Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

解像度を維持するBullet Timeの生成と評価

坂本 竜基[1,a)]　陳 鼎[1]

概要：被写体を中心に等間隔に並べた多視点カメラからの映像のうち、同時刻のフレームを順に切り替えるとBullet Timeと呼ばれるカメラワークをもった映像表現となる。この時、各カメラを被写体の一点が画像の中心となるように設置しなければ不自然な映像となってしまうため、各フレームを射影変換で補正する方法がよく用いられる。しかし、この射影変換は元のフレームを変形するため、そのままでは空白部分ができてしまう。これを回避するには変換後の画像を拡大すればよいが、過度な拡大をすると画像が劣化してしまう。そこで、本稿では、この拡大をなるべく抑えて元の画像の解像度を維持しつつ、Bullet Timeカメラワークとして自然な射影変換をおこなうアルゴリズムを提案する。また、それぞれの変換が既存手法に対してどの程度、解像度を維持できるのかを明らかにした上で、被験者実験によりカメラワークがユーザに与える影響を、主にテレプレゼンスと自然さの観点から検証した結果を報告する。

Generation Technique and Eveluation on High-Resolution Bullet-Time Camera Work

Ryuuki Sakamoto[1,a)]　Ding Chen[1]

Abstract: The multi-camera environment have been used in movie studio when they intent to apply the *"Bullet Time"* camera work to a scene. The camera work is realized with flipping through frames at same moment taken by multi cameras surrounding an object at even distances. For making outcome frames of the camera work, the Homography transformation is adapted for rectifying inaccurate camera poses. Therefore the Homography transformation, however, makes some blank spaces during distorting the frame, the scale up transformation should be applied after that. The scaling up, however, makes the quality of the outcame down. In this paper, we proposed a method to calculate Homography matrices for keeping the quality and naturality of outcome frames of the camera work. For measuaring the effectiveness of the method, we also describe the result of a user evaluation.

1. はじめに

時間を止めて被写体の周りを回っているかのような映像である、俗にBullet Timeと呼ばれるカメラワークは、映画をはじめとしてここ10年ほどで映像表現に広く用いられるようになった。Bullet Timeは、被写体の周囲に複数台のカメラを等間隔に並べ、ある時刻における各カメラのフレームを端に位置するカメラから順に切り替えることで得られる。ただし、例えば、光軸が交互に上下するといった、各カメラの映像を切り替えてみたときに、それらが滑らかに連続した変化として認知できないようにカメラが配置されている場合は、非常に違和感のある映像となってしまう。よって、各カメラの設置は厳密におこなう必要があるが、実際は、光軸が3次元空間中に位置する任意の1点を通過するように手動でカメラの向きを調整することは極めて難しい。そこで、ポストプロセスとして各カメラのフレームを、あたかもカメラを正しい向きに厳密に調整して撮影したかのように補正するのが一般的である[1]。

一方で、この補正は、あらかじめカメラを校正した上で適切な射影変換行列を掛けることで実現されるため、元の矩形のフレームは歪んだ四角形に変換されてしまう。そこで、その歪んだ四角形を矩形でクロッピングして、出力したい解像度まで拡大したものを出力とするが、これは結局、元のフレームの構成要素である全画素のサブセットである

[1] ヤフー株式会社
Yahoo Japan Corp. Mid9-7-1 Akasaka, Minato-ku
[a)] ryusakam@yahoo-corp.jp

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

ため実質的な解像度は低下してしまう。

しかし、過去の研究では、この実質的な解像度の低下を抑制する最適な変換行列に関する研究はなされていない。現在、8~ 12 台程度のカメラを同期させたうえ 30fps 以上で撮影する場合、解像度を XGA 以上にするとシステム全体が高価なものになる。つまり、システムのコストを考えればカメラの解像度は VGA 以下に抑える必要があるため、いまや FullHD が珍しくない TV や PC に配信することを鑑みれば可能な限り解像度は維持するほうが望ましいであろう。

そこで、本稿では、Bullet Time において実質的な解像度の低下を抑える射影変換法を提案する。また、このカメラワークを Web ブラウザ上のマウスドラッグ操作でおこなう実験システムを用意し、提案する射影変換の認知的影響の測定をおこなう。

さらに、そもそも Bullet Time をユーザがインタラクティブに操作可能なインタフェースを評価した研究はない。Bullet Time は視点の移動であるため移動時に運動視差が発生するが、運動視差は被写体のテレプレゼンスを増強させるとされている [2]。よって、このような操作をして Bullet Time をおこなうと被写体のテレプレゼンスは強化される可能性があるため、この観点から実施した被験者実験の結果も報告する。

