

多視点映像の切替え提示による注目対象誘導のための 視点間画像補間による滑らかな視点切替えの一手法

東海 彰吾 樋口 翔太 長谷 博行

福井大学 大学院 工学研究科

あらまし 多視点映像の切替え提示において、被写体の画面構成を制御し、連続性を持たせた画像に加工した上で切替え提示することによって映像観察者の注目を特定の対象に誘導する機能を映像に付加できる。しかし、撮影視点を切替える方法では、使用できるカメラの台数が十分でない場合、切替え前後の視点位置変化の大きさが観察時の違和感につながっていた。そこで本発表では、限られたカメラ台数の撮影映像から滑らかな視点切替え映像を生成するために、平行化に基づく撮影視点間の画像補間を用いた一手法を提案し、その注目対象誘導機能に対する効果を検証した。

キーワード 多視点映像, 切替え提示, 注目対象誘導, 中割り画像生成, モーフィング

A Method of Smooth View Transition with Image Interpolation for Attention Navigation by Switching Multiple Views

Shogo Tokai, Shouta Higuchi, Hiroyuki Hase

Graduate school of Engineering, University of Fukui

Abstract In this paper, we propose a method to realize smooth switching of multiple views with image interpolation. Switching multiple views is one of the most effective image media to show scene situations, and we can control attention of viewer navigated to an object with keeping image composition in switching. But, smoothness of switching depends on the number of cameras to shoot the scene. We developed a method to interpolate image for switching a view to the next one based on image parallelization. We also show experimental results to validate effectiveness of our method.

Keywords: Multiple view switching, Attention navigation, Image interpolation, Morphing

1 はじめに

シーン状況を複数の視点から同時撮影した多視点映像は、単一カメラでは観測できない多角的状況を、様々な視点や視線方向で観測したものである。多視点映像やそれを構成する静止画像群を統合・解析による、3次元形状や動きの計測や、シーン状況の理解など、多くの研究が報告されている。また、一旦フレーム群に分解し、カメラや撮影時刻を制御するように並べ直した映像は、例えば、静止した時間の

中で視点移動するような映像が得られ、映画などの映像作品でも使われている [1, 2]。

我々のグループでは、多視点映像の切替えによってシーン状況を観察者に提示する際に、その見せ方、特に被写体の画面構成を制御することで「何を見せたいか」を操作する方法について研究を進めている。しかし、撮影視点の数が十分多くない場合、切替えによって提示される映像内での仮想的な視点移動が離散的となり、連続性を持つ滑らかな視点切替えが行えない。結果として、注目対象の部分の画面構成の

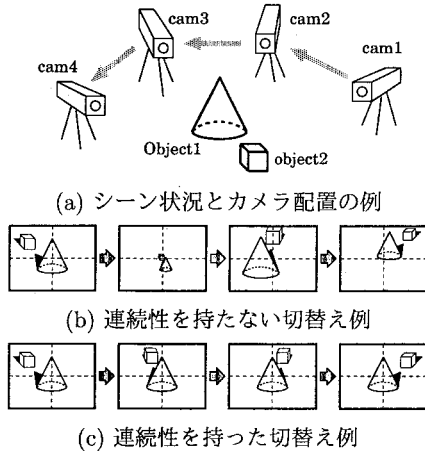


図 1: 多視点映像の切替えによる視点変化

連続性を確保したとしても、それ以外の対象物や背景については、その画面内での位置や大きさが極端に変わってしまい、映像観察者の違和感が増大する。

そこで本研究では、撮影視点間の中割り画像を合成して、切替え時に挿入し、離散的切替えの影響を軽減し、滑らかな切替え映像を得る一方法を提案する。以下、2では、切替え提示による注目対象誘導について概説し、3では、視点間の中割り画像を補間する方法について述べる。4では行った実験について述べ、5で本報告をまとめる。

2 注目対象誘導

複数のカメラの設置位置を離散的な視点位置とみなせば、その切替え提示は視点移動に対応する。実世界の人間の視点移動では、視野像は連続的で、さらに、何かを注視しながら移動することが多い。このとき、視野像では、注視対象物は静止しているように、背景が動いているように知覚される [3]。

