

遠隔操作ロボットのための複数 RGB-D センサを用いた 半隠消映像の生成

杉本 和也^{*1}, 藤井 浩光^{*1}, 山下 淳^{*1}, 浅間 一^{*1}

Half Diminished-Reality Image Using Multiple RGB-D Sensors for Remote Control Robot

Kazuya SUGIMOTO^{*1}, Hiromitsu FUJII^{*1}, Atsushi YAMASHITA^{*1}
and Hajime ASAMA^{*1}

^{*1} Department of Precision Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan

This paper presents a method to compose half-diminished reality images for remote control robots using multiple RGB-D sensors. Robots are designed to perform the work of humans during a disaster response to reduce secondary disasters. In these cases, an operator controls the robot with the aid of images from a mounted camera. However, operating efficiency decreases significantly when the robot arm occludes target work objects in those images. Recently, the *Half-Diminished Reality* technique has been proposed in the field of computer vision. This technique is used for seeing through foreground objects and viewing occluded backgrounds. Accordingly, composing half-diminished reality images can enable the operator to see through the target work objects occluded by the robot arm in only one image, which results in increasing the operating efficiency.

Key Words : Remote Control Robot, Multiple RGB-D Sensors, Half-Diminished Reality

1. 緒 言

本研究では、遠隔操作ロボットのための操作向上を目的とした半隠消映像の生成手法を提案する。

災害時には人間が立ち入ることが危険な場所が数多く発生する。そのため、このような危険環境下における復旧作業の際には、二次災害を軽減することが重要である。そこで近年では、人間の代わりにロボットによる作業が期待されている。中でも、瓦礫の撤去などの作業では、車体にアームを搭載したロボットが用いられている。この場合、図1に示すように、オペレータはロボットに搭載されたカメラの映像を用いて遠隔地から操作を行うことが一般的である。一方、カメラ映像を用いた遠隔操作の場合、目視による操作と比較して作業効率が著しく低下する問題がある [1]。この原因の1つとして、手前にあるアームに遮られ、作業対象物の映像が十分に得られないことが考えられる。

この問題を解決するためには、ロボットに複数のカメラを搭載することが有効である [2]。複数のカメラを搭載することで、死角となる領域が減少するため、オペレータは作業対象物の位置・形状が確認しやすく

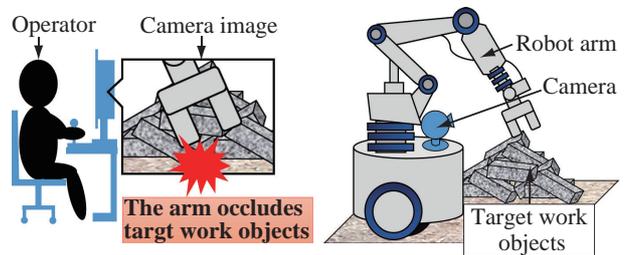


Fig. 1 Remote control system of a robot.

なる。しかしその一方で、多視点の映像を見比べながらロボットの操作を行う必要があるため、非常に高度な操作スキルを要することも報告されている [3]。そこでアームによって遮られて見えない作業対象物を、1つの映像上で認識可能にすることは有効である。

コンピュータビジョンの分野では、現実世界に存在する物体を映像上で隠蔽・除去する技術が存在する。中でも文献 [4] では、映像上の対象物体を完全に透明にするのではなく、あえて半透明の状態にすることで、対象物体に隠蔽されている背景が透視可能な表現のことを「半隠消現実感」と呼称している。このような表現は、See-through Vision [5,6] や X-Ray Vision [7,8] とも呼称されており、その多くはエンターテインメントや

^{*1} 東京大学大学院 工学系研究科 精密工学専攻 (〒113-8656
東京都文京区本郷 7-3-1) sugimoto@robot.t.u-tokyo.ac.jp