2. Bullet Time と射影変換

地面を xyz 座標系における xz 平面（$y=0$）、天方向を y 軸とした空間に被写体が立っている状態で、N 台のカメラで Bullet Time を実現するカメラの配置を考える。最も単純に実現するには、被写体が内包する 3 次元点 $\mathbf{G}$（以下、注視点と呼ぶ）から等距離かつカメラ間の距離を等間隔になるよう各カメラを並べ、各カメラを $\mathbf{G}$ に向ければよい。つまり、$\mathbf{G}$ を含む xz 平面に平行な平面上に $\mathbf{G}$ を中心とした円を考え、各カメラをピンホールカメラモデルとして捉え、k 番目のカメラの焦点位置を $\mathbf{C}_k$（$k=1,...,N$）とした時、$\mathbf{C}_k$ を円周上に等間隔で設置したうえで、$\mathbf{C}_k$ から画像平面の中央を通る直線（以下、光軸とよぶ）が $\mathbf{G}$ を通過するようカメラの向きを変えればよい。

しかし実際は、三脚などを使って設置をする場合、上記の手順に厳密に従うことは不可能である。その主な理由は以下の 3 点である。

イ　実空間における $\mathbf{C}_k$ が判らない
ロ　円周やカメラ間の間隔を厳密に測定できない
ハ　三脚の手動操作では $\mathbf{G}$ を光軸上に設定できない

このうち、イとロに関しては、被写体から比較的遠い位置から撮影するのであれば、カメラの大よその中心と巻尺程度の精度の計測で設置しても問題がない。しかし、ハに関しては、事前にモニタ上に出力した格子線を頼りに時間をかけて調整しても実際は十数ピクセルは誤差が生じてしまう。

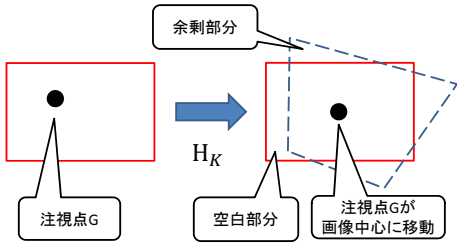

図 1　式 3 による変換
Fig. 1　Conversion by eq.3

しかし、もしカメラが強校正済みであるならば、本来撮影されたフレームを任意の 3 次元点を通過するよう光軸を向けて設置したかのような画像に射影変換する、いわばカメラに仮想的なパン・チルトをさせることができる [3], [4]。よって、厳密に光軸を $\mathbf{G}$ に向ける射影変換を各カメラの出力フレームに適用すればハの問題も解決する。

この変換をおこなう射影変換行列 $\mathbf{H}_k$（$k=1,...,N$）は以下のように求める。まず、各カメラは強校正されているので、k 番目のカメラの内部パラメータ行列 $\mathbf{A}_k$、外部パラメータ $\mathbf{R}_k$、$\mathbf{T}_k$ がわかっている。このうち、$\mathbf{A}_k$ は以下の要素で構成されているとする。

$$\mathbf{A_k} = \begin{bmatrix} f_k & 0 & u0_k \\ 0 & f_k & v0_k \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

ただし、f_k は焦点距離、$u0_k$、$v0_k$ は画像中心とする。ここで $\mathbf{C}_k = -\mathbf{R}_k^{-1}\mathbf{T}_k$ から $\mathbf{G}$ に向く新しい光軸にあたるベクトル $\mathbf{e_z}$ を求める。$\mathbf{e_z}$ と y 軸との外積から新しいカメラ座標系での x 軸 $\mathbf{e_x}$ が求められ、$\mathbf{e_z} \times \mathbf{e_x}$ から新しいカメラ座標系での z 軸 $\mathbf{e_z}$ を求める。これらから新しい回転行列 $\mathbf{R'}_k$ が以下のように求まるため、

$$\mathbf{R'}_k = \begin{bmatrix} \mathbf{e_x}/|\mathbf{e_x}| \\ \mathbf{e_y}/|\mathbf{e_y}| \\ \mathbf{e_z}/|\mathbf{e_z}| \end{bmatrix} \tag{2}$$

これらから射影変換行列 $\mathbf{H}_k$ が以下のように求まる。

$$\mathbf{H}_k = \mathbf{A}_k\mathbf{R'}_k\mathbf{R}_k^{-1}\mathbf{A}_k^{-1} \tag{3}$$

