

フレーム補間処理を用いた非同期多視点映像の滑らかな切り換え法

原崎 葵[†] 亀田 能成[‡] 北原 格[‡]

[†] 筑波大学 大学院システム情報工学研究科 〒305-8577 茨城県つくば市天王台 1-1-1

[‡] 筑波大学 計算科学研究センター 〒305-8577 茨城県つくば市天王台 1-1-1

E-mail: [†] s1520814@u.tsukuba.ac.jp, [‡] {kameda, kitahara}@iit.tsukuba.ac.jp

あらまし 非同期多視点映像から生成したバレットタイム映像で生じる被写体のブレを、フレーム補間処理を用いて低減することにより、滑らかな視点切り換え映像を生成する手法を提案する。各多視点カメラにおいて連続フレーム間で補間処理を行う際、初期補間画像において被写体の見え方を復元する手がかり情報を適切に与えることにより、補間処理結果を向上させる。その上で、視点移動時の被写体の観測位置の変化が最も小さくなるような補間画像を選択することにより、滑らかな視点切り換えのバレットタイム映像を生成する。

キーワード 自由視点映像, バレットタイム, フレーム補間, モーフィング, 非同期多視点映像

1. はじめに

被写体を様々な角度から観察可能な自由視点映像技術の研究が盛んに行われている[1]-[3]。自由視点映像では映像閲覧者が見たい位置から被写体を観察することが可能であり、被写体の状態や距離感のような空間的認識がしやすいという特長を有するため、映画などの映像産業やスポーツ科学などの応用分野で注目が集まりつつある。

従来の自由視点映像技術では、多視点映像から被写体の3次元CGモデルを生成し、3次元CG空間中で仮想カメラを操作することで様々な視点からの見え方を再現している[1][2][4]。しかし、3次元形状の推定誤差による生成映像の品質低下や、高い計算コストなどの問題が存在する。人物ビルボード法[5]を用いて被写体形状を2次元平面に近似することにより、推定誤差の影響を軽減することができるが、被写体の位置推定や領域分割の誤差による映像品質低下の問題は依然残っている。

バレットタイム映像は、被写体を取り囲むように複数台のカメラを配置し、撮影した多視点映像を連続的に切り換えていくことで、カメラ間を仮想的に視点が移動しているような視覚効果を再現することができる[6]。撮影映像をほぼそのまま提示するため、処理誤差による映像劣化が少なく高品質な映像提示が可能である。我々は、バレットタイム映像技術に基づいた自由視点映像生成提示技術に関する研究に取り組んでいる。

サッカースタジアムのような大規模空間における多視点映像撮影では、全カメラの同期撮影には膨大な労力と特殊な機材が必要となる。一方で、非同期撮影の多視点映像から生成したバレットタイム映像では、視点切り換えの際に同期ズレ

の影響により高速で移動する被写体の画像上での観測位置がブレてしまい、観察者に違和感を与えるといった問題が存在する。同一シーンを3視点から非同期撮影した例を図1に示す。同期ズレにより、高速で移動するサッカーボールとそれを蹴る選手との距離が画像毎に異なっている。このようなズレの発生した画像を用いてバレットタイム映像を生成すると、被写体の観測位置が画像上でブレたような映像が生成される。



図1 非同期撮影による観測位置の変化

本研究では、フレーム補間処理を用いて上述した観測位置のブレを低減し、滑らかな視点切り換え映像を生成することを目的とする。各多視点カメラにおいて注目フレームの前後フレームを対象としたフレーム補間処理を行い、その結果得られる補間画像群の中から、視点移動時の被写体の観測位置の変化が最も滑らかになるような補間画像を選択することにより、違和感の軽減を試みる。

本システムの流れを図2に示す。まず大規模空間への展開が容易な汎用カメラで構成される撮影システムを用いて非同期多視点撮影を行う。撮影した多視点映像から生成したバレットタイム映像では、視点切り換え時に同期ズレの影響により被写体の観測位置がブレる問題が発生する。提案手法を適応することで被写体の観測位置のブレが低減されたバレットタイム映像を生成する。

