情報処理学会論文誌　Vol. 52　No. 1　68–76 (Jan. 2011)

会議支援用 Visual Hull における顔部分の一貫性を確保するテクスチャリング手法

坂本竜基[†1]　宮田慎也[†1]

本論文では，自由視点映像を顔部分の品質維持が要求されるグループウェアに応用する目的で，Visual Hull における特定部分のテクスチャに起こる不整合を抑制するレンダリング手法を提案する．これは，顔部分のテクスチャが貼られるボクセルに対してテクスチャを取得するカメラ画像を単一に限定することで実現される．このカメラを選択する基準を選択されたカメラの数に依存する方法と仮想視点に依存する方法の 2 種類あげた．また，提案手法のレンダリング結果と通常のレンダリング結果について，どちらでレンダリングしたほうが顔の表情を読み取りやすいか被験者実験によって評価したところ，提案手法を用いるほうが優れているという結果を得た．

Facial Consistency Based Texturing of Visual Hulls for Conferencing Applications

Ryuuki Sakamoto[†1] and Shinya Miyata[†1]

This paper describes a rendering method for 3D free viewpoint videos with Visual Hull. The method attempts keeping quality and coherence of textures on human face mapping on its Visual Hull. For keeping coherence, one camera image is selected as the texture for whole voxels corresponding with facial portion. Two criteria of selection are proposed that depends on the number of selected cameras as a texture and virtual camera position. As a result of comparison between rendered images with another criterion and the normal rendering method, the proposed method is finer than the normal method.

†1 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University

1. はじめに

同期型のグループウェアとして，会議の様子を遠隔地に映像と音声で伝えるテレビ会議システムは古くから研究されている[1]．また，非同期型のグループウェアとして，会議やミーティングを映像と音声で保存するアプリケーションも研究されている[2)–4)]．どちらの研究においても映像はテキストでは欠落する非言語情報も含むため，会議の内容を伝達するメディアとして有効であるとされている[5),6)]．しかし，このようなアプリケーションの多くはカメラの映像をそのまま利用するか，全方位カメラの映像をトリミングしているだけなので，通常の映像しか閲覧することはできない．

一方で，複数台のカメラを用いて，閲覧者が実際にはカメラがない位置から見たかのような映像を生成する技術がコンピュータビジョンの分野で提案されており，その出力映像は自由視点映像と呼ばれている．近年では，自由視点映像は応用研究もさかんに行われ，たとえばスポーツを自由視点映像化する試みもなされている[7]．この技術を用いることで，会議やミーティングの様子を自由視点映像化することができれば，たとえば，カメラ位置をある会議参加者の視点に合わせ，参加者の目線から会議の様子を閲覧したり，自動的にカメラワーク[8] を付与したりすることにより通常の映像より理解しやすい映像を生成するといったことも可能になるであろう．

しかし，自由視点映像の生成によく用いられる Visual Hull[9] を用いたアプローチでは，その特性から映像に大きな劣化が起こり，観賞に耐えられない映像が生成される恐れがある．特に，人間は画像に写ったものの中で顔に注目する傾向があり[10]，グループウェアの分野でもゲイズアウェアネス[11] を筆頭に顔は非常に重要な情報であるとされているが，カメラの配置や仮想視点等の条件により顔部分のテクスチャに大きな亀裂が入った映像がしばしば生成されてしまう．

本論文では，主に会議やミーティングへの参加者の自由視点映像を生成する目的で，Visual Hull における顔部分に対するテクスチャの不整合を抑制するレンダリング手法を提案する．Visual Hull の品質向上には様々な手法が提案されているが，どれも比較的重い処理であるうえ，カメラ台数も多いスタジオのような環境での撮影を必要としている．これに対して提案手法は，従来手法に比べて処理時間が短く，かつ顔部分の補償に特化できる点が特徴であり，さらに，少ない台数のカメラで撮影した精度の低い Visual Hull にも対応しているため，将来的にテレビ会議システムへの応用も期待できる．

以下，2 章において Visual Hull においてテクスチャの不整合が起こる原因を述べ，3 章

でこの問題に対する既存アプローチを述べる．次に，4 章で本研究で提案するアルゴリズムを述べ，レンダリングの結果を 5 章で示す．6 章において，このレンダリング結果の優劣を被験者実験により明らかにし，最後に 7 章で全体を総括する．

2. 少数のカメラにより復元された Visual Hull の問題点

自由視点映像の生成には様々なアプローチが存在するが，その中でも Visual Hull と呼ばれる映像から被写体の 3 次元形状モデルを推定する方法を用いるのが一般的である．自由視点映像を実現する手法は Visual Hull のほかにも存在し，テレビ会議に応用するシステムはすでに提案されている．たとえば，Lanier のシステム[12] では，テレビ会議において被写体を取り囲むように 7 台のカメラを配置して，カメラから見た人物の深度マップを推定し，それにテクスチャマッピングを施すことで簡易的な形状モデルをレンダリングするアプローチを採用している．しかし，一般的に深度マップの作成を高速かつロバストに行うことは非常に困難なうえ，視点位置の自由度も低い．これに対して Visual Hull は視点位置の自由度が高い映像を頑健に生成できることから多くの研究で用いられている．1 度，被写体の形状モデルが推定されれば OpenGL 等の 3 次元コンピュータグラフィックス用ライブラリを用いてそれを 3 次元仮想空間内に復元することができる．これにカメラ画像を適切にテクスチャマッピングすると，ユーザは自由な位置から被写体を閲覧することができるため，映像を構成する各画像に対してこれらの処理を連続的に適用すると自由視点映像として出力される．