あとは、これをフレームに適用して変形させれば画像中心に $\mathbf{G}$ が正確に位置する仮想的なパン・チルトした画像が得られる。

しかし、**図 1** にあるとおり変換結果は空白部分を含むため、このままでは出力フレームとすることができない。出力フレームから空白部分を排除する単純な方法は、画像が元のフレームサイズを覆うまで拡大することである。しかしながら、拡大をすれば空白はなくなるが余剰部分が増え

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

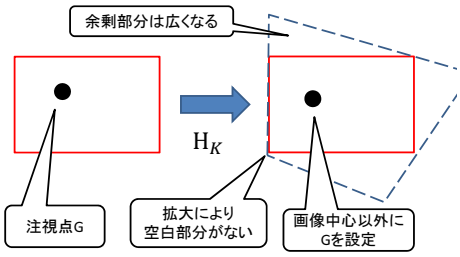

図 2　最適化
Fig. 2　Conversion by eq.3

る、すなわち元の画像にあった余剰部分の画素を捨てることになるため、実質的な解像度は低下してしまう。

この画質の劣化をなるべく抑えるには、元々 $\mathbf{G}$ が映っていた画像上の位置 $\mathbf{g}_k$ に変換後の $\mathbf{G}$ の 2 次元位置が近づくよう画像を平行移動させたうえ、拡大率を余白部分が丁度なくなるようにすればよい（**図 2**）。しかし、各カメラでこれらの最適化を個別におこなうと、隣のカメラへと画像を切り替えた時に認知的な連続性がなく大変違和感があるカメラワークとなってしまう。本稿では、この違和感をなるべく抑えた上で、実質的な解像度の低下を減らす $\mathbf{H}_k$ を提案する。

3. 拡大と平行移動を含む射影変換

前章で、解像度の低下を抑えるには、拡大と平行移動の二つの手段があることを述べた。アフィン変換における拡大のパラメータは式（1）における f_k、すなわち焦点距離である。内部パラメータにおいて焦点距離が増加するということは、画像面を $\mathbf{C}_k$ から離して $\mathbf{G}$ に近づける、すなわちズームしていることに他ならない。実際には f_k は $|\mathbf{C}_k - \mathbf{G}|$ という距離のみで決定され、$\mathbf{G}$ をその距離に応じた大きさにするのだが、$\mathbf{G}$ は固定であるため裏返せば $\mathbf{C}_k$ を $\mathbf{G}$ に近づけることに相当する。これは、カメラの物理位置を仮想的に前後させることになるため、先行研究では、イ、ロに起因する $\mathbf{G}$ と $\mathbf{C}_k$ 間の距離が各カメラで一致しない誤差を補正するのに用いている [1]。

一方、平行移動は、$(u0_k, v0_k)$ を変更することで達成され、これは $\mathbf{C}_k$ と $\mathbf{R'}_k$ はそのままに画像面を平行移動させることに相当する。すなわち、光軸が画像中心を通過しないカメラとなり、通常にはないカメラの構造になるが、実質的には光軸が画像中心を通った画像と大差がないため認知的な問題は少ない。

これらから、新しい焦点距離 f'_k と新しい画像中心 $(u0'_k, v0'_k)$ が設定された内部パラメータ $\mathbf{A}'_k$ を用いた新しい $\mathbf{H}'_k$ が導出される。

$$\mathbf{A}'_k = \begin{bmatrix} f'_k & 0 & u0'_k \\ 0 & f'_k & v0'_k \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

$$\mathbf{H}'_k = \mathbf{A'}_k \mathbf{R'}_k \mathbf{R}_k^{-1} \mathbf{A}_k^{-1} \tag{5}$$

4. 実質的な解像度の定義

ここで、元のフレームを構成していた全画素が変換後の画像の中にどの程度含まれるのか、すなわち実質的な解像度を計測する指標を作る。画像の四隅の 2 次元座標 $\mathbf{v}_{\alpha k} (\alpha = 0, 1, 2, 3)$ が $\mathbf{H}'_k$ によって $\mathbf{v}'_{\beta k} (\beta = 0, 1, 2, 3)$ へ移動するとする。

$$\lambda \begin{bmatrix} \mathbf{v}'_{\beta k} \\ 1 \end{bmatrix} = \mathbf{H}'_k \begin{bmatrix} \mathbf{v}_{\alpha k} \\ 1 \end{bmatrix} \tag{6}$$

