

## 2つのRGB-D映像の統合による遮蔽物の透視

橘高 達也† 藤井 浩光† 山下 淳† 浅間 一†

† 東京大学大学院 工学系研究科 精密工学専攻

E-mail: kittaka@robot.t.u-tokyo.ac.jp

### Abstract

本論文では遮蔽物を透視した映像を生成するシステムを提案する。災害現場などの危険な現場でのロボットの遠隔操作において、オペレータへの適切な視覚情報の提示は作業効率に大きな影響を与える。コンピュータビジョンの分野で提案されている「隠消現実感」は、遮蔽物を透視したような映像を生成する技術である。提案手法ではロボットの正面およびアームにRGB-Dセンサ（通常のカメラと同様の色情報に加えて、各画素の奥行き情報も取得可能なセンサ）を取り付け、アームとともにセンサを移動させながら環境の3次元情報を取得し、透視映像を生成する。従来手法と比べて、広い視野の映像を任意の視点から見た透視映像をリアルタイムで生成可能である。マニピュレータを用いた評価実験によって、提案手法の有効性が示された。

### 1 序論

近年、工事現場や災害現場などでの危険な作業において、人の代わりに作業を行うための遠隔操作ロボットのニーズが高まっている。例えば、2014年3月に東京電力株式会社が行った福島第一原子力発電所の2号機建屋内の調査で、Warriorというロボットが用いられた。Warriorは図1のようにフェンスの反対側にアームを伸ばし、電動はさみカッターにより反対側の根元部分を切断することで、作業の障害となるフェンスを撤去する作業を行った[1]。

しかし、遠隔操作による作業は搭乗での作業と比較して作業効率が低いという課題がある。遠隔操作による作業の一例である無人化施工に関する調査では、遠隔操作による作業効率は搭乗での作業の6割程度であると言われており、作業効率低下の原因は視覚関連の問題が29%と最大であると分析されている[2]。また、オペレータにとって適切な映像を提示することが作業効率の向上に重要な役割を果たすという報告もある[3]。作業対象が遮蔽物に隠れて見えない場合、オペレータは作業対象の形や位置を映像上で認識することが困難になり、作業効率の低下につながると考えられる。例

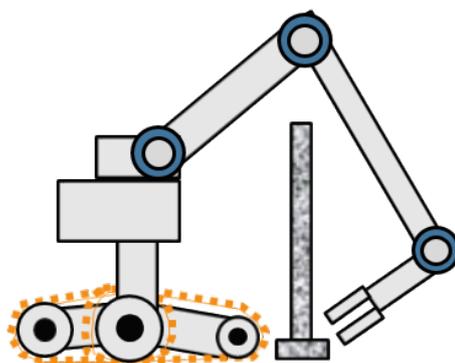


図 1: Warrior による遮蔽物越しの作業 [1] のイメージ

えば先に述べた Warrior のフェンス越しの作業を、ロボット本体の正面に取り付けられたカメラの映像を見ながら行うことは困難である。従来の遠隔操作では、ロボットに複数のカメラを搭載することで視野を広げることが一般的であるが、オペレータは複数の映像を見比べながら作業を行うこととなり、高い熟練スキルと集中力が必要となる[4]。よって、遮蔽物に隠れた領域をオペレータが見やすいように複数の映像を1つの映像に統合することは、有効であると考えられる。

コンピュータビジョンの分野では、隠消現実感 (Diminished Reality, DR) と呼ばれる、映像上で視覚的に不要な物体を隠蔽・消去、もしくは障害となる物体を透過させる技術が提唱されている[5]。この技術は、交通安全のための見通しの悪い道路における建物の透視[6]や、動画の補間[7]、Google ストリートビューの画像中からの歩行者の消去[8]といった様々な分野への応用のために広く研究されている。本稿では、DRを適用することで、映像中で遮蔽物に隠れて見えない領域を透視可能なシステムを提案する。

### 2 先行研究

以下、映像中で遮蔽物に隠れて見えない領域のことを「背景」と呼ぶ。背景の情報の取得には、背景の形や位置などについての事前知識を用いる方法[9]や、周囲の情報をもとに推定する方法（この技術は Inpainting と呼ばれる）[10]などがある。しかし、これらの方法で