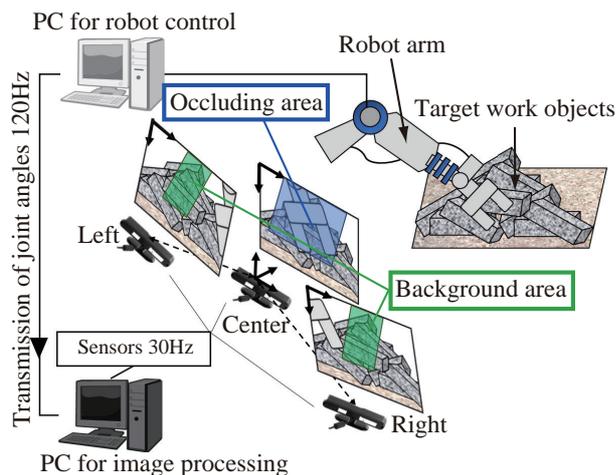


Fig. 2 Schematic view of our proposed method.

放映技術等に应用されている。ロボットに搭載されたカメラの映像に「半隠消現実感」を適用することで、アームによって遮られて見えない作業対象物が1つの映像で認識可能になる。そこで本稿では、複数のRGB-Dセンサを用いて、アームが搭載されたロボットを遠隔操作する際に、アームによって遮られて見えない作業対象物が透視可能な半隠消映像の構築手法を提案する。

2. 関連研究

半隠消映像を生成する際には、視野を遮蔽する物体に隠蔽されている背景の映像（以下、背景映像）を取得する必要がある。

背景映像の取得手法の1つに、視野を遮蔽する物体が映し出されていない場合の映像を、あらかじめ撮影しておく手法が存在する[9,10]。しかし、本研究のようにロボットを遠隔操作する場合、カメラにより撮影される映像は時々刻々と変化する。そのため、時系列映像を用いて生成した半隠消映像を、本研究に用いることは現実的ではない。

一方、多視点から撮影された映像を用いる手法も多数存在する[5-8,11]。多視点映像を用いることで、背景背景を別視点から撮影可能となるため、時々刻々と変化する環境へも適応が可能である。そのため、本研究においても複数のカメラによる多視点映像を用いることで背景映像を取得する。ただし、多視点映像を用いた多くの手法は、カメラから撮影対象物までの距離を十分に大きくとることで、背景の環境を2次元の平面と仮定している。そのため視野を遮蔽する物体の映像とその物体に隠蔽されている背景映像を、ホモグラフィ変換を用いて幾何変換を行った後、合成を行うことで半隠消映像の生成を可能にしている。中でも文

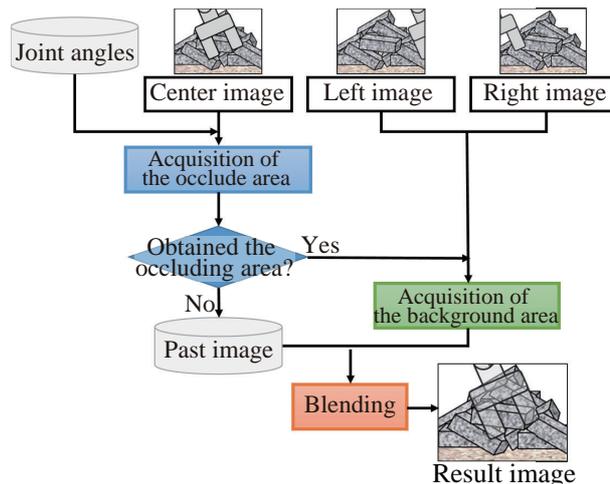


Fig. 3 Overview of procedures for composing image.

