

多視点カメラを用いたライトフィールドレンダリングによる 非平面隠背景可視化

前澤 桃子^{*1} 家永 直人^{*2} 森 尚平^{*1} 斎藤 英雄^{*1}

Abstract – 医療・教育・製造の分野において、その作業内容の教示を目的とした映像が利用されることがある。こうした映像は、作業者の様子と作業空間を同時に提示するため、一般に、教示者の視点とは別視点で撮影される。本稿では、映像を見る者が自身の視点と作業者の視点とをより直感的に照らし合わせられる映像を作成するため、作業者の一人称視点映像中の被遮蔽領域を可視化する手法を提案する。提案手法は、多視点カメラを用いたライトフィールドの概念に基づいており、特定の3次元位置付近を通過する光線を用いずに任意視点画像生成を行う。被遮蔽領域が平面でない場合も提案手法が有効に機能することを確認する。

Keywords : 隠消現実感, ライトフィールド, 多視点画像

1 はじめに

実物体を視覚的に隠蔽・消去・透過する技術は隠消現実感 (Diminished Reality; DR) として知られ、実空間を仮想物体によって拡張・融合する拡張/複合現実感 (Augmented/Mixed Reality; AR/MR) とは真逆の概念でありながら、技術的にはAR/MRの発展的課題として位置づけられている。これまでも、DR技術を用いた死角の可視化 [3] や家具再配置シミュレーション [4] 等への応用が提案され、こうしたDRによる可視化が特定の作業効率化にも効果的であることを示す文献も発表されている [5]。

我々は、こうした作業の効率化、遮蔽された空間の可視化による新しい映像表現を目指し、多視点カメラを用いたDR手法を提案してきた [1, 2]。特に、医療・教育・製造等の分野において、作業内容の教示を目的とした映像に写り込む作業者自身の手の視覚的除去及び作業空間の可視化を目的としたライトフィールドの構築法をこれまでに報告してきた [1, 2]。

これまでの手法 [1, 2] では、作業空間は平面で近似できるという隠背景平面拘束を設けていたが、本稿ではこの拘束を緩和し、かつ実時間で作業空間を可視化することを目的とする。提案手法は、多視点カメラを用いたライトフィールドの概念に基づいており、特定の3次元位置付近を通過する光線を用いずに任意視点画像生成を行う。この際、再構成する作業空間の焦点面を非平面にしてライトフィールドレンダリング [6] を行うことで、動的に変化する非平面で構成された作業空間の再構成を可能にする。また、隠背景平面仮定を置くこれまでの結果と、本稿での提案手法の作業空間可視化結果を比較し、その有効性を確認する。

2 提案手法

2.1 前報からの拡張内容

これまでの報告 [1, 2] では、作業領域を取り囲むように複数配置したカメラアレイを用いることを前提に、Buehler らの Unstructured Lumigraph Rendering [7] のレンダリングアルゴリズムに除去対象物体の代表位置をペナルティ点の一つとして与えるように拡張した Detour Light Field Rendering (DLFR) によってDR処理を実現した。この段階では、作業空間が1つの平面である、除去対象領域は固定である、という2つの制約を設けていたが、提案手法はこれらを緩和する拡張版である。以降、この拡張に際して追加した処理に関して述べる。

2.2 焦点面推定

既存手法 [1, 8] では、可視化したい空間にライトフィールドで再現した仮想焦点面を合わせることで除去対象物体の透視を可能にするため、焦点の合わない点に関しては背景の別々の点からの光線の合算となり、ボケが生じてしまう (図2)。そこで、体験者視点にお