Visual Hull はカメラ画像における被写体のシルエットから 3 次元空間における被写体の 3 次元形状が削り出されたものであり，一般的に凸包体の推定しかできないが，少なくとも数十台以上のカメラを用いれば比較的正しい形状を推定することができる．しかし，ビデオチャットや会議アーカイビング等のアプリケーションに応用することを考えると，カメラを大量に配置することは現実的ではない．たとえば，会議をアーカイビングする目的で長机の 3〜5 名程度を撮影したり，ビデオチャットのように着席した人物を撮影したりするのであればカメラは 4〜8 台程度の配置が現実的であろうと考えられる．

しかし，少数のカメラを用いる場合は Visual Hull による形状の推定精度が著しく低下することがある．**図 1** は 4 台のカメラを入力として Visual Hull を用いた直方体の推定を示した模式図であるが，本来の形状である長方形に対して，8 角形の形状が推定されてしまっている．このように少数のカメラで形状を推定する場合，本来の形状からはかけ離れた非常に荒い多角形として近似されてしまう．また，そもそも凸包体ではない部分は正しい形状を復元することができないため，髪と額の間や鼻の淵，人間の顔等の凹部分は最良の状態でもその部分を埋めるような平面として近似されてしまう．

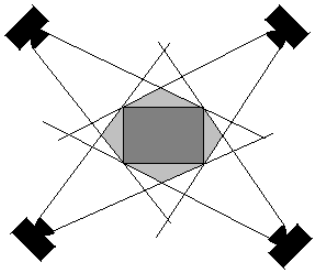

図 1　Visual Hull による形状復元
Fig. 1　3D model reconstruction with Visual Hull.

Visual Hull の復元は，実装上，撮影空間をボクセル空間として離散化した後に行われ，たいていは復元されたボクセルモデルをマーチンキューブ法[13] 等でメッシュ化し，さらにスムーシング等の処理を経て最終的にメッシュモデルとして生成されることが多い．しかし，このようなメッシュ化されたモデルは，元々のボクセルモデルの精度が低い場合にはテクスチャに不連続な部分が生じやすい．そこで，ボクセル部分に仮想視点に正対する微小面（以後，ファセットと呼ぶ）を立てて，そこにテクスチャマップを施すことで写実的に一貫性を持ったレンダリングを行うマイクロファセット・ビルボーディング法[14] と呼ばれる手法も提案されている．この手法を用いると，非スタジオ環境下において作成された比較的精度の低いモデルでも，メッシュ化する方法よりも写実的なレンダリング結果を得られるとされている[15]．

しかし，たとえマイクロファセット・ビルボーディング法を用いても，形状モデルの誤差はテクスチャマッピング時に大きな問題をもたらすことがある．たとえば，**図 2** のように仮想視点がカメラ A とカメラ B の中間部分にあったとしよう．形状が完全に正しい場合には C の部分は，カメラ A を用いてもカメラ B を用いても同じ情報が貼り付けられるため，どちらの画像をテクスチャとして用いても問題ない．しかし，形状が凸包体で近似された場合は，C の部分はカメラ A では円柱の右端，カメラ B では円柱の左端になってしまい，どちらの画像を用いても本来は中央部分のテクスチャが貼られるべき部分に，端のテクスチャが貼られてしまう．

この場合，そのファセットにカメラ A の画像をテクスチャとして用いるか，カメラ B の画像をテクスチャとして用いるかを決定する選択基準は，一般的にファセットから仮想視点

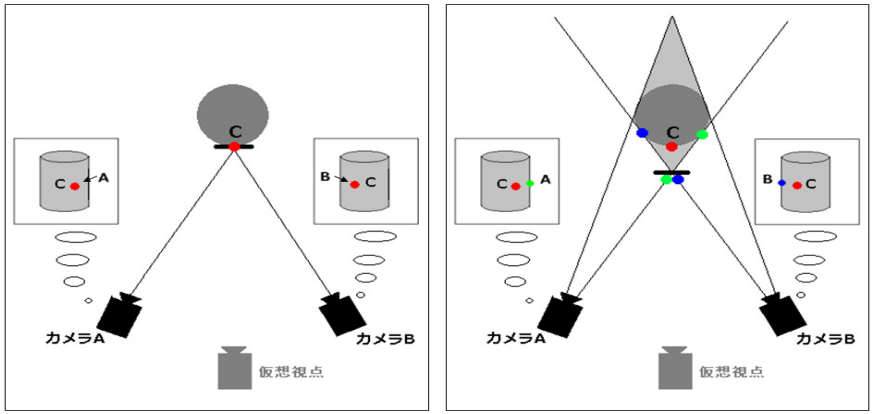

図 2　形状に誤差が引き起こすテクスチャの誤差
Fig. 2　Texture mapping errors caused by incorrect 3D shape.

