ただし複数カメラでの実際の撮影は困難であったため、スポーツシーンを模した3DCGを作成し、視点を設定することで複数カメラ映像の代替とした。利用した3つのシーンを図2に示す。野球については-45度から45度までの90度を、20台のカメラで生成した。サッカーについては0度から180度までの180度を、各々21台のカメラで生成した。以降の説明において、野球、サッカー1（青いユニフォームの選手が注目対象）、サッカー2（赤いユニフォームの選手が注目対象）と呼ぶ。

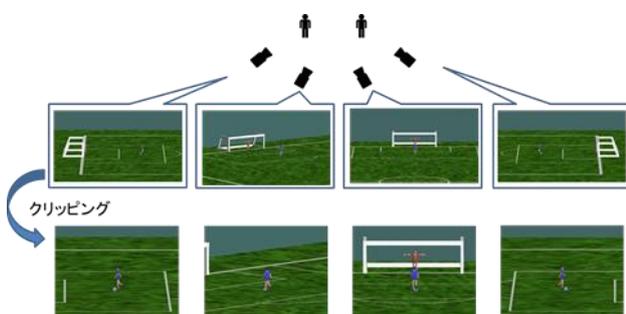


図1 システムの概念図；フィールド全体を写す複数の固定カメラ。各カメラの映像から注目被写体であるブルーの選手を切り出しぐるっと回り込む効果の映像を合成する

Fig. 1 System Overview; the system assumes the existence of multiple fixed cameras; target object, i.e., the player wearing blue, is clipped and resized to generate bullet-time view.

その時の人物検出結果を図3に示す。各図は水平方向および垂直方向の位置であり、横軸はカメラ位置を示す。真値との比較ではないものの、図3から検出にノイズが含まれていること、また検出を失敗する場面があることも見てとれる。このことから、検出ノイズは未だ解決がなされていない課題であるということが出来る。

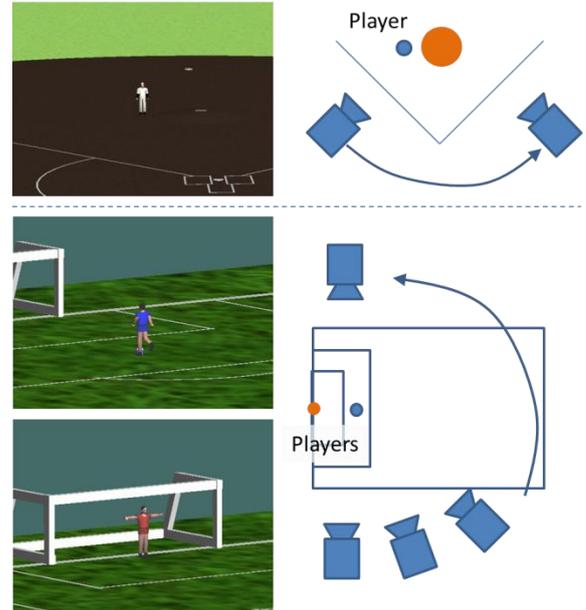


図2 コンテンツの例:上から野球, サッカー1, サッカー2. カメラの設定を右側に示す。

Fig.2 Example Content: Baseball, Soccer 1, and Soccer 2. The right side shows camera settings for the content.

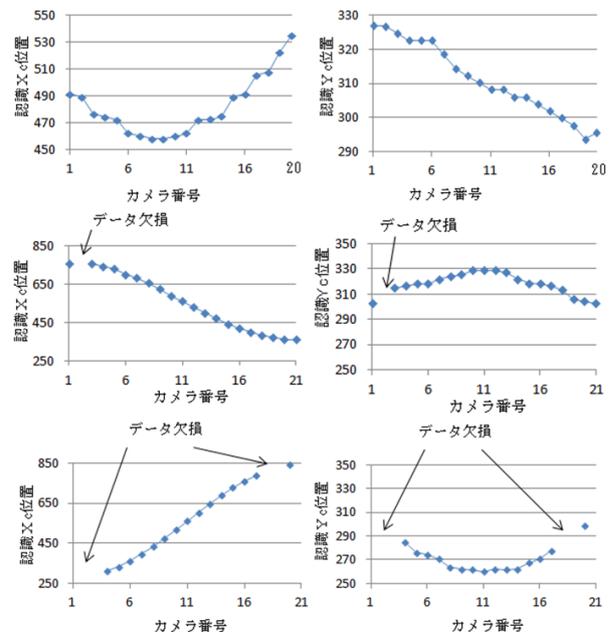


図3 人物検出結果; ノイズや欠損が確認できる  
 Fig. 3 Sample output of human object detection; we can see the noises and detection failure.