最終的な出力フレームから空白をなくすには、$\mathbf{v}_{\alpha k}$ すべてが、$\mathbf{v}'_{\beta k}$ で形成される四角形の内部に必ず存在する必要があるため、元の画像の全画素のうち、出力フレームに含まれる画素数の割合は $\mathbf{v}'_{\beta k}$ で形成される四角形の面積 S_k と元々の長方形の面積 $S0$ の比 $S0/S_k$ で定義できる。S_k は f'_k、$(u0'_k, v0'_k)$ に依存し、$S0$ は固定値であるため、ここでの問題は

$$S_k = \frac{1}{8} \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} (\mathbf{v}'_{(\beta \bmod 4)k} - \mathbf{v}_{\alpha k}) \times (\mathbf{v}'_{(\beta+1 \bmod 4)k} - \mathbf{v}_{\alpha k}) \tag{7}$$

という制約条件の下での面積比 E の最大化と定義できる。

$$argmax\ E(f'_k, u0'_k, v0'_k) = \sum_{k=1}^{N} \frac{S0}{S_k} \tag{8}$$

5. 滑らかなカメラワークを実現するアルゴリズム

E を維持しつつ f'_k、$(u0'_k, v0'_k)$ をどう設定すればスムースな Bullet Time になるかを検討した結果、以下の 4 つの戦略を考えた。

戦略 1: $(u0'_k, v0'_k)$ **の固定化**

$(u0'_k, v0'_k)$ とは、実質的に画像における $\mathbf{G}$ の位置を制御する変数であり、これらを各カメラで共通の値にすれば注視点は見た目上、不動になり違和感がない。この共通の値は、元の画像における $\mathbf{G}$ の位置 $\mathbf{g}_k$ に近いほど E は高くなるため、$\sum_{k=1}^{N} \mathbf{g}_k / N$ と設定できる。

戦略 2: $(u0'_k, v0'_k)$ **の変動化**

戦略 1 は、$(u0'_k, v0'_k)$ を各カメラ共通とする上では適切であるが、各カメラにおいて $\mathbf{g}_k$ が滑らかに移動しても経験上、違和感はそれほどない。そこで、k が 1 から N まで変化するときの $\mathbf{g}_k$ の軌跡を何らかの線形モデルに回帰させる。直感的に考えると、このモデルは単純であるほど違和感が少ないため実験では直線と

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

した。これは $\mathbf{g}_k$ の分布によっては残差が少なくなり戦略 1 よりも E を増大させる。

戦略 3: フォーカスの固定化

f'_k を増加させれば、擬似的にカメラを近づけたかのような見えになる。よって、f'_k を以下のように設定すれば注視点に居る被写体の見た目の大きさを一定に保つことができる。

$$f'_k = f_{average}\frac{|\mathbf{C}_k - \mathbf{G}|}{z_{average}} \tag{9}$$

$$f_{average} = \sum_{k=1}^{N}\frac{f_k}{N}$$

$$z_{average} = \sum_{k=1}^{N}\frac{|\mathbf{C}_k - \mathbf{G}|}{N}$$

被写体の大きさがカメラ間で不規則に大小すると違和感があるため、このような処理は自然な Bullet Time を演出する上で効果的である。

戦略 4: フォーカスの変動化

滑らかに $\mathbf{g}_k$ を変化させた戦略 2 と同じく、被写体の見た目の大きさも滑らかに変化させても違和感は少ない。これは、カメラが被写体を取り囲んでいるような配置を天から見下ろした場合、$\mathbf{C}_k$ を $\mathbf{G}$ と $\mathbf{C}_k$ を結ぶ直線上で前後に動かし、$\mathbf{C}_k$ が全体として何らかの滑らかな軌跡上に存在しているかのように f_k を設定すると達成され、その解の一つは、以下の α を何らかの線形モデルへの回帰させ f'_k を得ることで求まる。

$$\alpha = \frac{\mathbf{f_k}}{\mathbf{C}_k - \mathbf{G}} \tag{10}$$
$$f'_k = \alpha(\mathbf{C}_k - \mathbf{G})$$

実験では、天方向から見下ろした 2 次元空間において α を 2 次曲線で回帰させた。この回帰の結果得られる f'_k の残差が戦略 3 で決定した f'_k と f_k の距離の和よりも低い場合は E は戦略 3 より高くなる。