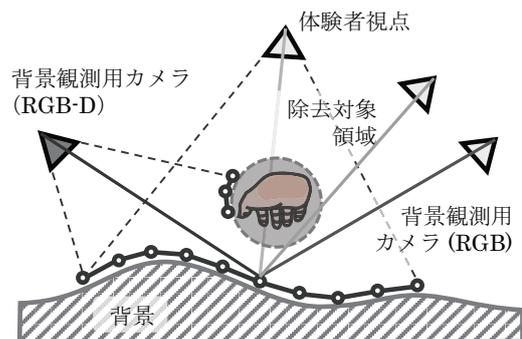


図 1: 非平面隠消現実感
Fig. 1 Non-planar DR

^{*1}慶應義塾大学 理工学部

^{*2}慶應義塾大学大学院 理工学研究科

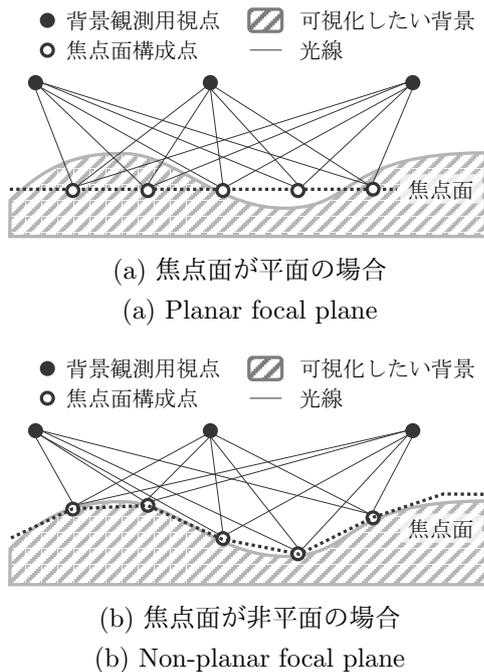


図 2: 焦点面の設定位置とボケの発生
Fig. 2 Focal plane and bokeh

ける隠背景の幾何形状を推定し、その表面に合わせて焦点面を設定する。このために、提案手法では RGB-D カメラを用いて背景の幾何形状を計測する (図 1)。RGB-D カメラは体験者視点とは別の視点位置に置くことで、作業領域を観測することを想定すると、体験者視点への変換と RGB-D カメラの性能の限界から生じるクラックが生じるため、これを軽減する必要がある。以降、各カメラの内部パラメータ及び位置姿勢は 2.3 節で述べる処理にて推定済みであるとする。

RGB-D カメラから得られる奥行画像を 3 次元空間に逆投影することで点群を生成し、これを体験者視点に投影することで体験者視点における奥行画像を生成する。まず、得られた奥行画像における奥行値 Z を持つ注目画素 $\mathbf{x} = (u, v)^T$ に対応する 3 次元位置 $\mathbf{X} = (X, Y, Z)^T$ を計算する。

$$\mathbf{X} = \begin{pmatrix} Z(u - c_x)/f_x \\ Z(v - c_y)/f_y \\ Z \end{pmatrix} \quad (1)$$

ただし、 f_x, f_y, c_x, c_y はそれぞれ焦点距離と投影中心に関する内部パラメータである。

次いで、注目画素 \mathbf{x} に関して、対応する 3 次元位置 \mathbf{X} 、その隣接する右と下の画素、 $\mathbf{x}_r = (u + 1, v)^T$ と $\mathbf{x}_b = (u, v + 1)^T$ に対応するそれぞれの 3 次元位置 $\mathbf{X}_r, \mathbf{X}_b$ から注目画素における法線 \mathbf{n} を計算する。

$$\mathbf{n} = \frac{(\mathbf{X}_r - \mathbf{X}) \times (\mathbf{X}_b - \mathbf{X})}{\|(\mathbf{X}_r - \mathbf{X}) \times (\mathbf{X}_b - \mathbf{X})\|} \quad (2)$$

RGB-D カメラ座標系から体験者視点カメラ座標系への変換行列が与えられたとき、この法線と 3 次元点を入力として、法線方向に向く面を Surface Splatting 法 [9] により体験者視点に合わせて描画する。この際、3 次元点群を作業領域とそれ以外に属するものに分ける。つまり、本節で述べる焦点面推定処理に利用する奥行画像と 2.4 節にて述べる除去対象領域推定処理に利用する奥行画像を生成する。例えば後に述べる実装では、作業台の上での作業を想定し、その作業台の上の面から一定距離に閾値を設けて、その閾値を下回る点を実作業領域内、上回る点を作業領域外に分けた。

こうして得られる体験者視点での奥行画像には RGB-D カメラの視点位置からは観測できない領域が存在し、これがクラックとなって表れる場合があるが、この領域は過去に観測された奥行画像を用いて埋める。この奥行画像は初期フレームで初期化した後、作業領域の奥行画像が得られる度に、これまでの奥行画像との加重平均により更新される (図 3)。

