

平成 22 年 5 月 31 日現在

研究種目：若手研究 (B)  
 研究期間：2008 ～ 2009  
 課題番号：20700122  
 研究課題名 (和文) ゲイズアウェアネスを考慮した自由視点映像によるミーティングの撮影に関する研究  
 研究課題名 (英文)  
 Gaze-awareness based rendering for 3D videos  
 研究代表者  
 坂本 竜基 (SAKAMOTO RYUUKI)  
 和歌山大学・システム工学部・講師  
 研究者番号：20395164

## 研究成果の概要 (和文)：

本研究では、自由視点映像を顔部分の品質維持が要求される会議支援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対する不整合を抑制するレンダリング手法を開発した。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を1つ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種類考案した。

## 研究成果の概要 (英文)：

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 are depended on the number of selected cameras as a texture and virtual camera position. As a result of comparison between rendered image with another criterion and the normal rendering method, the proposed method is finer than the normal method.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	1,500,000	450,000	1,950,000
2009 年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,800,000	840,000	3,640,000

研究分野：CSCW

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：グループウェア, ゲイズアウェアネス, 自由視点映像

## 1. 研究開始当初の背景

近年の映像機器の低価格化や記憶媒体の大容量化により、ミーティングを映像によって手軽に保存することが可能になってきた。これに伴い CSCW やグループウェアの研究分野において、単純なハンディカメラやビデオカメラによる撮影より手軽に扱えるツールの提案がなされてきた。例えば、卓上の全方位カメラや複数の小型カメラを用いて円卓状に集う参加者全員の上半身部分をキャプチャするものは、単純な監視カメラのように参加者の背中側から撮影する方法に比べ、カメラと被写体との距離が近いことにより参加者の表面積に対する解像度が高まり、表情や視線方向などを捉えやすい。このような参加者の表情や視線方向などの共有はアウェアネスと呼ばれ研究がなされてきた。アウェアネスは、遠隔地での協調作業や会議において単に文章や映像だけではなく、参加する成員が放つ臨場感、雰囲気、気配、熱気等といった非言語的情報の伝達・共有を目的としている。

数あるアウェアネスの研究の中でも、参加者の視線情報はゲイズアウェアネスと呼ばれ最も基本的な要素の一つとして古くから研究がなされてきた。しかし、既存のミーティングキャプチャシステムが撮影する映像では、カメラが参加者の視線以外の位置、すなわち卓上に固定されているため、参加者と視線を合わせるができない。これに対して本研究では通常の固定的な映像ではなく、仮想的な視点を任意の位置に設定可能な自由視点映像としてミーティングを撮影することで、あたかもある参加者の視点から会議に参加していたかのような映像の作成を実現する。

しかし、既存の自由視点映像は計算誤差や精度の問題から完全な映像を作成することはできない。この不具合は、テクスチャの断裂などによって発現し、目や顔の部分にこの不具合が起こればゲイズアウェアネスが著しく損なわれる恐れがある。

## 2. 研究の目的

本研究では、非同期・蓄積型のグループウェアとして視線情報（ゲイズアウェアネス）の保存及び再生を実現する自由視点映像によるミーティングの知的アーカイビングシステムの研究開発を目指し、特に自由視点映像におけるゲイズアウェアネスの補償技術を開発した。本研究によれば、既存の映像によるアーカイビングシステムに比べ、実際にミーティングの現場で共有されていた参加者間のコミュニケーションの機微を伝達することが可能となる。

## 3. 研究の方法

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アプローチは、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。一度、被写体の形状モデルが推定されれば、OpenGL などの 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。これにカメラからの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、これを映像に対して連続的に処理すると自由視点映像となる。

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、原理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、このようにカメラを大量に配置することは現実的ではない。例えば、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、このような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。例えば、4 台のカメラを用いて立方体の推定する場合、カメラの配置にもよるが、多くの場合で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う。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など人間の顔など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。

復元された形状の誤差がテクスチャの不整合を生み出すことが問題の本質であるが、大幅に幾何的な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他ない。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 1 つのカメラからの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ゲイズアウェアネスは確保されるはずである。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

トビルボーディングを用いることによって、この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おこなう
- (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). カメラ毎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

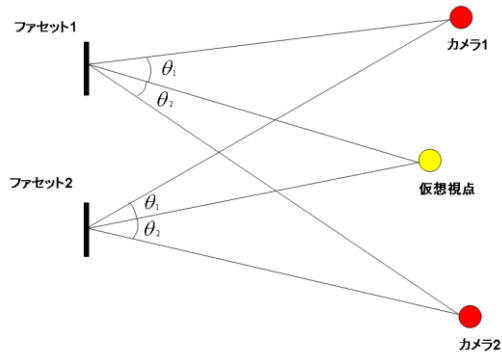


図1 テクスチャの選択

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1のように仮想視点、ファセット、カメラが成す角度 $\theta_i$ 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部分的にでも重複するのであれば、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、例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)。

顔の領域は、顔画像認識等によって各画像に対して顔領域のみを領域抽出した画像を用意してもよいが、例えば肌色の認識といった画素毎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問題がある。

#### 4. 研究成果

提案手法は、ボクセル毎に独立した処理が大半を占めるので並列化計算をおこなうと効率非常に向上する。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、カメラ毎のカメラ画像、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、カメラ毎のカウントを出力する。CUDA内部ではボクセル毎に処理(1)~(4)の処理を順次におこなうが、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大変な負荷となる。そこで、中身を削り抜きを処理(1)と処理(2)の間でおこなった。このため、全処理を並列におこなうのではなく、処理(1)と処理(2)の間で全スレッドに対して同期を取った。

図2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。上段が通常手法で、下段が提案手法である。また、それぞれ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ている。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上段では、顔の中央部分にテクスチャの切れ目があり、顔が断裂してしまっている。一方、提案手法の結果である下段では、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2下段は、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ることを確かめている。



図2 通常手法と提案手法の比較

次に、上述し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唯一のカメラを選択する基準[a][b]について、図3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3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



ダリング結果を示している。

図3 通常手法，選択基準[a][b]の比較

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成す角であり、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という事実に基づいている。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3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2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、非常に少数のマイノリティの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。

## 5. 主な発表論文等

〔雑誌論文〕(計 2 件)

- ① 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, 査読有。
- ② Hansung Kim, Ryuuki Sakamoto, Itaru Kitahara, Tomoji Toriyama, Kiyoshi Kogure, “Toward Cinematizing Our Daily Lives,” Springer International Journal on Multimedia Tools and Applications, vol. 44, No 1, pp. 87-110, 2009, 査読有。

〔学会発表〕(計 1 件)

- ① 宮田 慎也, 坂本 竜基, “ゲイズウェアネスを考慮した Visual Hull のレンダリング手法”, 情報処理学会 グループウェア研究会, 2010. 3. 18, 神奈川工科大学

## 6. 研究組織

### (1) 研究代表者

坂本 竜基 (SAKAMOTO RYUUKI)  
和歌山大学・システム工学部・講師  
研究者番号：20395164

### (2) 研究分担者

なし

### (3) 連携研究者

なし