へのベクトルと微小面からカメラへのベクトルがなす角度が最小のカメラを選ぶ方法が用いられる．よって，カメラの間に仮想視点がある場合は，C の部分の左側は，A のカメラの画像が，右側ではカメラ B の画像が使われるため，大きな不整合がおこる．実際，図のような極端な例ではなくても，テクスチャが別のカメラへと切り替わる境界部分では不整合が生じてしまうことが多い．

以後，本章で説明したようなマイクロファセット・ビルボーディング法で復元されたモデルに対して，仮想視点に依存したカメラ選択を行ってテクスチャマッピングするレンダリング手法を通常手法と呼ぶ．

3. 関 連 研 究

上述した Visual Hull における問題点を解決する手法としては，形状モデル自体を正しく矯正するアプローチが考えられる．このアプローチは問題の根本的な解決となりうるため，すでにいくつかの研究がなされている．延原らは，正しい形状モデル下では，どのカメラのテクスチャをマッピングしても結果が同じになるという特性を基に，初期状態の形状モデルを内側に徐々に矯正していく手法を提案している[16]．また，冨山らは，ステレオ視によって各カメラからモデルへの深度マップを作成し，そこから正しいモデルを推定する手法を提案している[17]．これらのアプローチは，非常に時間がかかる点が問題となるので，モデル自体は矯正せずにテクスチャ座標を適切にずらすことで，あまり時間をかけずにテクスチャ同士の不整合を解消するアプローチも提案されている[18]．

しかし，これらすべてのアプローチは非線形最適化を行っているため，実際は元々の形状モデルの精度がある程度高くなければ機能しないと考えられる．つまり，通常の部屋よりも広いスタジオ環境において，多くのカメラを用いて撮影されていることが前提となっている．一方で，本研究が撮影対象とするミーティングは通常の部屋で行われるため，実際は設置可能なカメラ台数には制限があり，精度良く Visual Hull を復元することは非常に困難である．

さらに，このような非スタジオ環境では，シルエットを正しくセグメンテーションできるとは限らず，間違ったシルエットが含まれると大幅な誤差がでてしまうという問題がある．特に，シルエットの一部が欠損した場合は，たとえば首から上が完全になくなるといった致命的な破綻を引き起こすことになる．そこで，このようなシルエットの精度に不安がある状況下では，ボクセル空間中へ各シルエットへの投影結果を投票し，ある程度の得票があればそのボクセルは存在しているとする実装方法がとられる．これにより，致命的な欠損は防ぐことができるが，本来はシルエットに問題のなかった部分も全体的に太ってしまい，結果としてモデル全体の精度が落ちてしまう．以上のような複合的な要因により，ミーティングを撮影する環境下において Visual Hull は正しく復元できないため，既存手法を適用して正しく補正することは困難と考えられる．

4. 提 案 手 法

2 章で述べたとおり，復元された形状の誤差がテクスチャの不整合を生みだすことが問題の本質であるが，大幅な幾何学的修正は困難である以上，テクスチャの見かけ上の不整合を防ぐ対策を講じるほかない．そこで，ミーティングの映像において特に重要である顔部分の整合性は必ず確保，つまり，顔部分については単一のカメラ画像のみを使用するようにする．これにより，全体的としては依然として不整合は残るものの，少なくとも顔部分の一貫性は保たれるため表情の判別等は容易になるはずである．また，最終的なレンダリングはメッシュを用いず，マイクロファセット・ビルボーディングを用いることによって，一貫性がより保たれるようにする．これらの処理の具体的な内容を以下に示す．

(1) Visual Hull を推定する．
(2) Visual Hull の各ボクセルにおいてマイクロファセット・ビルボーディングによるモデリングを行う．

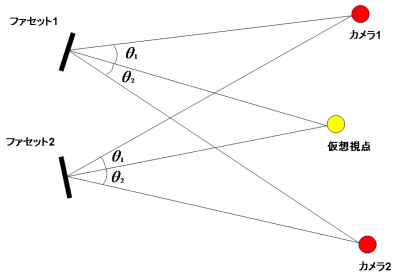

図 3　テクスチャ用カメラの選択
Fig. 3　Camera selection for texture mapping.