こうして、体験者視点における焦点面に関する奥行画像 D_f を取得した後、式 (1) と同様の方法で体験者視点における焦点面上設定用の 3 次元点群を任意の数だけ取得する。この頂点を三角形分割することで焦点面上に Geometric Proxy [7] (再構成基準面) を構築し、体験者視点における隠背景を再構成する [1, 2]。

2.3 カメラ位置姿勢推定

ライトフィールドを構築するために、RGB-D カメラ 1 台と RGB カメラ複数台の相対位置姿勢をバンドル調整を用いて推定する。体験者視点の各画素の画素値は、体験者視点の投影中心と各画素を結ぶベクトルが 2.2 節にて述べた焦点面と交わる点を計算した後、その点を他の RGB カメラに投影した位置に相当する画素の画素値の重み付き和で計算する。したがって、焦点面を基準にして各カメラの画素値を参照するため、その基準に合わせた位置合わせが必要となる。

提案手法では、RGB-D 画像から生成した 3 次元点群を用いて焦点面を構築するため、この 3 次元点群を利用して他の RGB カメラの相対位置姿勢を求める。まず、手動ないし特徴点マッチングによって RGB 画像間で対応点を取得する。RGB-D 画像内で検出された特徴点に関しては、その画素に対応する 3 次元位置が得られるため、その 3 次元位置と他の RGB 画像での対応点とを用いて Perspective-n-Point 問題を解くことでその RGB カメラの位置姿勢を求める。最後に、カメラ間での最適化を行うために、求めた位置姿勢と 3 次元点群を初期値としてバンドル調整を行う。

2.4 除去対象領域の決定

2.2 節にて述べた作業領域外の奥行画像から除去対象領域を決定する。提案手法では主として作業者の手

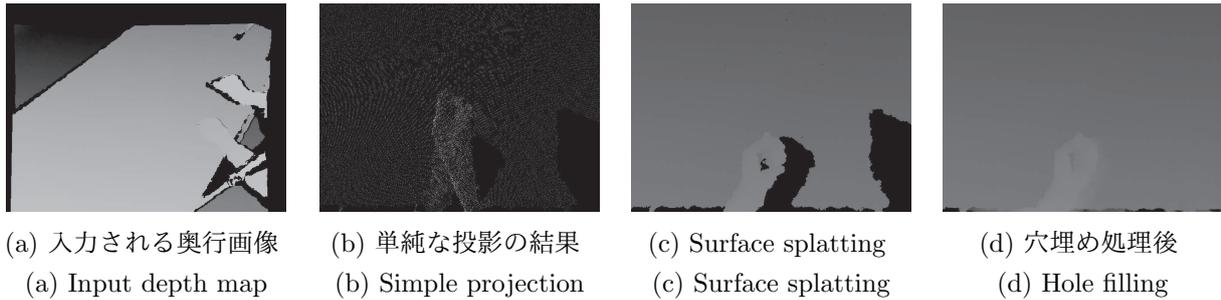


図 3: 焦点面推定処理 ((a) - (c) は手 (前景) と背景の分離後)
Fig. 3 Focal plane estimation (Fore- and background is separated in (a-c))

表 1: システム構成
Table 1 System specification

OS	Windows 10 Home 64bit
CPU	Intel Core i7-6567U 3.30GHz
Memory	16GB
GPU	Intel Iris Graphics 550

を除去して見せることを目的としているため、作業者視点においてその手は奥行画面中の下側から上側に向かって現れると想定する。そこで、奥行画像中の有効画素の内、縦方向に最も高い座標値を持つ画素に対応する3次元位置を手先の位置と考えた。実際の手の中心は手先よりも体験者の位置に近い位置に存在するため、一定のオフセットを与えて位置を調整した。

3 性能評価実験

3.1 評価方法及び機器構成

平面と非平面の作業領域に対して従来手法 [1] と提案手法を実行し、その結果を目視で確認する比較評価を行う。また、処理速度を計測し、提案手法が実時間性を保っていることを確認する。最後に、得られた結果について考察する。尚、従来手法では作業者の手が写っていない平面を写した初期フレームの奥行画像を与えてこれ以降更新せず、一方で、提案手法では毎フレーム更新した。