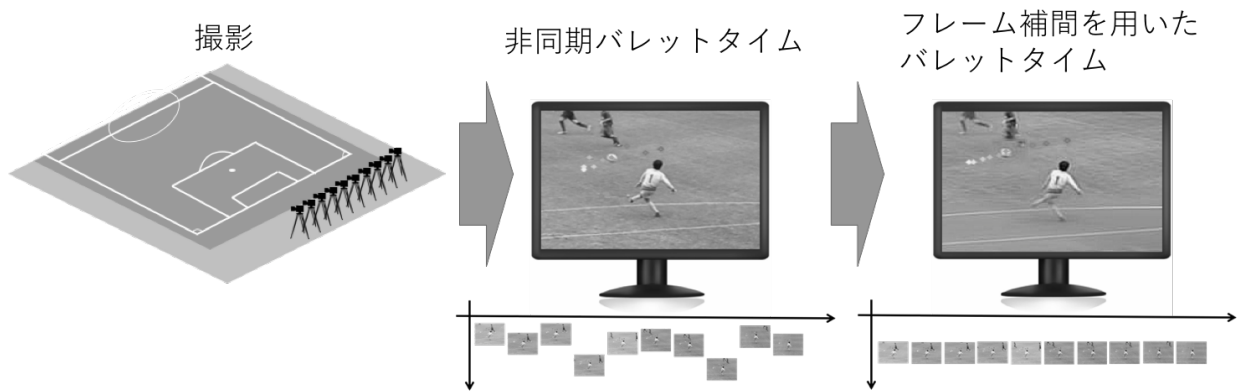


図2 撮影から提案手法の適応までの流れ

2. 関連研究

2.1. バレットタイム映像

Eye Vision[7]は実用事例の一つである。全てのカメラの光軸が撮影空間中の注目点で交差するようロボット雲台を制御することにより、アメリカンフットボール中継でのバレットタイム映像提示を実現した。Tomiyamaら[8]は、高価で設置に時間のかかるロボット雲台を使用せず、各多視点カメラを被写体のおおよその位置に向くように設置し、撮影映像に2次元射影変換を行うことで、仮想的なバレットタイム映像を実現した。明智ら[6]は、Structure from Motionを用いて注視点移動処理に必要なカメラパラメータを取得することにより、撮影空間中にランドマークを設置することなく多視点カメラの射影変換行列を推定し、その上で、注視点を自由に設定可能な多重解像度バレットタイム映像を提案した。

2.2. 多視点映像の同期撮影

バレットタイム映像などの自由視点映像生成のためには多視点映像撮影が必要である。Kanadeら[2]は直径5mの半円状のドームに複数台のカメラを設置してドーム内の被写体を撮影し、取得した実画像とデプス情報から、仮想空間での3次元CGモデルの復元を可能としている。Yanimotoら[9]は、100台のカメラを線形状や円形状、縦横の2次元アレイ状に配置した撮影システム（Free Viewpoint Television : FTV）を構築し、光線空間を用いた自由視点映像生成を実現している。これらの手法では、全てのカメラに信号配信用のケーブルを接続し、同期信号分配器から同期信号を配信しているため、大規模化に伴い設営が難しくなる問題が存在する。Kitaharaら[10]は、GPS信号から生成した外部同期信号を多視点カメラに配信することで、大規模空間に適用可能な同期撮影方式を提案している[11]。

明智ら[6]は、音声情報手がかりとして、大規模空間におい

て非同期撮影した多視点映像をソフト的に同期させている[12]。しかし、各カメラの内部クロックを同期させるわけではないため、微小ではあるが同期ズレが残る。また、音が反響しやすい環境ではその影響で同期精度が低下する。