献[11]は、多視点映像を用いて生成した半隠消映像をロボットの遠隔操作に利用した事例であるが、こちらの事例も背景の環境を2次元の平面と仮定しているため、本研究で想定するアームが作業対象物に接近する場合や、作業対象物が複雑な形状の場合は提案されている手法の適応が困難である。

そのため本研究では、環境の3次元計測を行うことで様々な環境に適応可能な半隠消映像の生成手法の構築を行う。そこで次章より、環境の3次元計測手法および半隠消映像生成手法についての詳細を述べる。

3. 提案手法

本研究では、環境の3次元計測を行うためにRGB-Dセンサを用いる。RGB-Dセンサとは、RGBの色情報に加え、センサから撮影対象物までの距離情報が取得可能なセンサである。本稿では、RGB-Dセンサを合計で3台使用した場合の半隠消映像生成手法を述べる。またその際、各センサを十分な視差がついた映像が得られるように設置し、区別のためにそれぞれ左センサ、中央センサ、右センサと呼称する。

ここで、本研究におけるシステムの概要図を図2に示す。本研究では、各センサの映像において、ロボットアームが映し出されている領域を遮蔽領域とし、遮蔽領域に隠蔽されて見えない領域を背景領域と定義する。以上の環境において、目的とする半隠消映像を生成するためには以下の3つの処理を行う必要がある。

- (1) 遮蔽領域 (Occluding area) の算出
- (2) 背景領域 (Background area) の算出
- (3) 遮蔽領域と背景領域の合成

処理のおおまかな流れを図3に示す。次節より中央センサ映像の遮蔽領域に隠れて見えない背景領域を左右のセンサ映像より取得する手法を述べる。

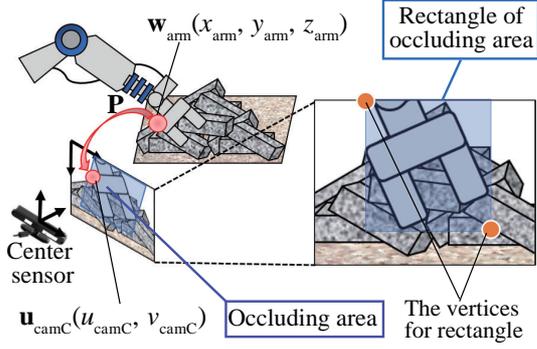


Fig. 4 Overview for acquisition of the occlude area.

3.1 遮蔽領域の算出 図4のように中央センサの光軸中心を原点として設定したワールド座標系において、ロボットアームの3次元形状データおよび各関節角度から計算可能なアーム位置座標 $\mathbf{w}_{\text{arm}} = [x_{\text{arm}}, y_{\text{arm}}, z_{\text{arm}}, 1]^T$ が、中央センサ映像の画像座標系において $\mathbf{m}_{\text{camC}} = [u_{\text{camC}}, v_{\text{camC}}, f_{\text{camC}}, 1]^T$ に像を結んだとき、行列 \mathbf{P} を用いると以下のような関係がある。

$$\mathbf{m}_{\text{camC}} \simeq \mathbf{P}\mathbf{w}_{\text{arm}}, \quad (1)$$

\mathbf{P} は中央センサの内部パラメータおよびワールド座標系に対するセンサの位置・姿勢を表す行列であり、カメラキャリブレーションにより事前の取得が可能である。そのため、中央センサ映像の各フレームにおいて、空間上のアーム位置座標を算出することで、中央センサ映像の遮蔽領域が取得可能となる。

ここで文献 [4] [10] によれば、視野を遮蔽する物体の輪郭を正確に算出し遮蔽領域を導くよりもあえて大きく領域をとることで、その後の合成処理を行ったときに画像合成特有の不自然さが緩和されるという報告がある。そこで計算量の減少を図ることで、リアルタイムに遮蔽領域を算出するためにも、画像中に投影されたアーム形状の輪郭に沿って領域を取得するのではなく、アーム形状を包括する矩形領域を取得する。そのため本研究では、ロボットアームの3次元形状を直方体の集合で近似し、この直方体の頂点のみを式 (1) を用いて透視投影変換する。その後、中央センサ映像上に投影された点のうち、図4に示すように画像中に投影されたアーム形状全体を包括する最小の矩形領域を本研究では遮蔽領域とする。