### 3. 提案手法

本稿では、クリッピングによるぐるっと回り込む映像表現を実現するために、検出の誤差が合成映像に与える影響を低減する方法を提案する。ここでは3つの方法を提案する。

#### 3.1 提案Ⅰ：拡大縮小によるノイズ低減

画像処理による人物検出の検出ノイズの大きさは、クリッピング前の画像のサイズに依存する。つまり、クリッピングサイズを変化させることで、相対的なノイズの大きさを変化させることが可能である。

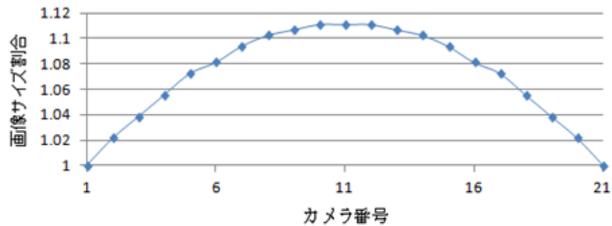


図4 サイズ変更の割合; クリップサイズが大きくなると拡大率は小さくなる

Fig. 4 Ratio of clipping size; large clipping size leads small magnification ratio.

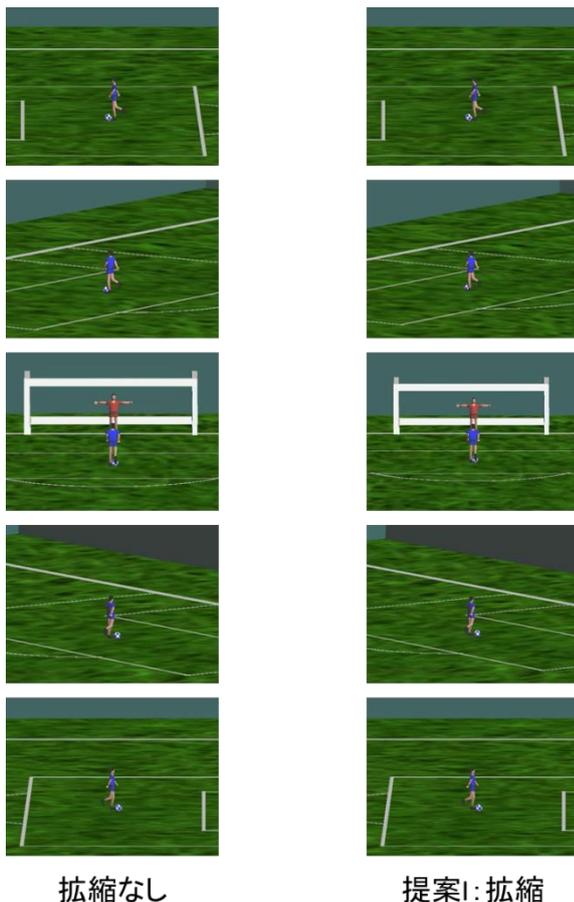


図5 拡大縮小前後の比較

Fig. 5 Comparison of images; before/after scaling

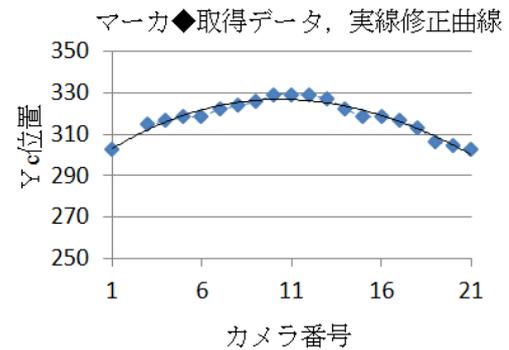
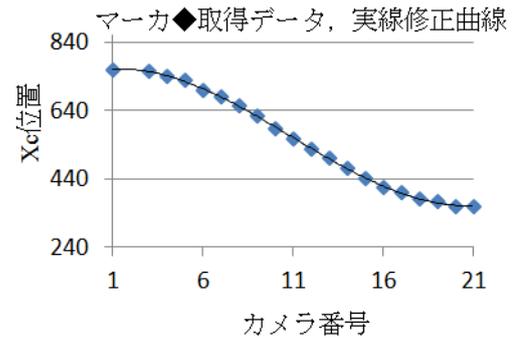