は災害現場などの未知環境で正確な背景の情報を得られるとは限らない。背景が時間的に変化せず、遮蔽物が動いている場合には、過去の映像を用いて背景情報を取得する方法 [7] を用いることができる。しかし、より一般的な条件での透視のためには、複数のカメラを用いて異なる視点からの映像を統合する方法 [6, 8, 11, 12] が有効である。

先行研究の多くに共通する課題として、背景は十分遠方にあるとして平面で近似している点がある。ロボットの遠隔操作に応用することを考えた場合、ロボットは作業対象の近くで作業を行うため、平面で近似された情報では不十分であり、3次元的な情報が必要となる。

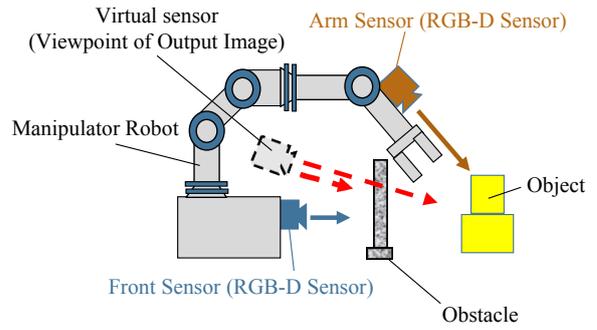
Jarusirisawadらは、多数のカメラを用いて3次元的な背景を取得した上で、カメラとは別に出力映像の視点も移動可能な遮蔽物透視システムを構築した [12]。このシステムでは出力映像を見る視点を表す仮想的なカメラを定義し、実カメラの間を自由に移動しながら映像を見ることができると、映像の奥行き情報の認識が容易となる。しかし、このシステムでは背景の3次元情報を取得するアルゴリズムの計算コストが高いという課題がある。

杉本らは、3台のRGB-Dセンサを用いてより低コストで背景の3次元情報を取得し、遮蔽物の透視を行うシステムを構築した [11]。RGB-Dセンサは、通常のカメラと同様のRGB情報に加えて、センサから画像中の各点までの奥行き情報も取得可能なセンサである。しかし、このシステムではセンサの位置は固定されていなければならないという制約がある。センサが固定されている場合、遮蔽物の位置や形状によっては、複数のセンサを用いたとしても死角が発生し透視が不可能であるようなケースが発生するという問題がある。

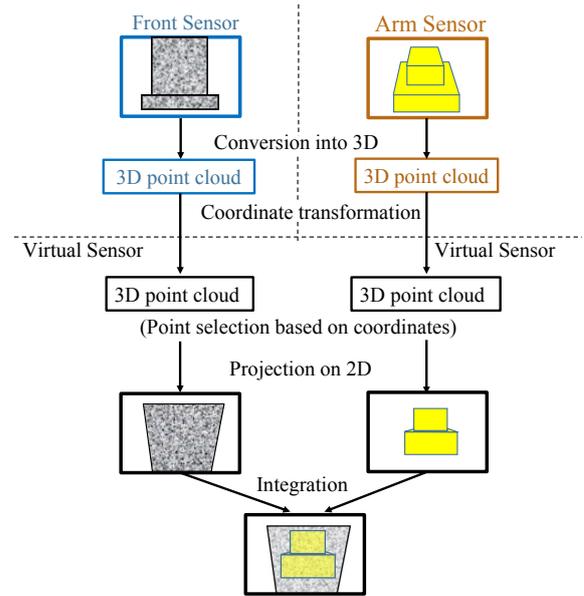
このように、従来のDRの研究にはそれぞれ長所と短所がある。これを踏まえ、本研究では、複数の移動可能なセンサを用いて、背景の3次元情報を取得し、さらに出力映像の視点も移動可能であるようなシステムを提案する。また、ロボットの遠隔操作への応用を考え、これらの機能をリアルタイムで実現することも本研究の目的とする。