3.2 背景領域の算出

3.2.1 左右のセンサを用いた背景領域の探索

背景領域を左右のセンサ映像より求める。本研究では、任意の位置に設置した各 RGB-D センサに対して、各 RGB-D センサの相対位置関係から得られるホモグラフィ行列を用いて、左右のセンサの映像を中央セン

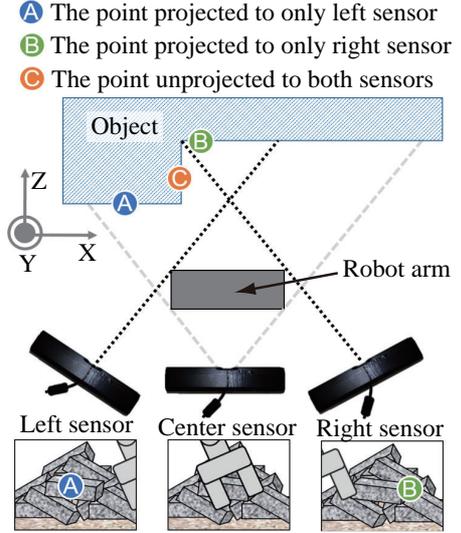


Fig. 5 Top view of experimental conditions.

サの光軸方向および視点の高さと等しくなるように仮想平面に投影する。

各 RGB-D センサの光軸方向をそろえ、視点の高さも等しく設置することで、平行ステレオを用いた三角測量の原理による位置計測が可能となる。そのため 3.1 節で取得した遮蔽領域の画像座標 $\mathbf{m}_{\text{camC}} = [u_{\text{camC}}, v_{\text{camC}}, f_{\text{camC}}, 1]$ に隠蔽されている背景物体の位置座標 $\mathbf{w}_{\text{back}} = [x_{\text{back}}, y_{\text{back}}, z_{\text{back}}, 1]$ が、左右のセンサ映像上で $\mathbf{m}_{\text{camL}} = [u_{\text{camL},i}, v_{\text{camL}}, f_{\text{camL}}, 1]$, $\mathbf{m}_{\text{camR}} = [u_{\text{camR},i}, v_{\text{camR}}, f_{\text{camR}}, 1]$ にそれぞれ像を結ぶとき以下の関係式が成立する。

$$z_{\text{back},i} - \frac{b_{\text{camLC}} \cdot f}{(u_{\text{camL},i} - u_{\text{camC}})} = 0, \quad (2)$$

$$z_{\text{back},i} - \frac{b_{\text{camCR}} \cdot f}{(u_{\text{camC},i} - u_{\text{camR}})} = 0, \quad (3)$$

f は、全てのセンサで共通の焦点距離の値であり、カメラキャリブレーションによって取得可能な既知の値である。また、 b_{camLC} は左センサと中央センサのセンサ間の距離であり、 b_{camCR} は中央センサと右センサのセンサ間の距離である。添え字の i は左右の映像上の画像座標において $v_{\text{camR}} = v_{\text{camL}} = v_{\text{camC}}$ のライン上における u 軸の値でありピクセル単位である。本研究では RGB-D センサを用いているため、各センサ映像の各ピクセルは RGB の色情報と任意のワールド座標系における位置座標を兼ね備えている。そのため、このラインに沿って、式 (2) および式 (3) をもとに作成した以下の評価関数 (4) および式 (5) を最小にする i の探索を行うことで、遮蔽領域に隠れている背景領域の取得が可能となる。