これらの方法を適用した全体の処理は以下のようになる。

(1) ユーザが $\mathbf{G}$ を入力する
(2) 方法 1 か 2 を適用して $(u0'_k, v0'_k)$ を決定する
(3) 方法 3 か 4 を適用して f'_k を決定する
(4) 式 (7) を満たす最小値まで全 f'_k を一定の割合まで増加させる
(5) 各フレームに $\mathbf{H}'_k$ を適用する

このアルゴリズムでは戦略 1、2 および戦略 3、4 はお互いに排他的であるため、実質的に 4 種類の変換行列が出力可能である。以下では、この**表 1** のように 4 種類の戦略の組み合わせをそれぞれコンビネーション A、B、C、D と呼ぶ。

表 1 戦略の組み合わせ
Table 1 Combination of each technique

	戦略 3	戦略 4
戦略 1	コンビネーション A	コンビネーション B
戦略 2	コンビネーション C	コンビネーション D

6. 関連研究とマイルストーン

映画ではなく現実に起こった出来事を多視点映像を処理することで効果的に表現するシステムは、古くから応用研究がなされ専らスポーツをその撮影対象に発達してきている [5], [6], [7], [8], [9], [10], [11]。このうち、多視点映像を自由視点映像化する場合は Bullet Time が可能であり、本稿のようにカメラを連続的に切り替える方法は 3 次元モデルを復元する必要がないため最も頑健な手法の 1 つである。

この手法を用いたシステムとして最も著名なのは Eye Vision[12] であろう。Eye Vision は、各カメラの光軸が写したい $\mathbf{G}$ で交差するように、各カメラをロボット雲台で制御することで、アメリカンフットボールの試合中継において Bullet Time を実現した。これに対し、冨山らは、高価で調整に時間がかかるロボット雲台を用いるのではなく、おおよその被写体の位置を向くようカメラを三脚で固定し、位置合わせをポストプロセスとして画像処理で補正することで仮想的に実現した例を紹介している [1]。この冨山らのシステムは Eye Vision と違い、被写体が大きく動くシーンには適用できないものの、ロボット雲台を導入するよりも圧倒的に設置が容易であるため全日本体操選手権において TV 放送用途での運用が可能であったと報告されている。本稿で提案した手法は、このようなシステムに適用することを想定している。なお、冨山らのシステムでは本稿での戦略 1 で $(u0'_k, v0'_k)$ を画像中心に固定したうえで、戦略 3 を適用するコンビネーション A の特殊ケースとなる射影変換が採用されている。以下の結果や実験では、この冨山らの変換をマイルストーンとして提案アルゴリズムによる E の改善を述べたい。

7. 変換結果

8 台のカメラにより撮影したデータセットに対して、提案アルゴリズムを適用した結果を示す（**図 3**）。図の 1 段目が撮影した元画像であり、画像におけるぬいぐるみの位置や大きさが微妙に異なることがわかる。これに対して、戦略 1、3 の組み合わせであるコンビネーション A で変換した 2 段目を見ると位置も大きさも揃い、Bullet Time の際に自然な切り替えになることが想像できる。3 段目のコンビネーション B は、戦略 3 の代わりに f'_k を回帰により最適化した戦略 4 を用いており、2 段目に比べて大きさが徐々に変化して、全体としてはぬいぐるみが小さく写っている。被写体が小さくみえるということは、実質的な解像

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

図 3 各シーンの Bullet Time
Fig. 3 Bullet Time camera work on each scene

表 2 各方法によって変換したときの E
Table 2 E values of each technique.

コンビネーション A	コンビネーション B	コンビネーション C	コンビネーション D	冨山の方法
0.33	0.60	0.72	0.78	0.26

度が高いということになり、**表 2** によると E は 0.33 から 0.60 へと改善している。**図 4** は元々の α とコンビネーション A、B の α をプロットした結果であり、コンビネーション B のほうが元々の α に近く滑らかに回帰されていることがわかる。

一方、コンビネーション C は、コンビネーション A の戦略 1 を戦略 2 に変更し、$(u0'_k, v0'_k)$ を回帰により最適化したものである。図の 4 段目を見ると、この効果により、2 段目と比べて大きさは一定であるが、全体的に小さくなっていることが判り、E は 0.72 へと改善している。**図 5** は、各カメラの $(u0_k, v0_k)$ とコンビネーション A、C による $(u0'_k, v0'_k)$ の位置をプロットしたものであり、コンビネーション C は直線上に回帰していることが判る。最後に 5 段目は f'_k および $(u0'_k, v0'_k)$ 両方を回帰させたコンビネーション D であり、E は全体で最高の 0.78 となっている。これらは何れも既存手法である冨山の方法よりも大きく改善しており、提案手法は元画像が持つ解像度を損なうことなく変換できているといえる。