図6 多項式近似曲線

Figure 6 Correction curve

提案Ⅰにおける切り出し対象領域のサイズの割合を図4に示す。切り出し対象の領域が大きくなるということは、リサイズ後の画像内での注目対象の被写体の画像中でのサイズは小さくなるということである。これを図5に示す。図5において左列は切り出し対象の拡大縮小を行っていない状態であり、右列は提案Ⅰにおける切り出し結果である。

#### 3.2 提案Ⅱ：多項式近似によるノイズ低減

水平方向位置および垂直方向位置を各々多項式近似する。図6に例を示す。本稿の範囲では3次曲線による近似を行うものとする。これにより、ノイズの影響を受けず滑らかな回り込み効果となることを期待する。

#### 3.3 提案Ⅲ：多項式近似後+拡大縮小によるノイズの低減

提案Ⅲでは、提案Ⅱの多項式近似による修正を行った後に、さらに提案Ⅰの被写体を中心とした画像の切り取りサイズを変える。多項式近似の効果と、拡大縮小により相対的にノイズを抑える効果の両者により滑らかな提示が実現することが期待される。

## 4. 実験

### 4.1 コンテンツの作成

被験者実験にあたっては、前述の3種類のコンテンツを用いた。すなわち、野球場に人物が1名存在するコンテンツで、注目対象をその選手にしたもの、サッカー場にキッカーとキーパーの2名が存在するコンテンツで、注目対象をそれぞれキッカー、キーパーとしたものである。

これらのコンテンツに対して、人物検出結果に基づいたクリッピングによりリサイズしたコンテンツ（以降、従来手法）と、提案手法（提案I、提案II、および提案III）によりノイズを低減しながらリサイズしたコンテンツを作成した。なお、リサイズ後のコンテンツは全て520ピクセル×390ピクセルとし、フレームレート30fpsに設定した。ただし、映像は時間としては進まず、ある時刻において、カメラが切り替わる映像である。

### 4.2 実験設定および視聴環境

本研究では、検出ノイズの影響を抑えて視聴者が主観的に、注目対象の被験者を中心として滑らかな回り込みを感じるか、という点が重要である。そこで被験者実験ではコンテンツおよび被写体に関して、

1. なめらかに動いているか、
2. 変化が気になるか、
3. 見やすいか、
4. 回り込んで見ているように感じるか、

の4項目を、それぞれ5段階評価させた。被験者は20歳台から50歳台の男女10名である。

実験における機器の設置環境を図7に示す。映像の提示には40インチディスプレイを利用した。被験者は、ディスプレイ前方2.0mの位置に設置されたソファに腰かけて映像を視聴した。

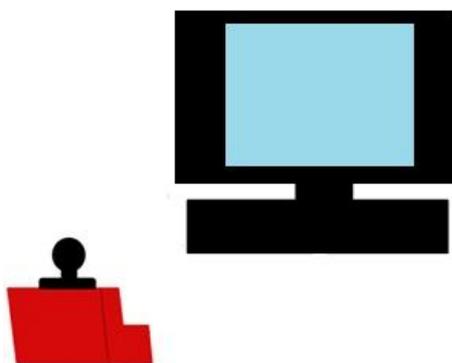


図7 設置環境

Figure 7 Installation environment

### 4.3 実験手順

実験の手順を示す。

1. インストラクション  
すべてのコンテンツが注目対象であるひとりの人物を映していること、そして、視聴後に前述の4項目について直感的に質問することを伝えた。
2. 事前視聴  
すべてのコンテンツを1回視聴させた。これは、視聴対象全体の品質の幅・範囲を視聴者に認識させることを目的としている。
3. 実際の視聴実験  
用意したコンテンツを視聴させ、評価項目について5段階評価させた。なお、順序効果を防ぐために提示順はランダムとした。被験者はひとつのコンテンツ視聴を行った直後に評価した。