実験には、PC (表 1)、三脚に固定した3台の USB カメラ D_1 , D_2 (体験者視点), D_3 と1台の Kinect (RGB-D カメラ) D_4 (それぞれ 640×480 画素), 作業台として配置した机を用いた (図 4)。作業者視点において作業者自身の手に遮蔽される机上の状況を従来手法及び提案手法によって可視化し比較する。尚、用いたカメラの色調補正処理や同期処理は特段行わなかったが、機器を配置した時点で自動ホワイトバランス・露光機能によってこれらの値を取得し、後の処理では固定値を用いた。

各カメラの内部・外部パラメータは 2.3 節で述べたバンドル調整によって取得した。また、図 4 において、

作業者から見て左手前の机上の適当な位置に世界座標系の原点を置き、右方向を X 軸、上方向を Y 軸、手前方向を Z 軸として設定した。図 5 に本システム構成で取得できるライトフィールド用画像群の例を示す。尚、RGB-D カメラにはマイクロソフト社製 Kinect を用い、そこから得られる奥行画像からカメラ位置姿勢推定を行ったため、世界座標系のスケールは実スケールに合ったものが得られているとする。2.2 節における焦点面推定処理において、 40×30 頂点を有する再構成基準面を構築した。また、作業領域内外を決定する閾値として $Y = 30\text{mm}$ を設定した。

実験に用いるデータセットは、各カメラ 640 フレームの画像を撮影することで取得し、これらを PC に読み込んで実験を行った。提案手法が動的かつ非平面な作業領域において機能することを示すため、撮影中に作業者の手を写すと同時に作業領域で手や紙を動かしながら撮影を行った。

3.2 結果と考察

DR 処理前、従来手法 (隠背景平面仮定) [1] 及び提案手法による DR 処理結果像を図 6 に示す。

図 6 (a) は隠背景が平面であった場合の結果である。従来手法、提案手法ともに、除去対象である作業者の右手を除去できている。提案手法では RGB-D カメラから得た3次元点群をもとに焦点面を決定しているが、平面が抜けなく再構成され、RGB-D カメラで観測できなかった領域の穴埋め処理が適切に行われた。

図 6 (b), (c) は隠背景が平面でない場合の結果である。従来手法による DR 画像では、提案手法のものに比べて、隠背景に存在する作業者の左手が本来の位置から右にずれている。これは、作業者の左側に位置する RGB-D カメラの画像を用いて非平面の背景に対して、実際より奥に位置する平面の焦点面を用いて再構成したため、右に引き延ばされるように写った結果であると考えられる。提案手法では RGB-D カメラを導入して隠背景の幾何形状に合わせた焦点面を設定し、光線を取得したため、隠背景が非平面の場合にも周囲のテクスチャと整合性のとれた DR 処理結果が得られた。



図 4: 実験システム構成
Fig. 4 Experimental system setup

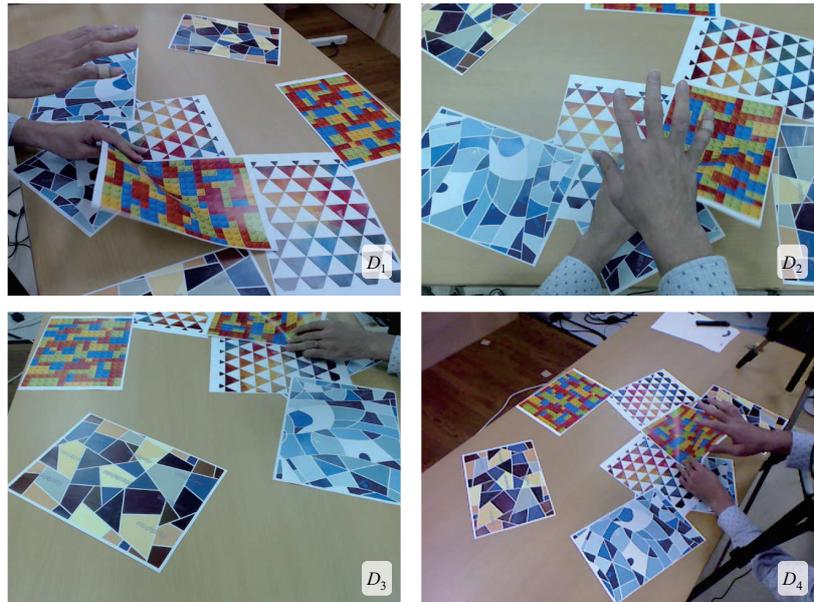


図 5: 実験システムで獲得できる画像の例
Fig. 5 Example images fetched with the experimental system

表 2: 各処理における実行速度
Table 2 Processing speed

処理内容	実行速度 (ms/frame)
焦点面推定 (2.2 節)	23.11
除去対象領域の決定 (2.4 節)	0.01
GPU への画像転送	0.09
DLFR[1]	1.20