しかし、切替え提示映像は、カメラの位置や姿勢に依存した見え方の映像であり、観察者にとっては注視箇所が不連続に感じられシーン状況が分かりにくくなる。例えば、図 1(a) の状況を切替え提示する場合、対象の位置や大きさが不連続な切替え提示では注目対象が定めにくい (図 1(b))。一方、一連の切替え提示映像において、実際に近い映像の連続性を持たせることができれば (図 1(c))、観察者は画面上の位置や大きさが連続的な対象を注視しながら

視点移動しているように感じられ、シーン状況や視点変化を無理なく知覚できることが期待できる。さらに、映像の製作者側が注目させたい部分の明確化などにも応用可能と考えられる。

そこで、我々は、被写体の画面内での位置や大きさ (以下、画面構成と呼ぶ) が映像の切替えの前後で連続的になるよう加工する、注目対象誘導を提案している [7]。撮影された画像に対して、撮影時と異なる視線方向や焦点距離の画像を生成する手法 [6] を利用し、それぞれの撮影像から抽出した対象領域の面積と重心を用いて、切替え提示用のパラメータを算出し、画像を加工した後に切替え提示を行うものである。また、それぞれのカメラからの対象領域抽出結果のばらつきの影響を抑えるため、カメラ群の校正情報を利用する方法を提案している [8]。

これにより、注目すべき被写体が複数存在する場合でも、一部の被写体の画面構成に連続性を持たせることができ、結果としてその被写体に注目を誘導することができる。しかし、撮影像を加工はするもののそれぞれのカメラの位置から撮影された像だけを提示するので、カメラの台数、すなわち、撮影視点の数が十分でない場合、その切替え映像は視点移動を表現するには不十分な数の画像しか生成できず、提示映像の不連続性が目立つ。そこで本研究では、撮影視点間の中割り画像を生成することで、撮影視点の不足を補い、滑らかな視点移動により近い、切替え提示の方法について考える。

2.1 画面構成制御

ここで、文献 [8] で述べている画面構成制御について簡単に説明する。画面構成制御の基本的考え方は、撮影時と同一の視点から異なる視野 (視線方向、焦点距離など) で撮影したかのような画像として、対象物に対して、設定した画面構成で捉えるように視野を制御した合成画像を、撮影画像からの投影によって生成するものである (図 2)。

あるカメラ C の画像 I について、 I 上の点 $m = (u, v, 1)^T$ とグローバル座標系の点 $M = (X, Y, Z, 1)^T$ の関係は、

$$wm = A \begin{bmatrix} R \\ -RT \end{bmatrix} M \quad (1)$$

と表される [4]。ただし、 A は透視投影変換を表す 3×3 の行列、 R, T はカメラの位置、姿勢を表す回転行列、および平行移動ベクトル、また、 w は 0 で

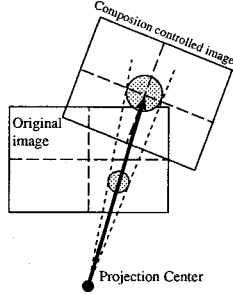


図 2: 画面構成制御の考え方

ない定数とする。なお、簡単化のために、以下の議論では、 A はそのカメラの焦点距離 f だけを使った

$$A_i = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

となるように u, v の座標系を定めることにする。

このとき、同一視点で視野の異なるカメラ C' では、 T は変わらないが、 A, R が A', R' に変わり、その画像 I' 上の M の投影点 $m' = (u', v', 1)^T$ は、

$$w'm' = A' \begin{bmatrix} R' & -R'T \end{bmatrix} M \quad (3)$$

の関係を持つ。これらの式より、

$$w''m' = A'R'R^{-1}A^{-1}m = Hm \quad (4)$$

を得る。つまり、 $H = A'R'R^{-1}A^{-1}$ を満たす 3×3 の行列 H で、両画像の画素は 1 対 1 に対応する。 A', R' は画面構成の設定と対象物の 3 次元的位置や大きさから定める。具体的には、以下の通りである。