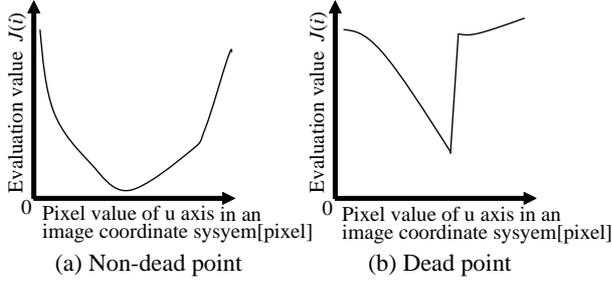


Fig. 6 Change of the evaluation value. The left side graph shows the case of a non-dead area, and the right side graph shows the case of a dead area.

$$J_{LC}(i) = \left| z_{back,i} - \frac{b_{camLC} \cdot f}{(u_{camL,i} - u_{camC})} \right|, \quad (4)$$

$$J_{CR}(i) = \left| z_{back,i} - \frac{b_{camCR} \cdot f}{(u_{camC,i} - u_{camR})} \right|, \quad (5)$$

実際に行う処理としては、3.1節で求めた遮蔽領域を縦に2分割し、それぞれ左遮蔽領域、右遮蔽領域と定義する。その後、左遮蔽領域は領域の左端を探索出発点とし、画像座標における u 軸の正方向に向けて一行ずつ探索を行う。反対に右遮蔽領域は領域の右端を探索出発点とし、 u 軸の負方向に向けて探索を行う。

一方、作業対象物の表面形状や、ロボットアームの位置・姿勢によっては、左右のセンサに投影されない点が存在する。そのため、以上の処理を行っただけでは、遮蔽領域に対応する背景領域が取得可能になったとは限らない。そこで、図5に示すように、遮蔽領域において隠れて見えない作業対象物の位置座標を以下の3通りに分類する。

- (A) 左センサのみに投影される位置
- (B) 右センサのみに投影される位置
- (C) いずれのセンサにも投影されない位置 (Dead Point)

例えば、作業対象物の位置座標が(B)または(C)のパターンの場合、上述の手法を用いて左遮蔽領域のあるピクセルに対して探索を行っても正しい結果は得られない。本研究では、このような位置座標を隠れの位置 (Dead Point) と定義する。そのため得られた背景領域に対応する作業対象物の位置座標が、隠れの位置にあたるのか否かを以下の手法を用いて判別する。

左遮蔽領域のあるピクセルに対して探索を行った結果、隠れがない場合、評価関数式(4)および式(5)は、図6(a)に示すように最小値付近では評価値 $J(i)$ が0に近い値をとる。一方、隠れがある場合は、図6(b)に示すように評価値 $J(i)$ が全体的に大きな値をとり、且つ

最小値付近では評価値 $J(i)$ が急速に増加する傾向がある。以上より、以下に示す2つの特徴量を用いて隠れの領域を判定を行う。

- (1) 評価値 $J(i)$ の最小値 e
- (2) 評価値 $J(i)$ が最小となるピクセル近傍での評価値 $J(i)$ の変化率 e'

ここで変化率 e' は、評価値 $J(i)$ が最小となるピクセル値 d とそれに隣り合ったピクセル値 $d-1$ 、 $d+1$ のうち $J(i)$ が大きいほうの値との差とし、以下のように定義する。

$$e' = \max(J(d-1) - J(d), J(d+1) - J(d)), \quad (6)$$

e および e' が任意の閾値より小さい場合を左センサに投影される位置とし、それ以外を左センサには投影されない位置とする。

その後、一度の処理で取得が不可能であった領域に対しては、探索に使用するセンサ映像を変更して再探索を行う。例えば、左センサを用いて左遮蔽領域のあるピクセルに対して探索を行い、隠れがあると判定された場合は、右センサを用いて上述した手法により再探索を行う。以上の処理を行った結果、取得不能であった遮蔽領域に関しては以下に示す手法により補間を行う。