各処理における実行速度を表 2 にまとめた。最も処理時間がかかったのは 2.2 にて述べた焦点面の推定に関する処理である。1 フレームあたりにかかる処理時間はおよそ 24.4ms (41.0fps) であったことから、一般的に必要とされるフレームレートである 30.0fps を上回り、実時間処理を達成できたと言える。

以上より、提案手法によって、従来手法 [1] の隠背景平面拘束を緩和し、動的に変化する非平面で構成された作業空間の可視化を実時間でできたと言える。

しかし、除去対象領域とその周辺とでの色調に関する光学的不整合が課題として残った。図 6 では、除去された手のあたりに、別視点の画像の画素を合成したこと起因する不連続な色調の変化が見受けられる。これを解決するためには、多視点画像の合成方法に更なる改善が必要である。また、2.2 節にて述べた通り、奥行画像を時間軸方向に関して平滑化しているが、それでも毎フレーム背景を更新するためジッターが発生する場面があった。加えて、カメラの台数が少ないため手に関する光線を除去しきれない場面が見受けられた。こうした理由から、ライトフィールド中の光線の密度を上げる必要があると考える。つまり、RGB-D カメ

ラ、RGB カメラともに、より多くの台数を用いるシステムへ拡張する必要性が感じられた。将来的により多くのカメラを導入する場合、先述の焦点面推定処理（主として Surface Splatting[9] とそれに伴う GPU への読み込みと書き込み処理）の高速化が望まれる。

4 むすび

本稿では、除去対象物体を通過する光線を避けて隠背景を再構成する従来手法 (DLFR) を拡張し、作業者視点映像における非平面の作業領域を実時間透視を可能にする DR 手法を提案した。そして、実画像を用いた実験により、隠背景平面仮定を置く従来手法と比較して提案手法が有効に機能することを確認した。

今後は、ブレンディング処理やカメラ配置の再考、そしてユーザスタディの実施に取り組む。

謝辞

研究の一部は学振・特別研究員奨励費 (課題番号 16J05114) 及び科学研究費基盤研究 (S) 24220004 の補助により行われた。

参考文献

- [1] 森 尚平, 前澤 桃子, 家永 直人, 斎藤 英雄: 多視点カメラを用いたライトフィールドレンダリングによる隠消現実感表示; 日本バーチャルリアリティ学会複合現実感研究会, MR2016-7, 2016.
- [2] S. Mori, M. Maezawa, N. Ienaga, and H. Saito: Detour light field rendering for diminished reality using unstructured multiple views; Proc. Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR),