2.3. モーフィング法

フレーム補間処理を実現するモーフィング法とは、二つの画像の中間の見え方を合成することにより、見え方が連続的に変化する一連の画像群を生成する手法である[13]。一般的なモーフィング法[14]では、2枚の入力画像間の対応点情報が必要となるが、毎秒30枚以上のフレームで構成される映像データを処理する場合、手動で正確な対応情報を付与することは現実的ではない。

本研究では、入力画像間の対応情報を用いずに補間画像の自動生成が可能なRegenerative Morphing[15]を採用する。Regenerative Morphingでは、モーフィング対象となる2枚の入力画像の単純なブレンディング画像を初期補間画像とし、その初期補間画像と入力画像間のパッチ領域の類似度に基づいて、補間画像を生成するため、被写体が高速に移動する場合には、初期補間画像上の適切な位置に被写体の見え方の手がかりとなる情報が存在しなくなり、補間処理結果が期待する見え方にならないといった問題が発生する。

3. 多視点映像切り換え処理

提案手法は主に二つの処理によって構成される。一つ目は、各視点においてフレーム補間処理を用いて連続する前後フレーム間に複数枚の補間画像を生成する処理である。二つ目は、生成された補間画像のうち、被写体の観測位置の変化の少ない画像を選択することで切り換え時の違和感を低減したバレットタイム映像を生成する処理である。

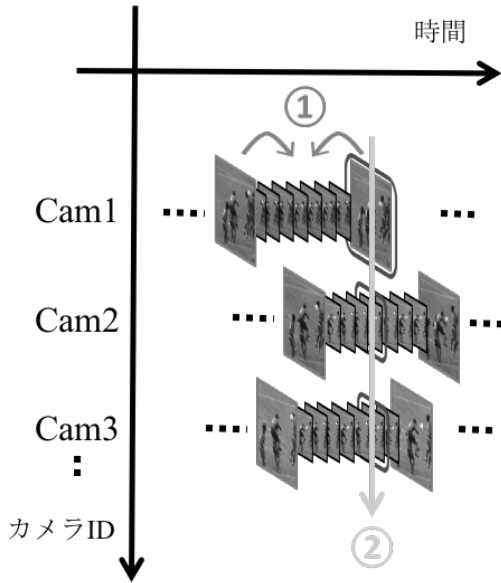


図3 提案手法の概略図

図3に提案手法の概略図を示す。横軸が時間軸で、両端の画像は撮影画像（左側が n 番目のフレーム、右側が $n+1$ 番目のフレーム）である。また縦軸はカメラIDであり、各カメラの撮影のタイミングに少しずつズレが生じていることを示している。まず、各カメラにおいてモーフィング法を用いて、撮影画像間にあるような画像で示す複数枚の補間画像を生成する。次にバレットタイム映像において、高速で移動する被写体の画像上の観測位置のブレが最も少ない画像を補間画像から選択し、切り換え画像として使用する。

4. フレーム補間処理法

各カメラにおいて注目フレームの前後フレームを対象としたフレーム補間処理について述べる。フレーム補間処理にはモーフィング法として **Regenerative Morphing** を採用する。その際、高速で移動する被写体に対しては、簡易的な対応情報を手がかりに被写体にモーションブレンダーがかかったような初期補間画像を使用することにより、適切な位置に被写体が観測される補間画像を生成する。

4.1. Regenerative Morphing の有用性と問題点

注目フレームの前後フレームに対して **Regenerative Morphing** を適用し補間画像を生成する。**Regenerative Morphing** では入力画像と出力画像（補間画像）間の双方向類似性を用いて補間画像を生成する。図4は入力画像 S_1 と S_2 の間の連続する複数枚の補間画像群 ($\dots T_{n-1}, T_n, T_{n+1} \dots$) を示している。注目補間画像 T_n では三角やひし形のように隣接する