3.2.2 時系列映像による補間 本研究では、環境が動的に変化することを想定しているため、 $t=T$ の状態の各センサ映像より半隠消映像を生成する。一方で、これまでの処理では取得が不可能な背景領域(以下、欠損領域)が存在することは避けられない。一般的にはロボットを遠隔操作することで作業を行う際には、ロボットは移動を停止し、アームのみを動かすことで作業を行なうことが基本である。そのため欠損領域が存在すると操作性に悪影響を及ぼすことが予想される。一方で過去の時刻 $t=T-1, T-2$ の映像を記録しているため、3.2.1項の処理で対応する背景領域が取得不能であった欠損領域においては、時系列映像を用いて補間を行う。

3.3 取得した領域の合成 3.1節で取得した遮蔽領域を半透明にし、3.2節で得られた背景領域と共に中央センサ映像に投影および合成処理を行う。投影位置は3.1節で求めた中央センサ映像上の遮蔽領域とする。本研究では、合成処理を行う際に以下の式を用いる。 I は各映像の各ピクセルにおけるRGBの色成分である。また α は不透明度を表す係数であり、 $\alpha=0$ で完全な透明となる。

$$I_{output} = \alpha I_{occluding} + (1 - \alpha) I_{hidden}. \quad (7)$$

4. 実験

4.1 実験内容 前章で述べた処理を用いて、視差のある3つ映像からアームに隠されて見えない背景が透視可能な半隠消映像を生成する。ロボットを遠隔操作する際に行う作業としては、前方の瓦礫の撤去などが存在する。そこで本実験では、ロボットを用いた瓦礫の撤去作業を想定し、3台のRGB-Dセンサ（ASUS: Xtion Pro Live）およびロボットアーム（YASKAWA: MOTOMAN-HP3J）を用いて、ロボットが前方の瓦礫にアクセスする際にアームに隠されて見えない瓦礫が透視可能な映像の生成を行う。

各機器の位置・姿勢を図7に示す。本実験では、図7のように設定したワールド座標系に対し、まずZ軸方向に150mm、Y軸方向に250mmの位置に中央センサを設置した。さらに中央センサからX軸方向に200mm離れている位置に左センサを設置し、X軸方向に-200mm離れている位置に右センサを設置した。取得映像サイズは640×480pixelとした。

実験環境の全体図を図8に示す。実際の作業現場で散在した瓦礫の撤去作業を行うことを想定し、瓦礫のモデルとして擬似的に再現した異なる2種類の形状の物体を用いた。1つ目としては画像が貼り付けられている1,000×800mmの板状のボードとした。2つ目としてはボードの手前に散在させた200×200×100mmの直方体形状の瓦礫を模したブロックとした。

使用したロボットアームモデルに対し、図9のように各関節角 $\theta_1 \sim \theta_6$ を設定した。ロボットアームが前方の瓦礫にアクセスする動作を想定し、 $\theta_1 \sim \theta_6$ を図9のように変動させた際の各センサ映像を入力映像とした。(7)式における透過値を表す係数 α は、半透明にする全ての領域で0.5とした。また、各センサのキャリブレーションは事前に済ませておいた。

ここで、実験の際に取得した入力画像の一例を図10に示す。これらの画像は図9に示す動作において $t = 5,000\text{ms}$ の際に各センサから撮影された画像である。本実験では、前章で述べた処理と同様、中央センサ映像（図10(b)）のアーム部分を遮蔽領域とし、この領域に対応する背景領域を左右のセンサ映像より取得する。

4.2 結果と考察 ロボットアームが前方の瓦礫にアクセスする際の、時間 $t[\text{ms}]$ におけるロボットアームの姿勢および中央センサから得られる入力映像と、入力映像に対応する出力映像を図11に示す。これらの画像群は、初期位置を $t = 0\text{ms}$ とし $t = 30,000\text{ms}$ に向けて、各関節角度を図9のように変化させた様子である。この結果、瓦礫が透視可能な半隠消映像がリ

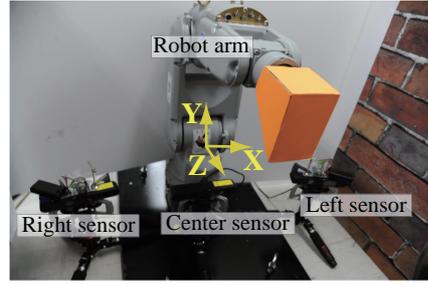


Fig. 7 The position of the RGB-D sensors.