いま、 N 台の中の第 k 番目のカメラ C_k ($k = 1, \dots, N$) の校正情報 A_k, R_k, T_k が既知とする。また、各画像上で対象領域の重心 c_k と面積 (画素数) s_k を求める。さらに、画面構成制御後の対象の画面上的重心位置 c' と面積 s' を設定する。

このとき、 c_k を対応点としたステレオにより、物体の 3 次元位置 M を推定できる。物体と各視点との距離 $L_k = \|T_k - M\|$ と焦点距離 f_k より、その物体を球体と見なしたときの半径 r を、

$$r = \frac{1}{N\sqrt{\pi}} \sum_k \frac{L_k}{f_k} \sqrt{s_k} \quad (5)$$

で推定する。これにより、それぞれの画像における領域抽出精度のばらつきを軽減する。ただし、対象

物は画面の極端に端で捉えられておらず、画面端で捉えることによる像の引き延ばしの影響は無視できるとする。

結果として得られた r を利用すると、合成視野のための焦点距離 f'_k は、

$$f'_k = L_k \sqrt{\frac{s'_k}{\pi r^2}} \quad (6)$$

で求められる。

求まった f'_k を使った A'_k を前提として、光軸回りの回転を行わないことで自由度を 1 つ少なくした上で、 M の投影点が c' となるように、 R' を定めることができる。最終的に得られた A'_k, R'_k を利用して各 I_k から I'_k を生成する。この I'_k は、設定した画面構成で、撮影と同一視点から対象を撮影した画像であり、得られた I'_k を切替え提示することで、最終的な映像を得る。さらに、撮影画像の外側の部分は予め用意しておいたパノラマ画像により補間する。

3 撮影視点間の中割り画像生成

次に、撮影視点の間の中割り画像を生成する。ここでは、2 視点画像の平行化を利用する。文献 [5] など、2 視点画像を平行化することで、画像上の点の対応付けに基づいて、中割り画像上での幾何学的に妥当な対応点位置が線形補間で求められるなどの利点があり、Image Based Rendering などでもよく用いられる。以下、校正情報を利用した平行化と、平面を利用した補間画像生成について述べる。

3.1 画像の平行化

ここでは、前述と同様にカメラの校正情報を利用して平行化を行う。2 つのカメラ視点の間の仮想視点を考え、2 台のカメラ C_i ($i = a, b$) の校正情報として、 A_i, R_i, T_i が既知とする。このとき、それぞれの画像平面上の 2 つの軸ベクトル x_i, y_i と奥行き方向のベクトル z_i は、 R_i の成分から直接知る事ができる。

いま、平行化のための画像平面として、2 視点を結ぶ直線に平行で、2 視点の視野になるべく近い視線方向を持つものを考える。具体的には、画像の 1 つの軸を表す単位ベクトル x' を

$$x' = \frac{T_a - T_b}{|T_a - T_b|} \quad (7)$$

とする。次に、もう1つの軸の単位ベクトル y' を、視線方向の平均ベクトル $(z_a + z_b)/2$ と x' に直交するように定める。さらに、奥行き方向の単位ベクトル z' を x', y' と直交するように定める。

得られた x', y', z' から、平行化画像のためのカメラの姿勢を表す回転行列 R' を構成できる。また、平行化画像の焦点距離として $f' = (f_a + f_b)/2$ とすると A' が定まる。この A', R' を用いて、前述の方法で撮影画像から平行化後の画像を生成する。

3.2 補間画像の生成

得られた平行化画像をもとに、中割りの画像を生成する。今、中割りの視点 C_t の位置を、2つの撮影視点 C_a, C_b を $t:1-t$ ($0 \leq t \leq 1$) に内分する点とし、撮影した2視点間で対応付けできる点 p_a, p_b を考える。この点の平行化中割り画像上での位置 p'_t は、平行化後の対応点 p'_a, p'_b を使って、

$$p'_t = (1-s)p'_a + sp'_b \quad (8)$$

と表される。この関係式を用いれば、内分比率 t に対して、2つの画像上の対応点が補間画像上でどの位置に現れるかが計算できる。さらに、 I'_a, I'_b の対応する画素位置の画素の値を t を使った比率でブレンドし、補間画像の画素値を計算する。