(3) 各ファセットに対してテクスチャ画像を取得するカメラを選択する．
(4) 各ファセットに対するテクスチャ座標を計算する．
(5) カメラごとにテクスチャ座標が顔の領域と交差しているファセットの数をカウントする．
(6) 顔領域とテクスチャ座標が交差しているカメラ群の中で，ある基準で選ばれたカメラ以外が選択されていた顔領域に相当するファセットのテクスチャ座標を，選ばれたカメラへのテクスチャ座標になるように再計算する．
(7) テクスチャ付きのファセットをレンダリングする．

このうち，処理 (3) で，各ファセットに貼るべきテクスチャを取得するカメラを選択するが，この選択には，テクスチャの一貫性を確保のために仮想視点に依存したカメラの選択方法を用いる．すなわち，**図 3** のように仮想視点，ファセット，カメラがなす角度 θ_i が最小となるカメラ i が各ファセットで選択される．

各ファセットの頂点を，ファセットに対して選択されたカメラの射影変換行列によって，カメラ画像に対するテクスチャ座標に変換する．これにより，ファセットの 4 頂点がカメラ画像に投影され，その 4 頂点による矩形領域を設定できる．もし，この矩形領域が顔の領域と重複する部分が存在するのであれば，そのカメラに設定されたカウンタをインクリメントする．また，同時にそのファセットに顔領域であることを示すマーキングしておく．この処理をすべてのファセットに対して適用した結果，ある基準に従って選択されたカメラの画像を顔領域に用いる唯一のテクスチャとして設定し，このカメラ以外が選択されていて，かつマーキングされているファセットのテクスチャと置き換える．この選択をする基準としては，たとえば以下のものが考えられる（以後，これらを選択基準 [a]，[b] と呼ぶ）．

[a] カウンタの値が最大のカメラ
[b] カウンタが 1 以上のカメラのうち，仮想視点の視線ベクトルとカメラの視線ベクトルのなす角度が最小のカメラ

顔の領域は，顔画像認識等によって各画像に対して顔のみを領域抽出した画像を用意してもよいが，たとえば肌色領域の抽出といった画素ごとに独立して判定可能な抽出方法を用いれば，処理 (4) の直後に判定が可能なので高速である．ただし，この場合は，単なる肌色認識であれば手や腕といった顔以外の肌色領域やノイズも顔領域と認識してしまうという問題がある．

5. 実　　験

5.1 GPGPU による実装

提案手法は，ボクセルごとに独立した処理が大半を占めるので並列化計算を行うと効率が非常に向上する．そこで，プログラマブルな GPU を利用する並列計算用ライブラリである CUDA を用いて処理 (1)〜(4) までの計算を並列的に処理し，その結果を OpenGL でレンダリングした．CUDA では，カメラごとのカメラ画像，シルエット画像，顔領域画像，射影変換行列を入力として，マイクロファセットの全頂点位置と各ファセットの頂点のテクスチャ座標，カメラごとのカウンタを出力する．CUDA 内部ではボクセルごとに処理 (1)〜(4) の処理を順次に行うが，このままでは Visual Hull の中身が詰まったモデルになり，処理 (7) のレンダリング時に負荷となる．そこで，中身の刳り抜き処理を処理 (1) と処理 (2) の間で行った．このため，全処理を並列に行うのではなく，処理 (1) と処理 (2) の間で全スレッドに対して同期をとった．

5.2 撮影環境と事前準備

個人用のブースに VGA の解像度を持つ IEEE1394 カラーカメラ（Pointgray 社製 Dragonfly2）を**図 4** のように 6 台設置した．これらのカメラは，あらかじめキャリブレーションをして射影変換行列を求めておいた．また，被写体のシルエット画像はレンダリング時にも作成可能であるが，その精度は一定ではない．よって，今回は各被写体に対する条件を一定にする目的で撮影後，事前に全カメラの画像に対して GrabCut [19] によるセグメンテーションを行い，比較的正確な被写体のシルエット画像を作成しておいた（**図 5**）．

5.3 結　　果

5.3.1 実 行 時 間

5.2 節の環境において 6 人の被写体を 1 人ずつ撮影し，その画像から CPU（Intel 社製 Core i7 930）と CUDA に対応した GPU（GeForce 280GTX）を装備した PC 上でレンダ

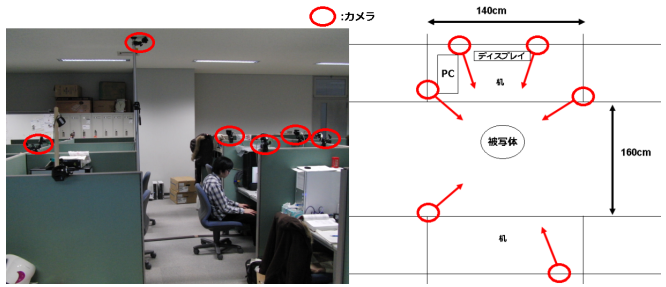

図 4　撮影環境とカメラ位置
Fig. 4　Environment and camera positions.

図 5　取得画像とシルエット画像
Fig. 5　Source images and silhouette images.

リングした実行時間の平均を示す．撮影対象空間は縦横奥行き方向にそれぞれ 128 分割したものをボクセルとした．(1) の結果である推定された形状のボクセル数は，どの被写体も約 30 万個であった．**表 1** は，それぞれの被写体を 6 カ所の仮想視点についてレンダリングした際に各ステップでかかった時間の平均を示している．ただし，各ステップごとに時間を計測するために無駄な処理が混入したり GPU と CPU 間の転送時間等で余分に所要するので

表 1　各ステップにおける実行時間（msec）
Table 1　Runtime by each step (msec).