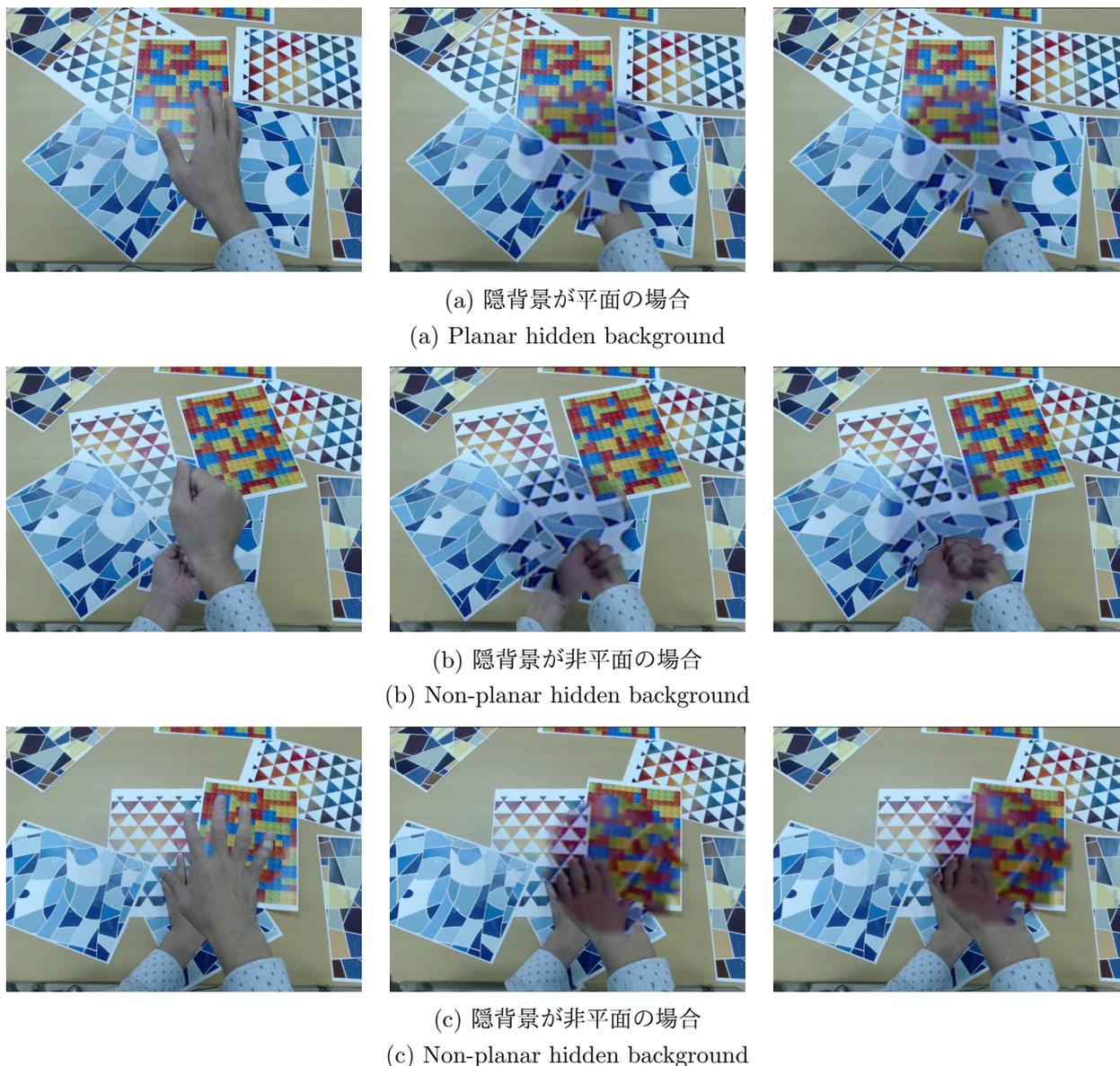


図 6: DR 処理結果の比較 (左から DR 処理前, DR 処理後 (隠背景平面仮定), DR 処理後 (提案手法))
 Fig.6 Comparison between DR results (From left: Input image, DR result (background planar assumption), DR result (the proposed method))

pp. 292 - 293, 2016.

[3] P. Barnum, T. Sheikh, A. Datta, and T. Kanade: Dynamic seethroughs: Synthesizing hidden views of moving objects; Proc. ISMAR, pp. 111 - 114, 2009.

[4] S. Siltanen, H. Sarasp, and J. Karvonen: A complete interior design solution with diminished reality; Proc. ISMAR, pp. 371 - 372, 2014.

[5] V. Buchmann, T. Nilsen, and M. Billinghurst: Interaction with partially transparent hands and objects; Proc. Australasian User Interface Conf., Vol. 40, pp. 17 - 20, 2005.

[6] M. Levoy: Light fields and computational imaging; *Computer*, 39(8):46 - 55, 2006.

[7] C. Buehler, M. Bosse, L. McMillan, S. J. Gortler, and M. F. Cohen: Unstructured lumigraph rendering; Proc. SIGGRAPH, pp. 425 - 428, 2001.

[8] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy: Using plane + parallax for calibrating dense camera arrays; Proc. CVPR, pp. 2 - 9, 2004.

[9] M. Zwicker, H. Pfister, J. van Baar, and M. Gross: Surface Splatting. Proc. SIGGRAPH, pp. 371-378, 2001.