前後の補間画像との一貫性と、丸や四角のように入力画像との類似性を用いた双方向類似関数[16]によって定義されるエネルギー最適化問題として補間画像生成を実現する。しかし、入力画像の単純なブレンド結果を初期の補間画像とするため、被写体が高速で移動する場合、初期補間画像上の適切な位置に被写体の見え方情報の手がかりが存在しなくなり、補間画像が期待する見え方にならない問題が存在する。

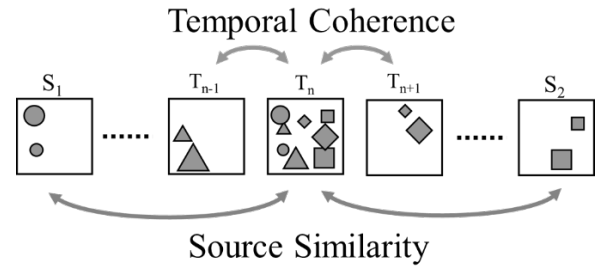


図4 Regenerative Morphing の概略図

図5は入力画像 S_1 と S_2 の間に連続的に複数枚の補間画像 ($\dots T_{n-1}, T_n, T_{n+1} \dots$) があることを示している。(a)の補間画像は入力画像間の内挿距離によってブレンドに重みをつけた初期画像となり、(b)はそこから生成された補間画像である。(a)の初期画像で被写体が観測されるべき (S_1 と S_2 での観測位置の中間) 位置付近に被写体の見え方の手がかりがないため、(b)の補間結果（補間画像 T_n ）では期待した位置（中央付近）に被写体像が再現されていない。この問題は被写体が高速で移動する場合に顕著に現れる。

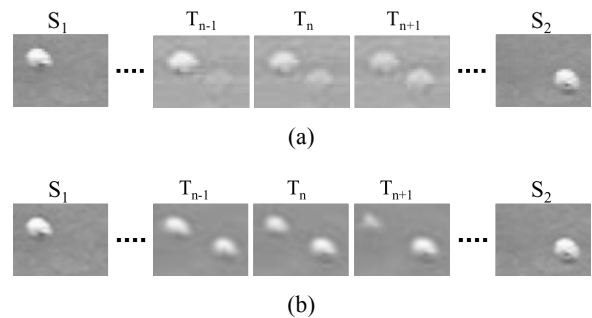


図5 Regenerative Morphing の補間画像

4.2. 初期補間画像生成法

高速で移動する被写体に対して **Regenerative Morphing** を実行した際に生じる問題の解決法を述べる。**Regenerative Morphing** では入力画像同士をブレンドした画像を初期の補間画像として使用し、見え方を更新しながら補間画像を生成する。初期補間画像の適切な位置に見え方の手がかりが存在しない場合、更新を何度繰り返しても期待した見え方に

はならない。本手法では、入力画像間で被写体の対応情報を一組だけ与え、そこからモーションブラーが生じたような初期画像を作成することにより、被写体の見え方の手がかりを与える。モーションブラーを付加する領域は被写体の大きさに依存する。図6に示すように被写体の大きさを H 、距離を D とし、焦点距離を d 、撮像素子サイズを x 、画素数を X とした時の画像上での被写体の大きさ h は式(2)によって求められる。

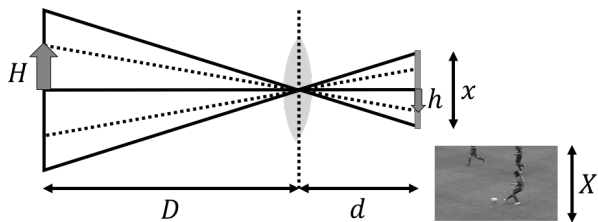


図6 画像上での被写体の大きさ

$$H:h \frac{x}{X} = D:d \quad (1)$$

$$h = Hd \times \frac{X}{Dx} \quad (2)$$

入力画像間で、内挿距離が異なる複数の補間画像を生成するため、式(3)に示す正規分布を用いて、2枚の入力画像との内挿距離に応じた被写体と背景のブレンド比率を決定する。入力画像間の位置を t とし、入力画像との内挿距離の割合を T_n 、分散値を σ^2 とする。