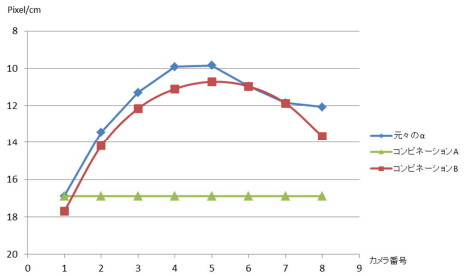

図 4 コンビネーション A、B における α
Fig. 4 α on combination A and B

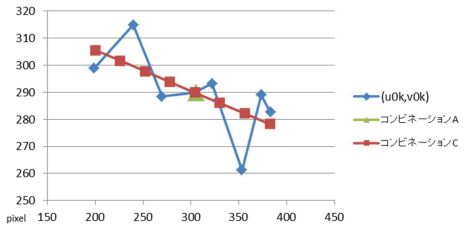

図 5 コンビネーション A、C における $(u0'_k, v0'_k)$
Fig. 5 $(u0'_k, v0'_k)$ on combination A and C

8. 評価実験

提案手法は先行研究よりも E を高く維持できるものの、f'_k、$(u0'_k, v0'_k)$ の変化がユーザに不自然な印象を与えている可能性は否定できない。また、そもそも多視点カメラを用いて Bullet Time をユーザに提供する利点は明らかにされていない。そこで、Web ブラウザ上におけるマウスのドラッグ操作に応じて、その時刻における他のカメラの画像に順に切り替わる実験用インタフェースを用意し、これに対して「Bullet Time の有用性」および「提案手法での変換の自然さ」について被験者実験をおこなった。

8.1 実験環境

実験に用いる映像を Point Grey Research 社の Flea2 を 8~ 12 台用いて 4 種類の多視点映像を撮影した。カメラは、被写体を 90~ 150 度程度の角度で一定の距離で取り囲むよう三脚を用いて手作業で設置した。

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

それぞれの映像は 30~ 90 フレーム程度であり、撮影と同時におこなったキャリブレーションのファイルと共に全フレームを画像として保存した。このデータセットに対して提案した射影変換をおこなった結果の画像群を別途それぞれシーン A~ D と名前をつけて http サーバに保存した。

一方、これらのデータを閲覧するビューワとして、読み込んだあるシーンのあるカメラのフレームを時系列に沿ってパラパラ漫画のように切り替える Web ブラウザで動作する HTML5 と JavaScript ベースのコードを書いた。これに、ユーザが画像上でマウスドラッグをすると、同時刻に写した隣のカメラのフレームに表示が切り替わる JavaScript ベースのインタフェースを組み込み、画像の幅 80 %程度を端から端までドラッグすると、その時刻における全カメラの Bullet Time を閲覧できるようにした。

図 6 は、この 4 つのシーンをコンビネーション A により変換した結果のデータを、ある時刻において Bullet Time をした時における一部のカメラの画像である。各シーン上段から下段にかけて視点が徐々に切り替わっていることがわかる。

実験では、このインタフェースを組み込んだ 5 種類の Web ページの下段に、それぞれラジオボタンを付した質問項目を設置し、被験者には Bullet Time を体験してもらった上で質問に最も適した回答項目を選択してもらった。**図 7** がそのうちの 1 ページのスクリーンショットである。画面の上段には、「左の映像」「右の映像」とラベルが明記された二種類の映像が表示され、例えば、「左の映像が右の映像よりも○○ですか？」という質問項目があった場合でも、両者を同時に目視したうえで回答できるデザインとした。

被験者は社内イントラネットによる掲示で募り、そのまま自席で参加する形式をとったので、被験者が本当に Bullet Time をおこなってから回答したかを確認できるように、画像の切り替えを記録するログも同時に取得した。次節以降に述べる結果は、このログから実際に端のカメラから端のカメラまで Bullet Time をおこなわなかった被験者のデータを除いたものである。

被験者は実験に際し、ポータルページで実験内容の説明とインタフェースに対するインストラクションを受け、実際にサンプル映像を動かして十分に操作に慣れた上で実験に臨んだ。

8.2 実験 1：Bullet Time の有用性

8.2.1 質問項目

まず、1 台のカメラによる映像に比べて Bullet Time がテレプレゼンスを増強するという仮説を検証したい。このため、被験者が回答した 5 種類の Web ページのうち、前半の 4 ページには、図 6 の 4 シーンをコンビネーション A で変換したデータセットをそれぞれ読み込むようにし、「右の映像」は Bullet Time がマウスドラッグで可能に、「左の映像」は一番左端のカメラの映像だけを繰り返し流すようにした。このことを明示するため右の映像の下には「通常の映像（ドラッグできません）」右の映像の下には「ドラッグするとカメラを切り替えられる映像」という注意書きを設けた。

図 7 実験用 Web ページ
Fig. 7 Screenshot of an evaluation page

表 3 存在感に関する質問結果
Table 3 Result of telepresence.