手順		(1)	(2)	(3)	(4)	(5)	(6)	(7)	合計
選択基準	[a]	5.79	15.81	1.26	12.39	15.49	4.00	8.14	73.82
	[b]						5.64		75.46

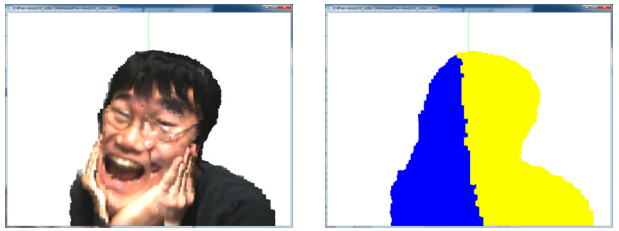

図 6　通常手法によるレンダリング結果
Fig. 6　Rendered image with normal method.

表の合計時間は正確ではなく，実際は全体として表の合計より約 5 msec ほど時間がかかる．

結果，既存手法でも必ず必要となる (1) の形状推定と既存の研究手法ではメッシュ生成に相当する (2) のモデリング部分が多くの時間を占めている．ただし，この時間は GPU の性能向上や最適化により大幅に短縮化する可能性があり，たとえば，既存研究では Homography による射影変換と補間によって高速にモデリングできることが報告されている[20]．実質的には，(5) と (6) が提案手法によって通常手法よりも余分に計算時間がかかる部分であるが，選択基準 [a]，[b] ともに約 20 msec とあまり計算コストがかかっていないことが分かる．

5.3.2　通常手法との比較画像

図 6 と**図 7** に，同条件下での通常手法と提案手法によるレンダリング結果を示す．これらの図における左側はレンダリングの結果であり，右側は各ファセットにおいて選択されたカメラを示した図である．右側では，同じカメラが選択されたファセットが同じ色で塗られている．

結果，通常手法によってレンダリングした結果である図 6 では，顔の中央部分にテクスチャの切れ目があり，顔が断裂している．一方，提案手法の結果である図 7 では，体部分にはテクスチャの切れ目があるものの，顔に関しては同一のカメラからのテクスチャのみで構成されているため，少なくとも顔に関しては違和感が少ない．なお，図 7 は，選択基準 [a] を用いた結果であるが，この条件下では選択基準 [b] を用いても同じレンダリング結果とな

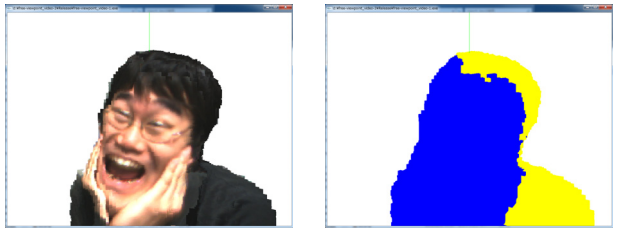

図 7　提案手法によるレンダリング結果
Fig. 7　Rendered image with proposed method.

図 8　通常手法と選択基準 [a]，[b]
Fig. 8　Comparison among normal mothod, selection criterion [a] and [b].

ることを確かめている．

5.3.3　カメラの選択基準に対する考察

ここでは，4 章で述べた顔部分に対する統一化されたテクスチャマッピングのために単一のカメラを選択する基準 [a]，[b] について，**図 8** に示すレンダリング結果とともにその特性を検証する．図 8 は，左が通常手法，中央が選択基準 [a]，右が選択基準 [b] を用いたレンダリング結果を示している．

まず，基準 [a] は，本来貼られるべきと主張するファセットが最大になる，換言すればマイノリティなファセットをマジョリティが選択しているカメラで塗りつぶす方針であり，処理 (6) で再計算する座標が少ないため多少高速である．しかし，視線方向とファセットの分布する位置によっては，マジョリティなファセットでカバーしきれない部分まで無理にテクスチャマッピングをすることになり，不自然な結果となることがある．これは，本来の各ファセットのテクスチャを選択する基準は，仮想視点とファセットとカメラがなす角であり，必ずしも仮想視点に対して近いカメラが選択されるわけではないことに起因する．極端な例では，仮想視点とあるカメラが非常に近い位置にいても，顔の位置が視線方向から遠くにあれば，隣のカメラのテクスチャが選択されることもありうる．図 8 は，このような状態を示しており，通常手法では顔の途中で，左側のカメラのテクスチャと右側のそれに分断されている．これを選択基準 [a] ではマジョリティである左側のカメラですべて置き換えてしまっているが，実際は仮想視点は左側のカメラの近くに位置しているため，特に右側のカメラが選択されていたファセット部分で不自然な結果となっている．