### 3 提案手法

提案手法の概要を図2に示す。本研究では、災害現場などで多く用いられているアームを有するロボットを想定し、図2(a)のようにロボットの正面およびアームにRGB-Dセンサを取り付けたシステムを提案する。図2(b)に、処理の流れを示す。まず、ロボットの正面に取り付けたRGB-Dセンサ（以下、正面センサ）と、アームに取り付けたRGB-Dセンサ（以下、アームセンサ）からの映像を3次元点群の情報に変換する。次に、



(a) システムの概念図



(b) 処理の流れ

図2: 提案手法の概要

3次元点群を座標変換し、空間中の任意の位置に設定した仮想的なセンサ（以下、仮想センサ）から見た座標系へと変換する。そして、3次元点群を2次元の出力画像へと投影する。最後に、正面センサ、アームセンサから生成した映像をアルファブレンディングによって統合することで、遮蔽物を透視したように見える映像を提示できる。

次節以降で、本手法の構成要素である3次元点群化、座標変換、投影、統合について詳しく説明する。

#### 3.1 2次元画像の3次元点群化

RGB-Dセンサの内部パラメータと、センサから得られる距離情報を利用して、3次元点群の情報を生成する。空間中のある点 $\mathbf{p}$ が、ピクセル単位で表された画像上の2次元同次座標系 $\Sigma_M$ （以下、画像座標系）で ${}^M\mathbf{p} = (u, v, 1)^T$ の位置に映っているとする。この点の、mm単位で表されたセンサを中心とする3次元同次座標系 $\Sigma_S$ （以下、センサ座標系）における座標 ${}^S\mathbf{p} = (x, y, z, 1)^T$ を求める。事前のキャリブレーションによって求めた内部パラメータとRGB-Dセンサから

得られる距離情報  $z$  を用いて  ${}^M\mathbf{p}$  と  ${}^S\mathbf{p}$  の関係を求め、未知数である  ${}^S\mathbf{p}$  の要素  $x, y$  を求めることができる。

### 3.2 仮想センサ座標系への座標変換

回転行列と並進ベクトルを用いて、正面センサおよびアームセンサ座標系から、仮想センサ座標系へと座標変換する。

座標系  $\Sigma_S$  から座標系  $\Sigma_{S'}$  へと変換するためには、式 (1) のように、行列  ${}^{S'}\mathbf{R}_S$  とベクトル  ${}^{S'}\mathbf{t}_{S \rightarrow S'}$  が必要である。

$${}^{S'}\mathbf{p} = \begin{bmatrix} {}^{S'}\mathbf{R}_S & {}^{S'}\mathbf{t}_{S \rightarrow S'} \\ \mathbf{0}^T & 1 \end{bmatrix} {}^S\mathbf{p} \quad (1)$$

${}^{S'}\mathbf{R}_S$  は、座標系  $\Sigma_S$  から  $\Sigma_{S'}$  への回転を表す  $3 \times 3$  の回転行列である。 ${}^{S'}\mathbf{t}_{S \rightarrow S'}$  は、座標系  $\Sigma_{S'}$  で見た座標系  $\Sigma_S$  の原点の位置を表す、 $3 \times 1$  の並進ベクトルである。

本稿では、ロボットに固定された世界座標系  $\Sigma_W$  を用いて、任意の座標系  $\Sigma_S$  の位置・姿勢を、回転行列  ${}^S\mathbf{R}_W$  と並進ベクトル  ${}^S\mathbf{t}_{W \rightarrow S}$  で表現する。すなわち、 ${}^S\mathbf{R}_W, {}^S\mathbf{t}_{W \rightarrow S}$  と  ${}^{S'}\mathbf{R}_W, {}^{S'}\mathbf{t}_{W \rightarrow S'}$  を用いて  ${}^{S'}\mathbf{R}_S, {}^{S'}\mathbf{t}_{S \rightarrow S'}$  を求めることになる。