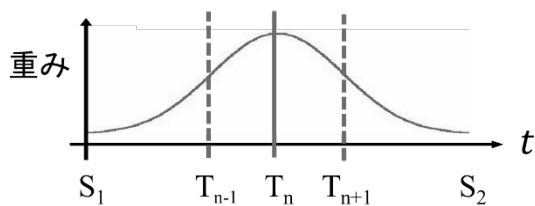


図7 被写体と背景のブレンド比率

$$f_{T_n}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-T_n)^2}{2\sigma^2}\right) \quad (3)$$

図8は、(a)が入力画像との距離に合わせてモーションブラーに重み付けをした初期補間画像、(b)がそれを用いて生成された補間画像を示している。補間画像 T_n を見ると、期待している画像中心付近に被写体が生成されていることがわかる。また T_{n-1} と T_{n+1} からは被写体の観測位置が連続的に移動していることがわかる。

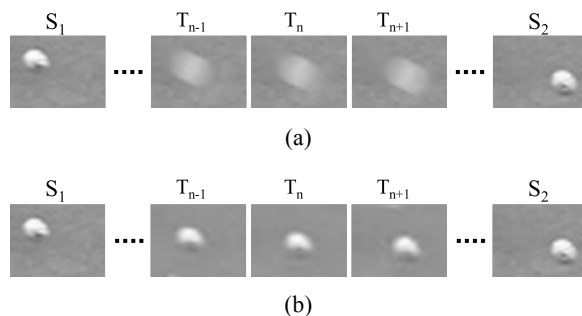


図8 提案手法での補間画像

5. 補間画像選択法

5.1. 視点切り換え時の被写体の動き

図9にバレットタイム映像における視点切り換え時の被写体の動きを示す。多視点カメラを切り換えた際に被写体が観測される位置を白色から黒色へと変化する点で示している。左の図は点線で囲んだサッカーボールを蹴った選手、中央は点線で囲んだ画像上部の選手、右はサッカーボールが被写体の場合である。注視点であるサッカーボールを蹴った選手は画像が切り換わっても、画像上での観測位置がほぼ一箇所に集まっている。画像上部の選手の移動はカメラのシャッター速度に対して十分に遅いため、同期ズレの影響が小さく、結果として観測位置はある曲線上を一定方向に移動している。一方で、高速で移動するサッカーボールは、同期ズレの影響を大きく受けるため観測位置がブレながら移動している。本手法では、この観測位置のブレを抑えることにより、滑らかな視点切り換えを実現する。