	シーン A	シーン B	シーン C	シーン D
n	173	142	113	142
平均	4.82	5.23	5.03	5.30
分散	2.04	1.30	1.89	1.35
T 値	7.55	12.93	8.00	11.52
全体の平均	5.08			
全体の分散	1.80			
全体の T 値	19.34			

この上で、「まず、右の映像の上でマウスドラッグができることを確認してください（左の映像はドラッグしても何も起こりません)。その後、下記のの質問に対し最も当てはまる項目を選択してください。」というインストラクションをして、下記を項目を記載した。

質問 右の映像において映画マトリックスのようなカメラワークを体験していると、左の通常の映像に比べて被写体の存在感をより強く感じる

この質問への回答として 7 段階のチェックボックスを置き、その下部に左から「まったくあてはまらない、あてはまらない、どちらかといえばあてはまらない、どちらともいえない、どちらかといえばあてはまる、あてはまる、非常にあてはまる」というスケールを記述した。

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

図 6 各シーンの Bullet Time
Fig. 6 Bullet Time camera work on each scene

8.2.2 結果

それぞれのスケールに 1 から 7 までの得点を設定したときの集計結果を**表 3** に示す。左の通常の映像に対して、右の Bullet Time 付き映像のほうが存在感があるかどうか調べるには、両者がまったく同じ存在感である場合「どちらともいえない」と選択されるはずなので、各データの平均と「どちらともいえない」の値である「4」との間で t 検定（片側検定）をおこなった。この結果は、すべてのシーンで 1%有意であった。また、個々のシーンの得点に関して差があるか一元配置分散分析をおこなったところ、有意差が認められた（$F(3,566)=4.23$, $p<.01$）。その後、Tukey の方法で多重比較したところ、シーン A、B 間とシーン A、D 間に有意差 5%が認められた。よって、シーン A から D はテレプレゼンスにとって影響がでるほど多様なシーン群であったにも拘らず、すべてのテレプレゼンスを増強させることがわかった。念のため、シーン A から D 全体の平均との比較をおこなっても 1%有意であったことから、Bullet Time は、テレプレゼンスを増強させるといえる。これは、先行研究にあるとおり Bullet Time が与える運動視差が影響を及ぼしたと考えられる。

8.3 実験 2：4 手法の比較

8.3.1 質問項目

5 種類の Web ページの残り 1 つには、シーン D に対し

Vol.2014-GN-90 No.6
Vol.2014-CDS-9 No.6
Vol.2014-DCC-6 No.6
2014/1/23

て 4 手法のうち何れかの手法を適用して変換したデータを読み込むように設定した。どの手法が適用されるかはランダムであるが、ページ下部には共通で以下のアンケート項目を表示した。この項目は、4 手法それぞれが認知的に適切な変換となっていることを確認することが目的である。上述した実験 1 の結果から、コンビネーション A による変換は自然な Bullet Time となるといってよいであろう。よって、すべての戦略が自然な変換となっているのであれば、コンビネーション B、C、D がコンビネーション A に比べて差がないはずである。なお、1 人に 1 種類の手法のみを閲覧させる理由は、その被験者個人内での学習効果をなくすためである。

この実験用ページには、「左の映像」に変換前のデータセットを、「右の映像」に 4 種類のコンビネーションの何れかの変換結果のデータセットを読み込ませ、左右どちらの画像上でドラッグしても、両方の画像が同期して視点が切り替わるように設定した。その上で実験 1 と同じく「まず、マウスドラッグで左右の映像が連動して動くことを確認してください。その後、下記の質問に対し最もあてはまる項目を選択してください」とインストラクションし、以下の 2 種類の質問をした。なお、質問の回答となるチェックボックスの下には、両者とも実験 1 における質問と同じスケールを記述した。