提案手法では、変換後の座標系  $\Sigma_{S'}$  は仮想センサ座標系を表し、その位置は空間の任意の位置に設定できるため、 ${}^{S'}\mathbf{R}_W$  と  ${}^{S'}\mathbf{t}_{W \rightarrow S'}$  も任意に設定できることになる。従って本節の目標は、 ${}^S\mathbf{R}_W$  と  ${}^S\mathbf{t}_{W \rightarrow S}$  を求めることである。提案手法では2つの RGB-D センサを用いるため、それぞれのセンサについて  ${}^S\mathbf{R}_W, {}^S\mathbf{t}_{W \rightarrow S}$  を求める必要がある。

まず、 $\Sigma_S$  は正面センサ座標系を表すとして説明する。正面センサは、ロボット本体に固定されており、ロボットに対する位置・姿勢は変化しないと仮定する。よって正面センサ座標系から仮想センサ座標系への座標変換は、 ${}^S\mathbf{R}_W, {}^S\mathbf{t}_{W \rightarrow S}$  を事前のキャリブレーションで求めておくことで達成される。

次に、 $\Sigma_S$  はアームセンサ座標系を表すとして説明する。アームセンサは、ロボットアームの運動に伴って空間中での位置・姿勢が変化するため、 ${}^S\mathbf{R}_W, {}^S\mathbf{t}_{W \rightarrow S}$  の値は常に更新する必要がある。提案手法では、ある基準となるロボットアームの姿勢（以下、基準姿勢。図 3 (a) 参照）を定め、そのときのアームセンサの位置・姿勢  ${}^{S_0}\mathbf{R}_W, {}^{S_0}\mathbf{t}_{W \rightarrow S_0}$  をあらかじめ求めておく。また、アームセンサはロボットアームに固定されており、ロボットアームの手先座標系  $\Sigma_H$  からアームセンサ座標系  $\Sigma_S$  への回転行列  $\mathbf{R}$  および並進ベクトル  $\mathbf{t}$  は一定であると仮定し、これもあらかじめ求めておく。基準姿勢のときの手先座標系が  $\Sigma_{H_0}$ 、ロボットアームの手先移動後の新しい手先座標系が  $\Sigma_H$  で表されるとする（図 3 (b)）。ロボットアームの順運動学を解くことで、手先の移動を表す回転行列  ${}^H\mathbf{R}_{H_0}$ 、手先の移動を表す

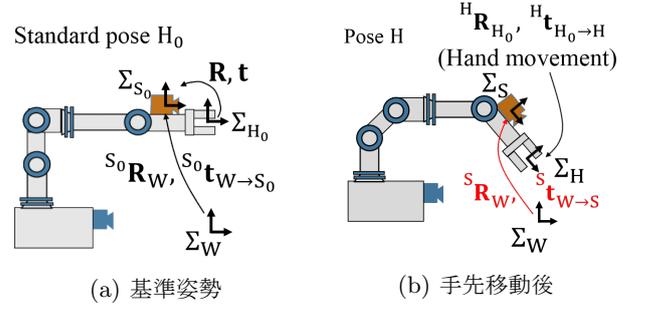


図 3: アームセンサの位置・姿勢の更新

並進ベクトル  ${}^H\mathbf{t}_{H_0 \rightarrow H}$  を求めることができる。このとき、新しいアームセンサの位置・姿勢  ${}^S\mathbf{R}_W, {}^S\mathbf{t}_{W \rightarrow S}$  は、 ${}^{S_0}\mathbf{R}_W, {}^{S_0}\mathbf{t}_{W \rightarrow S_0}, {}^H\mathbf{R}_{H_0}, {}^H\mathbf{t}_{H_0 \rightarrow H}, \mathbf{R}, \mathbf{t}$  を用いて計算可能である。

### 3.3 出力画像への投影

仮想センサ座標系で表された点群を、仮想センサの位置・姿勢から見たときの画像へと変換する。仮想センサ座標系  $\Sigma_{S'}$  で表された点  ${}^{S'}\mathbf{p} = (x', y', z', 1)^T$  が、仮想センサの画像座標系  $\Sigma_{M'}$  で  ${}^{M'}\mathbf{p} = (u', v', 1)^T$  の位置に映るとする。 ${}^{M'}\mathbf{p}$  と  ${}^{S'}\mathbf{p}$  の関係は、3.1 節で触れた  ${}^M\mathbf{p}$  と  ${}^S\mathbf{p}$  の関係と同様であるので、仮想パラメータの内部パラメータを任意に設定することで  ${}^{M'}\mathbf{p}$  を求めることができる。