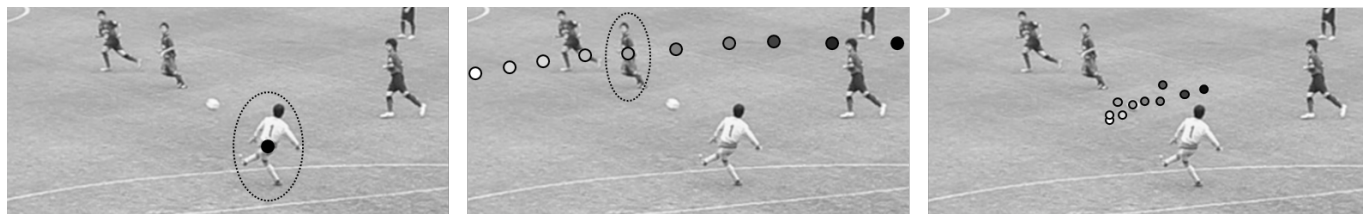


図9 被写体の画像上の観測位置

5.2. 切り換え画像選択手法

多視点カメラの入力画像と全ての補間画像において、被写体の観測位置を求め、図 10 に示すように、カメラ n の画像上での座標 $P_n(x_n, y_n)$ とカメラ $n+1$ の画像上での座標 $P_{n+1}(x_{n+1}, y_{n+1})$ から求めたベクトル \mathbf{a}_n と、カメラ $n+1$ の画像上での座標 $P_{n+1}(x_{n+1}, y_{n+1})$ とカメラ $n+2$ の画像上での座標 $P_{n+2}(x_{n+2}, y_{n+2})$ から求めたベクトル \mathbf{a}_{n+1} のなす角度 θ_n を算出する。式(4)に示すように、全カメラ (Cam1,...,CamN) について θ_n を累積する。同様の処理を多視点カメラの入力画像と補間画像の全てについて行い、累積値が最小となる画像を選択する。二つのベクトルのなす角が小さいほど3点の座標が一定の曲線状に存在するので、式(4)が小さいほど被写体の観測位置の変化が小さく、滑らかな画像切り換えになると考えられる。

$$\sum_{n=1}^{N-1} \theta_n = \sum_{n=1}^{N-1} \cos^{-1}\left(\frac{\mathbf{a}_n \cdot \mathbf{a}_{n+1}}{|\mathbf{a}_n| |\mathbf{a}_{n+1}|}\right) \quad (4)$$

$$\mathbf{a}_n = (x_{n+1} - x_n, y_{n+1} - y_n) \quad (5)$$

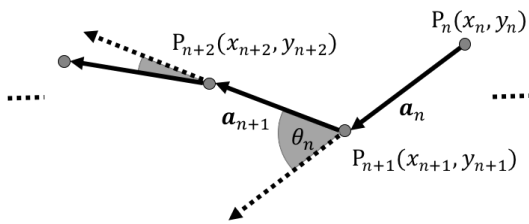


図 10 観測座標の変化

6. 実証実験

6.1. 実験環境

カシマサッカースタジアムで開催されたサッカーの試合のペナルティーエリア付近を、図 11 に示すゴール裏観客席に配置した 10 台のカメラで撮影した。隣り合うカメラの輻輳角度は約 10 度である。カメラはキヤノン社 EOS 5D Mark II を使用し、1,920 画素×1,080 画素の動画を毎秒 30 フレーム撮影した。映像処理は、CPU : Intel Core i7 3.4GHz、GPU : Intel HD Graphics Family、メモリ : 8.0GRAM を搭載した計算機を用いて処理する。

6.2. 実験結果と考察

図 12 に上記の環境で撮影したシーケンスに提案手法を適

応した処理結果を示す。(a)は撮影した多視点画像の切り換えによるバレットタイム映像、(b)が提案手法を用いたバレットタイム映像である。サッカーボールの観測位置を白色から黒色へと変わる点で示している。画像を切り換える際、(a)では、観測位置の軌跡が大きくブレているが、(b)ではそれが解消されていることがわかる。3点の観測座標から求めた角度の合計を最小にすることで、全ての観測座標が統一感のあるものになり、より滑らかな切り換えが実現されていることがわかる。