しかし、このためには、2つの画像上の対応付けが必要となる。ここでは、カメラ校正に利用したマーカを利用した平面状の背景部分の画像補間と、それらの平面に乗らない領域部分の画像補間を組み合わせる方法を提案する。

まず、カメラ C_a, C_b の入力画像 I_a, I_b を平行化し、画像 I'_a, I'_b を生成する。校正に利用したマーカの内、同一平面上に乗っているもののグループを考えると、その平面上の点の投影点は、それぞれの画像上の対応する点 m_a, m_b の間に、

$$\alpha m_a = H m_b \quad (9)$$

の関係が成り立つ。ここで、 H は 3×3 の行列で、ある平面上の4点の対応を使って求められ、 α は0でない定数である。

この H は空間内の平面毎に定められるもので、これを用いると、一方の画像の平面領域をもう一方の画像に正しく重ね合わせることができる。逆に、その領域に映っていても、平面上にない部分では、同じ H で変換しても像は重ならない。この性質を利用する

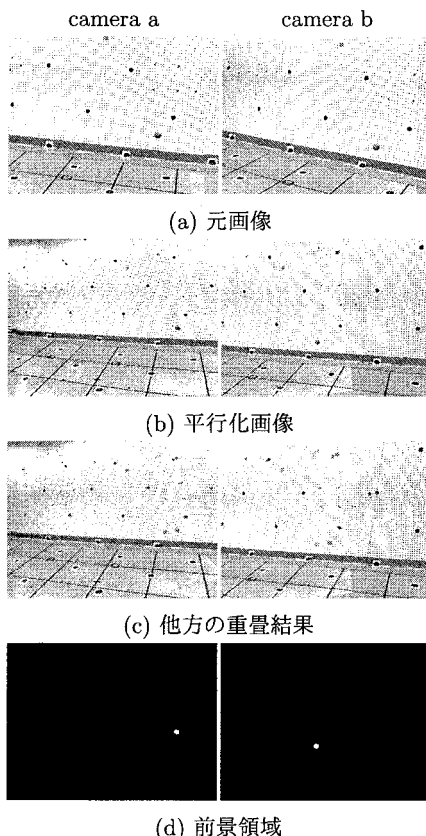


図3: 前景抽出処理例

と、画像の中で、(1) 想定する平面上である部分（平面背景）を抽出できる。

また、平面上にない部分も明確化される。1つの物体につき、2つの画像上にそれぞれ2つの領域が生じるが、平面との前後関係とテクスチャの類似度に基づいて判断し、(2) 前景領域と、(3) 前景により隠される平面背景領域を抽出する。この処理例を図3に示す。同図(c)では、式9で一方の画像に他方を重ねた時の平均画像だが、平面背景部分はほぼ正しく重なるが、前景部分（ボール）ではずれており、この部分から同図(d)の前景領域が抽出されている。

補間画像の生成では、まず、(1)の平面背景については、校正用マーカを使った点の対応づけを利用し、中割り画像内の平面背景領域を生成する。この領域は、2視点から同時に観察できるので、内分比率 t を用いたブレンドで画素値を求める。次に(3)の部分は、平面背景ながら一方の画像では観

測できない部分なので、観測している側のみを利用して中割り画像上の画素値を決定する。

最後に、(2)の対象候補物体の領域は、対象領域単位でのモーフィングによる画像補間を行う。画像上の対象領域から、前述の方法で対象の3次元位置とおよその大きさが得られる。その位置を通り、平行化画像平面に平行な仮想面を考え、対象がその面に「貼り付いている」と見なして補間する。仮想面上にそれぞれの画像から対象領域を投影し、さらに、今作ろうとしている中割り画像の視点位置の平行化画像上に投影し、対応する画素値を内分比率 t でブレンドする。