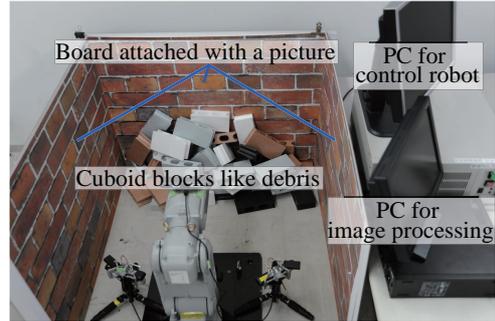


Fig. 8 Experimental condition.

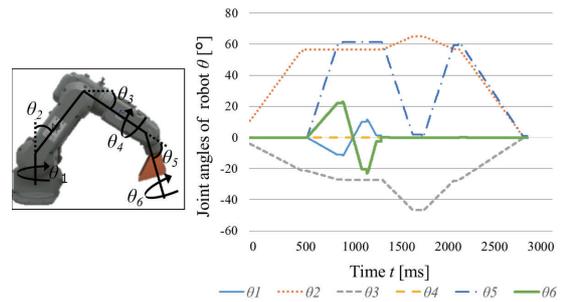


Fig. 9 The orbit of the robot.



(a) Light image (b) Center image (c) Right image

Fig. 10 Input images

アルタイムに生成可能となった。また、比較のために図10(b)に対応する出力画像、およびアームが中央センサに映り込んでいない場合の映像を図12に示す。本実験では、環境を動的に変化させてはいないため、図12(b)の画像が、生成した図12(a)の画像における真値である。この一例を比較しても分かる通り、瓦礫の多少の大きさや位置のずれがあるもの、背景が複雑な形状の物体の場合でも半隠消映像が生成可能となった。

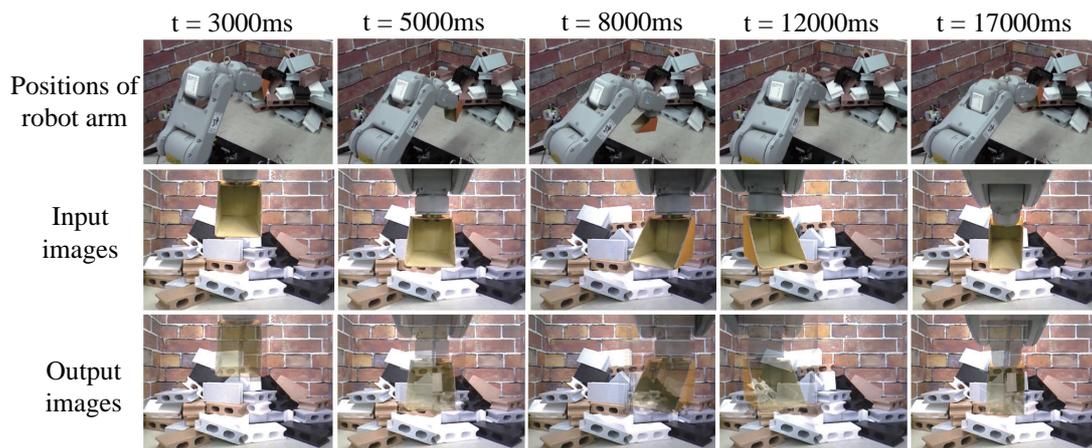


Fig. 11 The time series images when the robot copied the removal motion (図 9).