質問項目の詳細は以下の通りである。

- なめらかに動いているか  
5:強くなめらか、4:ややなめらか、3:どちらでもない、2:ややなめらかでない、1:強くなめらかでない
- 変化が気になるか  
5:全く気にならない、4:やや気にならない、3:どちらでもない、2:やや気になる、1:強く気になる
- 見やすいか  
5:見やすい、4:やや見やすい、3:どちらでもない、2:やや見づらい、1:見づらい
- 回り込んで見ているように感じるか  
5:強く感じる、4:やや強く感じる、3:どちらでもない、2:やや弱く感じる、1:弱く感じる

## 5. 実験結果と考察

3つのコンテンツを通しての評価結果を図8に示す。また、コンテンツ毎に集計した結果を図9~11に示す。すべての図において、5%水準で有意傾向が確認できた箇所\*を記す。

従来手法と提案I、および提案IIと提案IIIを比較すると、視点変化における映像の滑らかさ、見やすさなどの項目で評価があがっていることが分かる。これは、全体に対する評価（図8）とコンテンツ毎の評価（図9~11）とで共通して確認できる傾向である。ただし、効果はさほど大きくなく、拡大縮小の割合と検出ノイズの程度などを考慮しながら更なる検証を行っていく必要がある。

一方で、従来手法と提案II、提案Iと提案IIIを比較すると、すべての評価項目において大幅な評価スコアの上昇が確認できる。このことから、検出ノイズが存在する環境下では、近似により切り出し位置を調整することで視聴者の体感品質を大幅に向上させる効果があることが分かった。

さらに詳しくみると、従来手法と提案 I では、いくつかのコンテンツ・項目で評価の低下がみられる。しかし、その場合であっても、提案 II から提案 III では評価が上昇している。注目被写体の拡大縮小は、コンテンツ生成における被写体の魅力などを向上させる効果があることが指摘されており [7], ノイズ低減の効果は小さいものの、多項式近似によりノイズが十分に減らされた後に、拡大縮小効果が加わったことで、コンテンツ全体の魅力が増した結果、評価の上昇につながった可能性もあると考えている。この点に関しても今後より詳細な実験を行ってきたい。

実験後にインタビューを行っており、そこでの意見を紹介する。ぐるっと回り込む映像の体験が新しくコンテンツに没入してしまったり、全体的にダイナミクスを感じてしまったりした結果、あまり差が無い評価になってしまったという意見があった。今回の実験でも事前にコンテンツを確認させ評価しやすくする配慮を行っていたが、今後評価実験の方法についてもさらに工夫を行ってきたい。