出力画像へ投影する際、 $x', y', z'$  に関してある条件を満たす点のみ投影するような処理を行うことができる。これによって、遮蔽物の背後など、空間中の特定の領域のみに着目した出力映像を得ることが可能となる。 $x', y', z'$  の条件としては様々なものが考えられるが、本稿では計算を簡略化するため、 $z'$  がある区間  $[z_{\min}, z_{\max}]$  に入っている点のみを投影する処理を行うことで、任意の奥行き領域に注目した映像を生成可能とした。

### 3.4 正面センサおよびアームセンサからの出力映像の統合

正面センサおよびアームセンサから生成した画像を、アルファブレンディングによって1つの画像に統合する。正面センサおよびアームセンサから生成した画像の RGB 値  $\mathbf{I}_{\text{front}}, \mathbf{I}_{\text{arm}}$  から、出力画像の RGB 値  $\mathbf{I}_{\text{out}}$  を式 (2) のように求める。

$$\mathbf{I}_{\text{out}} = \alpha \mathbf{I}_{\text{front}} + (1 - \alpha) \mathbf{I}_{\text{arm}} \quad (2)$$

ここで  $\alpha$  は  $0 \leq \alpha \leq 1$  を満たす不透明度のパラメータである。なお、正面センサ、アームセンサのうち一方の画像からのみ画素が得られる領域はその画素の色が出力され、正面センサ、アームセンサのいずれの画像からも画素が得られない領域は黒色として出力される。

## 4 実機による検証および定量評価

### 4.1 実験環境

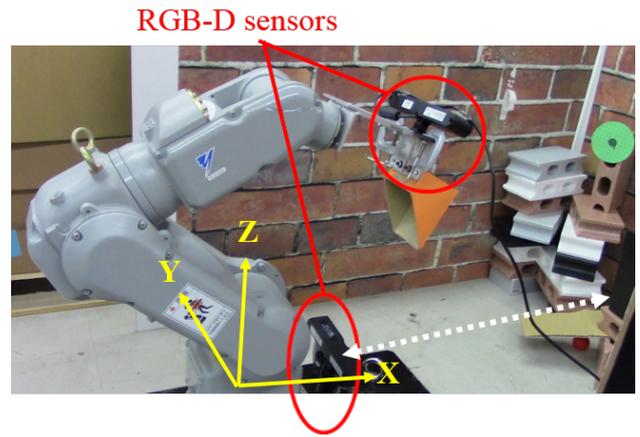
提案手法の有効性を検証するために実機実験を行った。実験環境を図4に示す。本実験では、図4(a)中に示す2台のRGB-Dセンサ(ASUS: Xtion Pro Live)とマニピュレータ(YASKAWA: MOTOMAN-HP3J)を用いた。実験環境には図4(b)中に示すように、遮蔽物、透視する対象である直径0.05 mの小球、出力画像の評価の目印に用いる直径0.1 mの円盤3つがある。図4(b)では小球は1個しか見えないが、図4(c)で確認できるように、遮蔽物に隠れた位置に、全部で3個の小球が置かれている。正面センサから遮蔽物までの距離(図4(a)中に点線矢印で示した)は約0.7 mであり、正面センサから実験環境の最奥にある壁までの距離(図4(b)中に点線矢印で示した)は約1.0 mである。マニピュレータの関節角の情報の取得レートは30 Hzであり、RGB-Dセンサの取得映像のサイズは640×480 pixel、フレームレートは30 fpsとした。マニピュレータの手先は、図4(a)中に示したマニピュレータの根元を基準とした座標系でX, Y, Zそれぞれの方向に0.15 m程度動いてから元の位置に戻るような、事前に与えた運動を行った。マニピュレータの手先の軌跡を図5に示す。図5(a)は手先の座標の時系列データである。同様に図5(b)は手先の角度の時系列データである。ここで手先の角度はZ-Y-Xオイラー角で表した。すなわち任意の姿勢を、手先がX軸に並行な状態からZ軸まわりに $\alpha$ だけ回転させ、その回転後の座標系のY軸まわりに $\beta$ だけ回転させ、さらにその回転後の座標系のX軸まわりに $\gamma$ だけ回転した姿勢として表現している。