これらの処理で、平行化されている中割り視点画像と、その画像上の対象領域の重心、面積が求められる。最終的には、 t を適切な刻みで連続的に変えながら生成した平行化中割り画像から、前述の画面構成制御を施した画像を生成し、それらを順に提示することで切替え提示を行う。

4 実験

4.1 画像生成実験

カメラ5台を図4のように配置し映像生成実験を行った。撮影された多視点映像のある時刻のフレームを図5に示す。ボールを被写体として考えている。それぞれの撮影画像に対して、ボールの領域を抽出し、得られたボールの三次元位置と大きさに基づいて、撮影画像の画面構成を制御した例を図6に示す。図5と図6を比べると、特定の対象の画面構成の連続性を持つ切替えが実現されていることがわかる。

さらに、視点間の補間画像の例を図7に示す。平面背景の合成では、一方の画像で観測できない部分を他方の画像で穴埋めされていることも分かる。比較的単純な形状の被写体ではあるが、対象部分も鑑賞に耐える合成結果が得られている。しかし、 $t = 0.5$

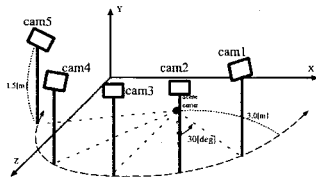


図4: 実験のカメラ設置状況

の合成結果では本来重なるべき部分にずれが生じている。この主な原因はレンズ歪みなどの影響と考えられ、今後これらの影響を軽減する必要がある。

4.2 手法とその効果の比較

注目対象誘導、ならびに本手法による中割り画像の補間による効果を検証するために、以下の3種類の映像を用いて主観評価実験を試みた。

手法1 撮影像の切り替えのみ (図5)

手法2 撮影像+画面構成制御 (図6)

手法3 撮影像+画面構成制御+画像補間 (図7)

前述と同じ状況とし、(項目1)映像の見やすさ、(項目2)映像の品質、(項目3)注目対象(ボール)の注視のしやすさ、および、(項目4)映像としての違和感について9名の被験者にアンケート調査した。2手法の映像を同時に提示し、各項目で比較し、 $>$ 、 $=$ 、 $<$ の三種の記号で回答してもらった。それぞれの記号を0, 1, 2点に換算した平均値を表1に、評価の分布を図8に示す。値が0に近いほど左側の手法が、2に近いほど右側の手法が優位で、1が中立である。

この結果より、項目3は手法1より2, 2より3の評価が高く、画面構成制御の効果を裏付けている。また、手法2と3の比較では手法3の評価が高く、中割り画像補間の効果も分かる。全体として、被験者数が少ないものの、手法1<手法2<手法3の順により評価が得られ、本手法の効果が確認できた。

一方で、見やすさや品質の項目では、手法1が良いという回答が見られた。これは、画像処理操作が少ないほど画質の劣化が抑えられることによる影響と考えられ、適切な解像度設定や画素位置計算の精度の向上など、合成品質の向上が必要である。

5 まとめ

本報告では、提示映像の滑らかな切替えを実現するために、画像の平行化とモーフィングに基づく処理で、撮影視点間の中割り画像を合成する方法を示した。従来法と比べて、カメラ台数に依存しにくい滑らかな切替え提示が行えた。また、比較評価実験で本提案手法の有効性を確認した。