これに対して選択基準 [b] は，仮想視点とカメラの関係性にのみ依存しているため，このようなことは起こらない．図 8 においても，右側のカメラですべて置き換えられており，選択基準 [a] に比べて自然である．しかし，カウンタの値を比較に利用していないため，マイノリティなカメラによって他のすべてのファセットが塗り替えられる可能性があり，そのような場合は，選択基準 [a] よりも無理なテクスチャマッピングとなることがある．また，マイノリティによって塗り替えられる場合は処理 (6) の再計算に多少の時間がかかる．表 1 は 6 種類の形状モデルの平均なのでほとんど差が現れていないが，条件によっては 3 msec ほどの差があることがあった．もしこの時間を短縮したいのであれば，カウンタの値がある一定数以上の場合にのみ選択基準 [b] を用いて，それ以下では選択基準 [a] を用いる方法も考えられる．

6. 評 価 実 験

6.1 実 験 概 要

提案手法を用いて図 4 の環境で撮影されたマルチカメラ映像をレンダリングした結果を評価する被験者実験である実験 A〜D を行った．自由視点映像は，動画であることが念頭におかれているが，実験 A と実験 B では，被験者にテクスチャの品質をより正確に判別させるように，つまり，被写体の動きによって不整合が誤魔化されないように VGA の静止画としてレンダリングした 1 枚の画像を評価対象とした．一方で，静止画と動画での違いがないことを確認する目的と，本来目的としている自由視点映像によるアプリケーションへ応用した場合を想定して，実験 C では 15 frame/sec で撮影およびレンダリングされた VGA の動画を評価対象とした．さらに，実験 D では，提案手法をアプリケーションへ応用することを想定して，キーボードにより視点変更が可能なブラウザを作成し，それを用いて被験者に被写体を自由な視点位置から閲覧するタスクを課したうえで評価をしてもらった．

どの実験も，被験者は 21〜22 歳のコンピュータサイエンスを専攻している大学生である．画面のサイズが 11 インチのノート PC に異なる手法によってレンダリングされた同一条件下（同一の被写体かつ同一の仮想視点）での静止画もしくは動画を左右に並べて被験者に閲

図 9　実験 A に用いた画像の一部
Fig. 9　Images used for experiment A.

覧させ，左右どちらのほうが表情が判別しやすいか回答してもらった．1 回の評価実験で計 20 組の静止画および動画を次々に閲覧，もしくは 1 種類の自由視点映像を閲覧してもらったが，ある手法が右側や左側に偏らないよう，ランダムに位置の入れ替えをしてカウンタバランスをとった．すべての実験における帰無仮説は，両手法によるレンダリング結果に有意差はないということであり，その場合，同じような得票数となる．

6.2　実験 A：カメラの選択基準の比較

まず，カメラの選択基準 [a]，[b] 間で比較を行った結果を示す．被験者は 5 人であり，1 人につき 20 組の画像を閲覧してもらったので，計 100 個のデータを収集できた．その結果，49 組のデータにおいて選択基準 [a] が選ばれ，51 組のデータにおいて選択基準 [b] が選ばれた．また，χ^2 検定を行った結果，有意差は認められなかった．

本実験は，**図 9** に示すように図上段のように選択基準 A が有利であると考えられるケースと図下段のように選択基準 B が有利であると推測できるケースの両方が閲覧対象として含まれていた．これらの画像を出力する際に選択した仮想視点の位置は，選択基準 [a] と選択基準 [b] でレンダリング結果に差異があり，かつミーティングを撮影するカメラワークとして不自然ではない場所を選んだのであるが，全体的にはどちらのほうが優れているということはいえないようである．さらに，ほとんどの仮想視点において選択基準 [a] と選択基準 [b] の結果は同じになり，異なるのはカメラ間に仮想視点がある狭い区間だけであることも考慮すると，実際にアプリケーションに応用する際は，どちらの選択基準でも差はあまりないと結論付けられるであろう．

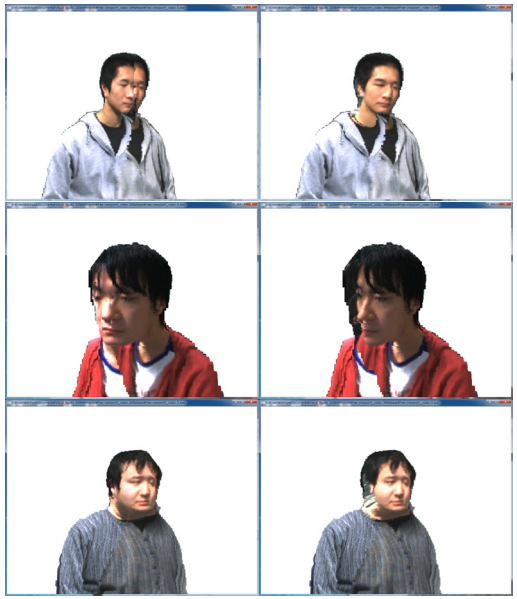

図 10　実験 B に用いた画像の一部
Fig. 10　Images used for experiment B.