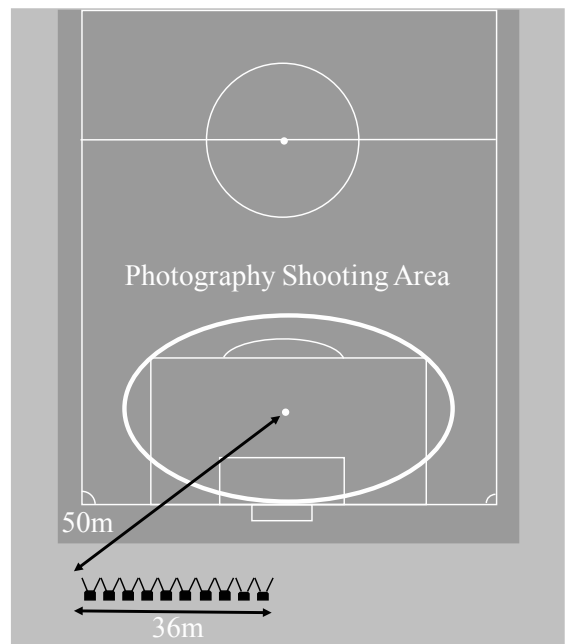
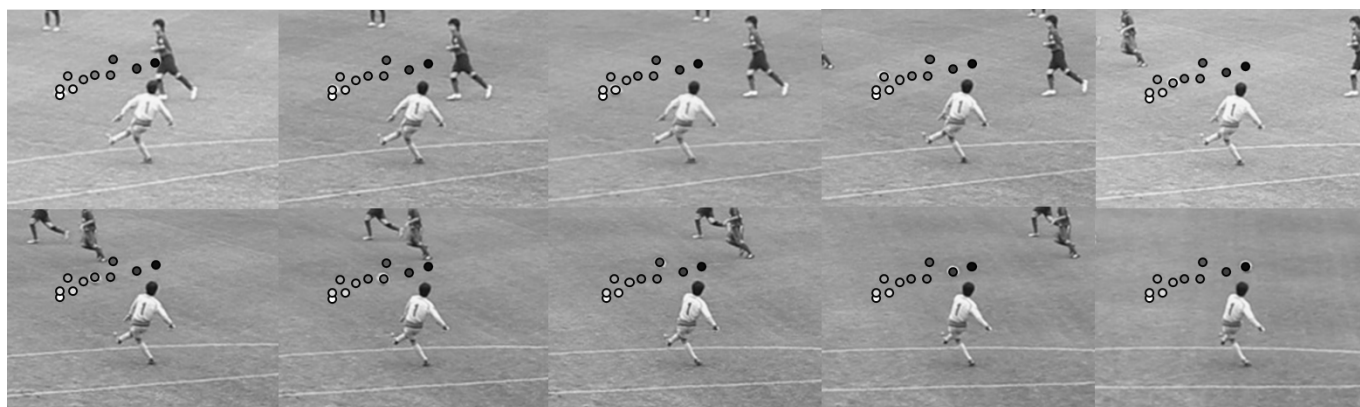