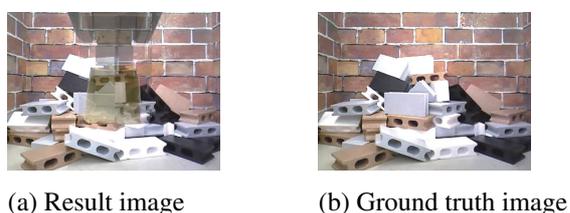


Fig. 12 Output images

5. 結 言

本稿では、複数台の RGB-D センサを使用し、中央に設置したセンサ映像中の遮蔽領域を半透明にすることで、遮蔽領域に隠蔽されている背景が透視可能な半隠消映像の生成手法を提案した。また本手法を用いることで、ロボットアームが前方の瓦礫にアクセスする際に、瓦礫が透視可能な半隠消映像がリアルタイムに生成可能なことを確認した。今後は、生成した半隠消映像を用いた際の遠隔操作性の評価を行う。

6. 謝 辞

本研究の一部は、総合科学技術・イノベーション会議により制度設計された革新的研究開発促進プログラム (ImPACT) 「タフ・ロボティクス・チャレンジ」の援助を受けた。また、マニピュレータを提供して頂いた株式会社安川電機の関係者各位に感謝致します。

参 考 文 献

- (1) M. Moteki, K. Fujino, T. Ohtsuki and T. Hashimoto: "Research on Visual Point of Operator in Remote Control of Construction Machinery", Proceedings of the 28th International Symposium on Automation and Robotics in Construction, pp. 532–537, 2010.
- (2) 古屋 弘, 栗山 楊雄, 清水 千春: "3D 画像と体感型操縦を用いた「次世代無人化施工システム」", 大林組技術研究所報, No. 76, 2012.

- (3) A. Nishiyama, M. Moteki, K. Fujino and T. Hashimoto: "Reserach on the Comparison of Operator Viewpoints between Manned and Remote Control Operation in Unmanned Construction Systems", Proceedings of the 30th International Symposium on Automation and Robotics in Construction, pp. 772–780, 2013.
- (4) 森 尚平, 一刈 良介, 柴田 史久, 木村 朝子, 田村 秀行: "隠消現実感の技術的枠組みと諸問題", 日本バーチャルリアリティ学会誌, Vol. 16, No. 2, pp. 239–250, 2011.
- (5) P. Barnum, T. Sheikh, A. Datta, and T. Kanade: "Dynamic seethroughs: Synthesizing Hidden Views of Moving Objects", Proceedings of the IEEE International Symposium on Mixed and Augmented Reality 2009, pp. 111–114, 2009.
- (6) T. Hashimoto, Y. Uematsu and H. Saito: "Generation of See-Through Baseball Movie from Multi-Camera Views", Proceedings of IEEE International Workshop on Multimedia Signal Processing 2010, pp. 432–437, 2010.
- (7) B. Avery, C. Sandor and B. H. Thomas: "Improving Spatial Perception for Augmented Reality X-Ray Vision", Proceedings of the IEEE Virtual Reality Conference 2009, pp. 79–82, 2009.
- (8) C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila: "An Augmented Reality X-Ray System Based on Visual Saliency", Proceedings of the IEEE Virtual Reality Conference 2010, pp. 27–36, 2010.
- (9) Y. Shen, F. Lu, X. Cao and H. Foroosh: "Video Completion for Perspective Camera Under Constrained Motion", Proceedings of the The 18th International Conference on Pattern Recognition, pp. 63–66, 2006.
- (10) F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa and M. A. Otaduy: "Augmented Touch without Visual Oubtrusion", Proceedings of the IEEE International Symposium on Mixed and Augmented Reality 2009, pp. 99–102, 2009.
- (11) H. Tatsumi, Y. Mae, T. Arai and K. Inoue: "Translucent View for Robot Tele-operation", Proceedings of IEEE International Workshop on Robot and Human Interactive Communication 2003, pp. 7–12, 2003.