6.3　実験 B：通常手法との比較（静止画）

次に，提案手法と通常手法を用いてそれぞれレンダリングされた静止画間の比較を行った結果を示す．ここでは提案手法として，上記実験でわずかながらも低い評価となった選択基準 [a] を用いた．被験者は 14 人であり，1 人につき 20 組の画像を用いたことから，280 個のデータが得られた．**図 10** は実験に用いた画像の一部であり，左側が通常手法，右側が提案手法である．図の最上段は比較的明確に提案手法の効果が見られるが，中段の画像は提案手法のほうがやや不自然な結果となっている．下段は，通常手法において顔の中央部分にテクスチャの切れ目があるが，目立たない．一方，提案手法でも比較的不自然なマッピングになっていないため，違いが分かりにくい結果となっている．このように，明らかに有利な条件だけでなく，様々な条件の画像も用いて実験を行った．その結果，提案手法による結果画像が 201 組のデータにおいて選択された．これは全体の 71%にあたり，χ^2 検定によって有

意水準 1%で有意差が認められた．この結果は，提案手法によるレンダリングが通常手法と比較して表情の判別等，顔部分のテクスチャが重要となるテレビ会議システムやミーティングのアーカイビングシステム等のアプリケーションにおいて Visual Hull を応用する際に有効であることを示唆している．

提案手法は，顔領域でのテクスチャの一貫性を維持する手法であるがゆえに顔と顔以外の領域間の不整合は起きる．これが大きな違和感につながらないかどうか実験後に聴き取り調査を行ったところ，全員が通常手法に比べ提案手法によるレンダリングに特に違和感は感じなかったと回答した．これは，顔以外の領域における不整合は，本手法によって悪化するわけではないという事実と，やはり顔領域に注目するという人間の習性に起因すると考えられる．

6.4　実験 C：通常手法との比較（動画）

15 fps で撮影された 15 秒の映像群に対して提案手法と通常手法を用いてレンダリングした映像に対して評価した結果を示す．被験者は 5 人であり，1 人につき 10 組の映像を閲覧してもらったことから，50 個のデータが取得できた．この実験も実験 B と同じような様々な条件下でレンダリングした映像を用いた．この結果，40 組の映像において，提案手法が優れているとの回答が得られ，χ^2 検定により有意水準 1%で有意差が認められた．この結果は，提案手法の優位性が静止画だけではなく動画でも発現することを示している．

6.5　実験 D：通常手法との比較（自由視点映像）

最後に，実験 C で用いた映像のうちの 1 つを自由視点映像として閲覧した場合の通常手法と提案手法との比較結果を示す．本実験の被験者は 9 人であり，各被験者には視点をキーボードで動かすための説明および練習をしたうえで，時間制限を設けずに様々な位置から閲覧した場合について比較してもらった．本実験では 9 個のデータのみしか得られないので有意差を認めるまでには至らないものの，7 人が提案手法を選び，1 人が通常手法，1 人がまったく違いが分からないと回答した．また，実験後に両者にどのような違いがあったか口頭で回答してもらった結果，提案手法では顔の切れ目が気になった，提案手法の方が見やすいとの回答が多かった．この結果は，アプリケーションへ Visual Hull を応用する際に提案手法を適用する意義があることを示唆している．

7. おわりに

本論文では，Viual Hull を用いた自由視点映像をゲイズアウェアネス等の顔部分の品質保証が要求されるグループウェアに応用する目的で，顔部分のテクスチャに一貫性を持たせるレンダリング手法を提案した．これは，顔部分のテクスチャが貼られるファセットにおいて，テクスチャを取得するカメラを限定することで実現され，この限定するカメラの選択基準を 2 つあげた．この 2 つの基準のうち，どちらが顔の表情を読み取りやすいか被験者実験によって評価したところ，どちらかに優劣はつけられないことが示唆された．また，提案手法によるレンダリング結果と通常手法によるレンダリング結果を比較した結果，提案手法が優れているという結論が得られた．これにより，提案手法をもちいて Visual Hull をレンダリングすると，通常手法に大きな計算負荷を加えることなく，顔の表情が読み取りやすい自由視点映像が生成可能であることが示された．

提案手法は，品質の低いモデルに対してあまり計算コストをかけずにグループウェアにおいて重要視される顔部分のテクスチャの一貫性を，単一のカメラのテクスチャを用いるという戦略で補償した．しかし，被験者実験において少数ながら通常手法のほうが優れているとされた画像では，2 つのカメラのテクスチャが顔に含まれているにもかかわらず，さほどの不整合はおこっておらず，むしろ単一のカメラを無理にマッピングするよりも自然な結果となった．そこで，整合性を安価に計測できれば，無理に提案手法で補正する必要がなくなり，レンダリング結果がより安定すると考えられるが，これは今後の課題とする．