しかし、今回の補間画像は撮影視点を折れ線ではないもので、切替え提示の仮想的な視点移動は滑

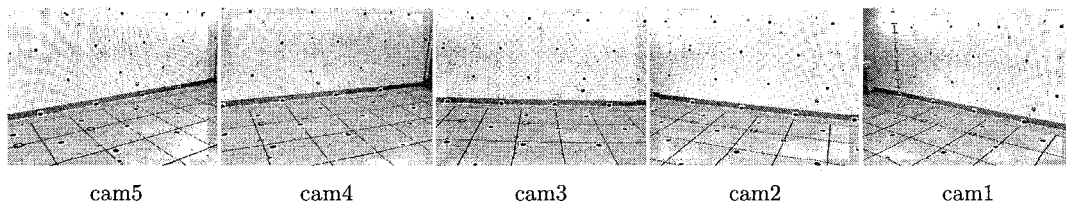


図 5: 元画像

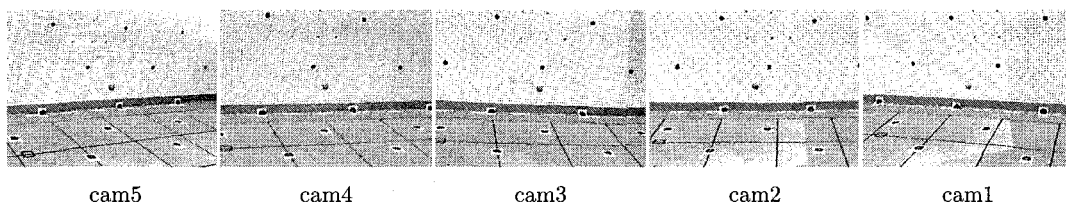


図 6: それぞれのカメラに対する画面構成制御画像

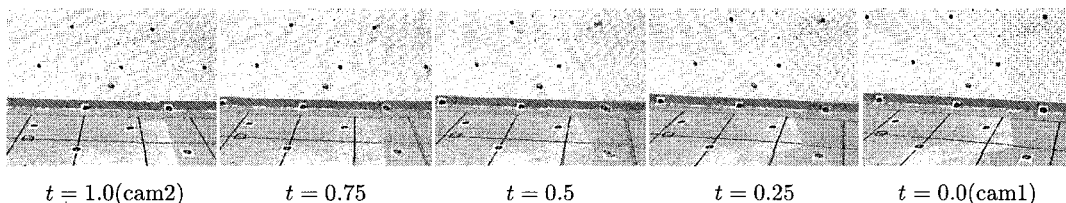


図 7: ボールを中央に捉える画面構成での、cam1 と cam2 間の補間画像例。

らかとは言えない。今後、仮想視点移動の経路を考慮した画像内挿法を検討する必要がある。また、平面背景という仮定を除き、より一般的な状況に対する拡張が必要である。また、評価実験の設問、被験者数とも十分とはいえず、評価法の検討も今後の課題である。

表 1: 評価実験結果

手法	項目 1	項目 2	項目 3	項目 4	平均点
1 対 2	0.78	0.67	1.44	0.89	0.94
1 対 3	1.56	1.11	2.00	1.56	1.56
2 対 3	1.11	1.78	1.67	1.44	1.50

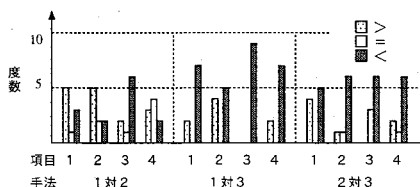


図 8: 評価実験の評価値分布

参考文献

- [1] Carnegie Mellon Goes to the Super Bowl, <http://www.ri.cmu.edu/events/sb35/tksuperbowl.html>
- [2] 富山仁博, 宮川勲, 岩館祐一, “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-”, 信学技報, Vol.106, No.429, PRMU2006-178 (pp.43-48) (2006)
- [3] 松田隆夫, “視知覚”, 培風館 (1995)
- [4] 出口光一郎, “ロボットビジョンの基礎”, コロナ社 (2000).
- [5] S.M.Seitz and C.R.Dyer, “View Morphing”, Proceedings of SIGGRAPH'96, pp.21-31 (1996)
- [6] 東海彰吾, 宮川栄一: 首振りカメラによる対象追跡映像からの仮想カメラワーク映像生成, 情処論, Vol.44, No.SIG9(CVIM7), pp.11-20 (2003)
- [7] S.Tokai and H.Hase, “Attention Navigation by Keeping Screen Layout for Switching Multiple Views”, ICPR 2006, D04_0539.PDF in CD (2006.8)
- [8] 東海彰吾, 樋口翔太, 長谷博行, “多視点映像切り替え提示による注目対象誘導における校正情報を用いた画面構成の安定化”, MIRU 2007, p.589 (IS-3-27-P0255.pdf in CD) (2007).