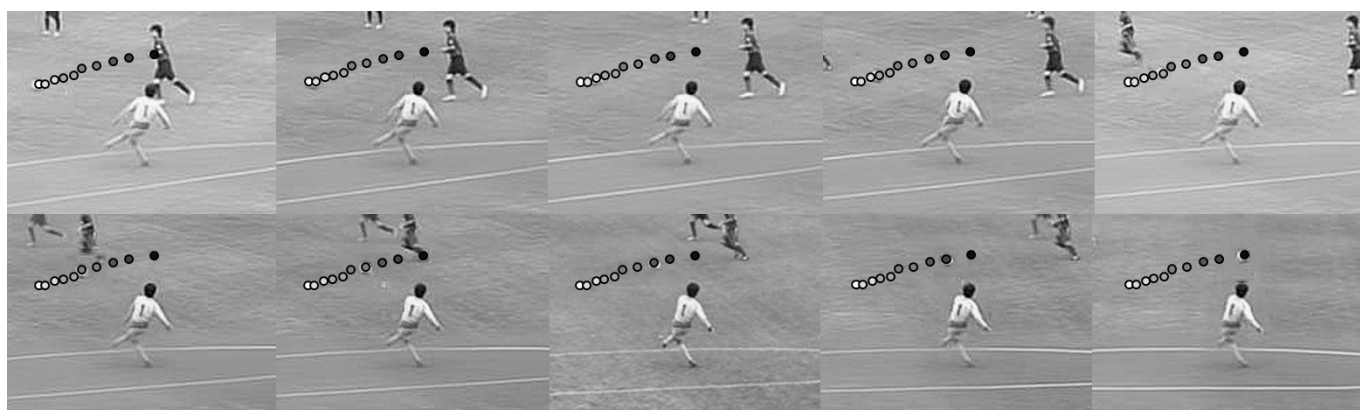
図 11 撮影環境の概略図

7. おわりに

本研究では非同期多視点映像から生成されたバレットタイム映像中で観察されるブレを、フレーム補間処理を用いて低減することにより、滑らかな視点切り換え映像の生成法を提案した。連続する前後フレーム間で高速で移動する被写体に対してフレーム補間処理を行う際、初期補間画像においてモーションブラー効果を用いて被写体の見え方再現の手がかりとなる情報を付与することにより、補間処理結果を改善した。視点移動時の被写体の観測位置の変化が最も小さくなるような補間画像を選択することにより、滑らかな視点切り換えのバレットタイム映像を生成することが可能となった。



(a)



(b)

図 12 従来手法(a)と提案手法(b)の比較

文 献

- [1] J. Carranza, C. Theobalt, M. A. Magnor and H.-P. Seidel, "Free-Viewpoint Video of Human Actors", *ACM Transactions on Graphics*, Vol.22, No.3, pp.569-577, 2003..
- [2] T. Kanade, P. Rander and P. J. Narayanan, "VirtualizedReality: Constructing Virtual Worlds from Real Scenes", *IEEE MultiMedia*, Vol.4, No.1, pp.34-47, 1997.
- [3] Y. Liu, Q. Dai and W. Xu, "A Point-Cloud-Based Multiview Stereo Algorithm for Free-Viewpoint Video", *IEEE Transactions on Visualization and Computer Graphics*, Vol.16, No.3, pp.407-418, 2010.
- [4] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, S. Sullivan, "High-Quality Streamable Free-Viewpoint Video" *ACM Transactions on Graphics*, 34(4), 2015.
- [5] T. Koyama, I. Kitahata and Y. Ohta, "Live Mixed- Reality 3D Video in Soccer Stadium", *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp.178-186, 2003.
- [6] N.Akechi, I.Kitahara, R.Sakamoto and Y.Ohta, "Multi-resolution bullet-time effect" *SIGGRAPH-ASIA*, 2014.
- [7] T. Kanede et al., "EyeVision", <http://www.ri.cmu.edu/events/sb35/tksuperbowl.html>, 2001.
- [8] K. Tomiyama and Y. Iwadate, "Development of Multi-View HDTV Image Generation System", *The Journal of the Institute of Image Information and Television Engineers*, Vol.64, No.4, pp.622- 628, 2010.
- [9] M. Tanimoto "Free-Viewpoint TV", *IEEE Signal Processing Magazine*, Vol.28, No.1, pp.67-76, 2011.
- [10] I. Kitahara, H. Saito, Y. Ohta, T. Ono, "Large-scale Virtualized Reality", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Technical Sketches, 2001.
- [11] T. shin, I. Kitahara, Y. Kamada, Y. Ohta, "Comparative Experiments Between Two 3D Modeling Methods -Human Billboard and Shape-From-Silhouettes-", *電子情報通信学会総合大会*, 2009年, pp.159, D-12-80
- [12] RedGiant, "Plural Eyes", <https://www.redgiant.com/products/pluraleyes/>.
- [13] T. Beier and S. Neely, "Feature-Based Image Metamorphosis", *ACM SIGGRAPH Computer Graphics*, Vol.26, No.2, pp.35-42, 1992.
- [14] S.Y.Lee and K.Y.Chwa, "Image Morphing Using Deformation Techniques", *The Journal of Viualizaaiion and Computer Animation*, Vol.7, pp.3-23, 1996.
- [15] Shechtman E, Rav-Acha A, Irani M, Seitz S, "Regenerative Morphing" *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San-Francisco CA, June 2010.
- [16] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *CVPR*, Anchorage, AK, USA, 2008.