参 考 文 献

1) 井上智雄，岡田謙一，松下　温：空間設計による対面会議と遠隔会議の融合：テレビ会議システム HERMES，電子情報通信学会論文誌 D，Vol.J80-D2, pp.2482–2492 (1997).
2) Stiefelhagen, R., Chen, X. and Yang, J.: Capturing Interactions in Meetings with Omnidirectional Cameras, *International Journal of Distance Education Technologies*, Vol.3, No.3, pp.34–47 (2005).
3) Rui, Y., Gupta, A. and Cadiz, J.J.: Viewing meeting captured by an omnidirectional camera, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, pp.450–457, ACM Press (2001).
4) Lee, D.-S., Erol, B., Graham, J., Hull, J.J. and Murata, N.: Portable meeting recorder, *MULTIMEDIA '02: Proc. 10th ACM International Conference on Multimedia*, New York, NY, USA, pp.493–502, ACM Press (2002).
5) 坂本竜基，金　韓成，伊藤禎宣，鳥山朋二，北原　格，小暮　潔：全方位カメラによる会議撮影システムが意思決定の非同期的伝達に及ぼす影響の評価，情報処理学会論文誌，Vol.50, No.1, pp.289–301 (2009).
6) Takao, S. and Innami, I.: The effects of the two modes of video-conferencing on the quality of group decisions, *Proc. SIGCPR '98*, New York, NY, USA, pp.156–158, ACM Press (1998).

7) 北原　格，大田友一，斎藤英雄，秋道慎志，尾野　徹，金出武雄：大規模空間における多視点映像の撮影と自由視点映像生成，映像情報メディア学会誌，Vol.56, pp.1328–1333 (2002).
8) 井上智雄，岡田謙一，松下　温：テレビ番組のカメラワークの知識に基づいた TV 会議システム，情報処理学会論文誌，Vol.37, pp.2095–2104 (1996).
9) Magnor, M., Pollefeys, M., Cheung, G., Matusik, W. and Theobalt, C.: *Video-based rendering*, AK Peters (2005).
10) Cerf, M., Harel, J., Huth, A., Einhäuser, W. and Koch, C.: Decoding What People See from Where They Look: Predicting Visual Stimuli from Scanpaths, *5th International Workshop on Attention in Cognitive Systems*, Berlin, Heidelberg, pp.15–26, Springer-Verlag (2009).
11) Ishii, H. and Kobayashi, M.: ClearBoard: A seamless medium for shared drawing and conversation with eye contact, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp.525–532 (1992).
12) Lanier, J.: Virtually There: Three-dimensional tele-immersion may eventually bring the world to your desk, *Scientific American*, Vol.4, pp.66–75 (2001).
13) Lorensen, W.E. and Cline, H.E.: Marching Cubes: A High Resolution 3D Surface Construction Algorithm, *Computer Graphics*, Vol.21, No.4, pp.163–169 (1987).
14) 山崎俊太郎，佐川立昌，川崎　洋，池内克史，坂内正夫：微小面ビルボーディングを用いた複雑なシーンの表示手法，画像の認識・理解シンポジウム（MIRU2002）予稿集，Vol.1, pp.127–132 (2002).
15) Kim, H., Kitahara, I., Sakamoto, R. and Kogure, K.: An immersive free-viewpoint video system using multiple outer/inner cameras, *3rd International Symposium on 3D Data Processing, Visualization and Transmission*, pp.782–789 (2006).
16) 延原章平，和田俊和，松山隆司：弾性メッシュモデルを用いた多視点画像からの高精度 3 次元形状復元，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，Vol.43, No.11, pp.53–63 (2002).
17) 冨山仁博，片山美和，折原　豊，岩舘祐一：局所的形状特徴に拘束された 3 次元形状復元手法とそのリアルタイム動画表示，映像情報メディア学会誌：映像情報メディア，Vol.61, No.4, pp.471–481 (2007).
18) 高井勇志，松山隆司：Harmonized texture mapping，映像情報メディア学会誌，Vol.63, No.4, pp.488–499 (2009).
19) Rother, C.: GrabCut: Interactive foreground extraction using iterated graph cuts, *ACM Trans. Graphics*, Vol.23, No.3, pp.309–314 (2004).
20) Wada, T., Wu, X., Tokai, S. and Matsuyama, T.: Homography based parallel volume intersection: Toward real-time volume reconstruction using active cameras, *Proc. Computer Architectures for Machine Perception 2000*, pp.331–339 (2000).

(平成 22 年 4 月 19 日受付)
(平成 22 年 10 月 4 日採録)

坂本　竜基（正会員）

1974 年生．2003 年北陸先端科学技術大学院大学知識科学研究科博士後期課程修了．同年 ATR 知能ロボティクス研究所研究員，ATR 知識科学研究所研究員を経て，2008 年より和歌山大学システム工学部講師．ATR 客員研究員（兼任）．CSCW，グループウェアの研究開発に従事．ACM，日本バーチャルリアリティ学会各会員．博士（知識科学）．

宮田　慎也

1985 年生．2010 年和歌山大学システム工学部情報通信システム学科卒業．現在，同大学大学院システム工学研究科博士前期課程に在学中．自由視点映像を応用したグループウェア・テレプレゼンスシステムの研究に興味を持つ．