質問 1 映画マトリックスのようなカメラワークを体験していると、右の映像は左の通常の映像に比べて被写体の存在感をより強く感じる

質問 2 左の映像に比べて右の映像は、マウスドラッグした場合のカメラワークが自然である

8.3.2 結果

まず、シーン D に対して冨山の方法を適用したときの E は 0.45 であり、**表 4** に示した通り、各手法の E はそのよりも何れも高いことから、実質の解像度は既存手法に比べて大きく改善されている。この状態で質問 1 の得点を各手法に関して一元配置分散分析をおこなった結果、有意差は認められなかった (F(3,135)=0.48, n.s.)。また、質問 2 に対しても一元配置分散分析をおこない有意差が認められなかった (F(3,135)=1.58, n.s.)。有意差がないことは直ちに差がないことにはならないが、少なくとも本サンプル数においては各コンビネーションは近しい印象を与えたことになる。これは、すべてのコンビネーションによる変換によって自然な Bullet Time として認知されるという仮説を支持している。

表 4 各コンビネーションによって変換した結果
Table 4 Results of each technique.

コンビネーション名		A	B	C	D
E		0.74	0.77	0.82	0.75
質問 1	平均	4.12	4.17	3.79	4.13
	分散	2.05	2.03	2.59	2.72
質問 2	平均	4.14	4.31	3.72	3.71
	分散	1.64	2.57	1.84	2.21

9. 結論

本稿ではカメラ映像の切り替えによって Bullet Time を実現するシステムにおける射影変換による補正を解像度維持の面から最適化する 4 種類の戦略を提案した。そのうちの 1 つを用いて変換した画像群をマウス操作により切り替えることにより Bullet Time を体験可能なインタフェースを用いた被験者実験により、そもそも Bullet Time がテレプレゼンスを強化することを確認した。また、提案した戦略のすべての可能な組み合わせで変換したデータ同士を比較し、何れの組み合わせも自然さとテレプレゼンスの観点では認知的に差がないことも確認した。本研究では、各戦略を段階的に適用する近似的な解法をとったが、今後はこの結果を初期値とした非線形最適化に取り組みたい。

参考文献

[1] 冨山仁博, 宮川勲, 岩舘祐一 : 多視点ハイビジョン映像生成システムの試作: 全日本体操選手権での中継番組利用, 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. 106, No. 429, pp. 43–48 (2006).

[2] Nakanishi, H., Murakami, Y. and Kato, K.: Movable cameras enhance social telepresence in media spaces, *Proceedings of the SIGCHI*, ACM, pp. 433–442 (2009).

[3] Hartley, R. and Zisserman, A.: *Multiple view geometry in computer vision*, Cambridge Univ Press (2000).

[4] Szeliski, R.: *Computer vision: algorithms and applications*, Springer (2011).

[5] Kanade, T., Rander, P. and Narayanan, P.: Virtualized reality: Constructing virtual worlds from real scenes, *MultiMedia*, Vol. 4, No. 1, pp. 34–47 (1997).

[6] Kitahara, I. and Ohta, Y.: Scalable 3D representation for 3D video in a large-scale space, *Presence: Teleoperators and Virtual Environments*, Vol. 13, No. 2, pp. 164–177 (2004).

[7] Inamoto, N. and Saito, H.: Intermediate view generation of soccer scene from multiple videos, *Proc. on 16th International Conference on Pattern Recognition*, Vol. 2, IEEE, pp. 713–716 (2002).

[8] Hilton, A., Guillemaut, J., Kilner, J., Grau, O. and Thomas, G.: 3d-tv production from conventional cameras for sports broadcast, *Broadcasting, IEEE Transactions on*, Vol. 57, No. 2, pp. 462–476 (2011).

[9] Hashimoto, T., Uematsu, Y. and Saito, H.: Generation of see-through baseball movie from multi-camera views, *IEEE International Workshop on Multimedia Signal Processing*, IEEE, pp. 432–437 (2010).

[10] Kimura, K. and Saito, H.: Video synthesis at tennis player viewpoint from multiple view videos, *Proceedings. VR 2005.*, IEEE, pp. 281–282 (2005).

[11] Tomiyama, K., Miyagawa, I. and Iwadate, Y.: Prototyping of HD Multi-Viewpoint Image Generating System-Live broadcasting use at gymnastics competition (60'th'National Championships)-, *IEIC Technical Report*, Vol. 106, No. 429, pp. 43–48 (2006).

[12] Kanade, T. et al.: EyeVision, *Web, http://www. pvi-inc. com/eyevision.*