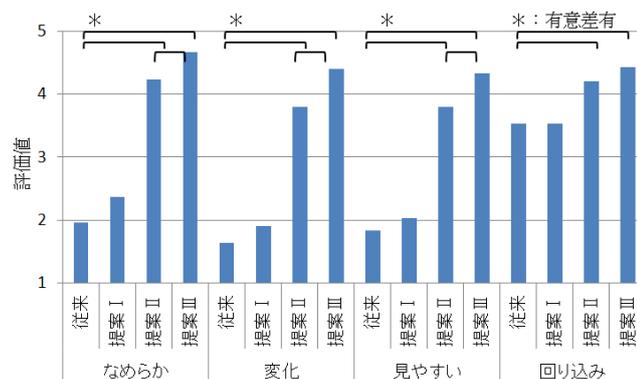


図 8 評価値／全体

Figure 8 Evaluation value/ Overall

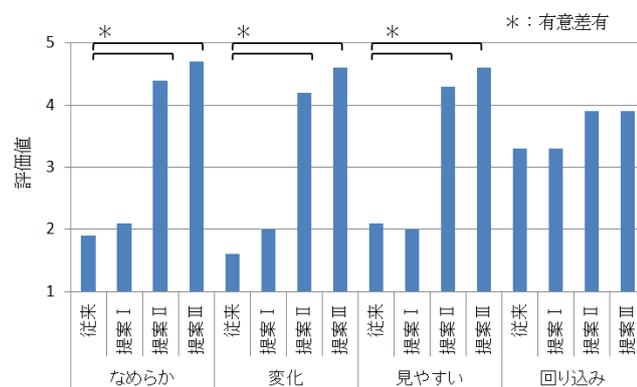


図 9 評価値／野球

Figure 9 Evaluation value/baseball

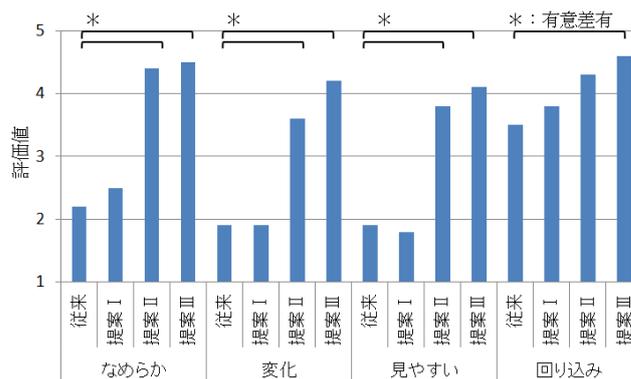


図 10 評価値／サッカー 1

Figure 10 Evaluation value/ soccer1

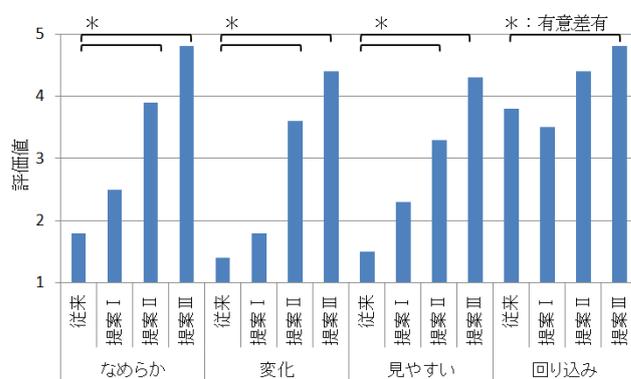


図 11 評価値／サッカー 2

Figure 11 Evaluation value/ soccer2

## 6. まとめ

人物検出の誤差に伴う、回り込み映像の誤差を抑制する方法として、クリッピングを行う画像サイズに対する拡張効果の付与 (提案 I), 多項式近似によるクリッピング位置の修正 (提案 II), さらにそれらの組み合わせ (提案 III) を検討し、主観評価実験を行った。

- 画面サイズを変えることで、誤差比率は小さくなるものの、その合成映像の品質に対する影響は限定的である。
- 一方、多項式近似では被験者実験評価値の大幅な上昇がみられた。従来法から提案法 II, 従来法から提案法 III については、なめらかさ、変化、見やすいにおいて全て有意な差が確認でき、大きな効果があると推測される。
- また、提案 II から提案 III では、従来手法から提案 I と比較して、評価値の上昇が見られている。この原因などについては今後より多くの実験などを通して確認してきたい。

今回は、撮影の都合などで、3DCGにより作成した空間に仮想的なカメラを配置して映像生成を行った。しかし、より現実に近い状態での被験者実験を行うためには、実写あるいはより精巧なCGでの実験が望ましいと考えており検討したい。

また、現状では、拡大縮小の大きさや、多項式近似の際のパラメータなど精査が不足している点がいくつか存在する。この点の詳細な実験などを同時に行うことでより現実的な環境に即した検証を行っていききたい。

## 参考文献

- [1] Yuichi Ohta, Itaru Kitahara, Yoshinari Kameda, Hiroyuki Ishikawa and Takayoshi Koyama, "Live 3D Video in Soccer Stadium", *International Journal of Computer Vision (IJCV)*, vol.75, no.1, pp.173-187, 2007.
- [2] 三功他, "選手領域の抽出と追跡によるサッカーの自由視点映像生成", *映メ学会誌* 68 卷 3 号, p. J125-J134, 2014.
- [3] Hideo Saito, Shigeyuki Baba and Takeo Kanade, *Appearance-Based Virtual View Generation From Multicamera Videos Captured in the 3-D Room*, *IEEE Transactions on Multimedia*, Vol. 5, No. 3, pp. 303 - 316, September, 2003
- [4] 池谷他, 多視点ロボットカメラを用いたぐるっとビジョンシステム, *映メ学会年次大会講演*, 2013.
- [5] Ochi et al., "A Study on Making Camera Trajectory from Panorama Watching Manipulation", *ACMMM2012*, 2012.
- [6] Zhe Cao and Tomas Simon and Shih-En Wei and Yaser Sheikh, *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*, *CVPR*, 2017.
- [7] 武田利明, 三上弾, 小澤史朗, 小島明, "被写体サイズの変化によるコンテンツおよび被写体の主観評価に対する影響の考察," *日本感性工学会論文誌*, Vol. 14, No. 2 p. 335-341, 2015.