仮想センサの内部パラメータは正面センサと同じ値を使用した。また、アルファブレンディングの不透明度 $\alpha$ は0.5に設定した。

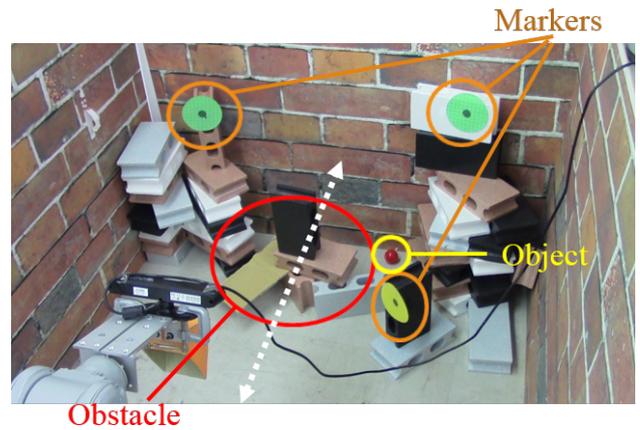
### 4.2 出力結果

実験の結果、得られた出力画像を図6に示す。図6(a)は、ある瞬間における正面センサおよびアームセンサからの入力画像である。正面センサおよびアームセンサからの画像では、小球はそれぞれ1個しか見えない。図6(b)は、この瞬間における仮想センサの出力画像を、仮想センサの位置を変えて取得したものである。正面センサ、アームセンサそれぞれ単独では見えなかった小球(図中に円で囲んで示した)が透視できていることが分かる。また、図6(c)は別の瞬間における仮想センサの出力画像である。このとき、アームセンサからは奥にある2つの小球が見えるため、仮想センサを正面センサの位置に置いた場合、3つの小球全てを1つの画面で確認できる。

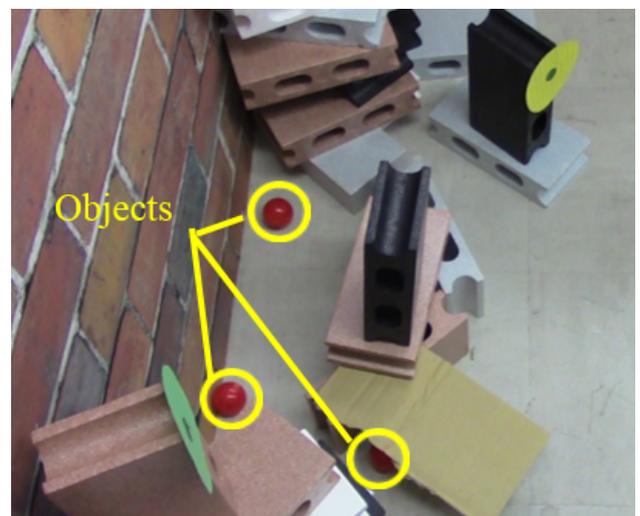
また、仮想センサの位置・姿勢を移動させた場合の出



(a) マニピュレータおよびRGB-Dセンサ



(b) 遮蔽物, 小球, 目印の円盤

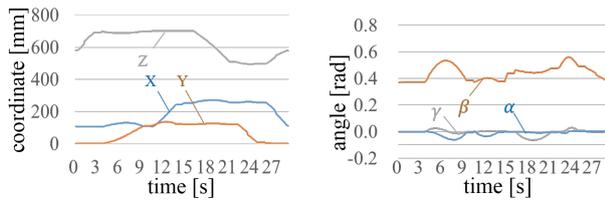


(c) 別視点から見た小球

図4: 実験環境

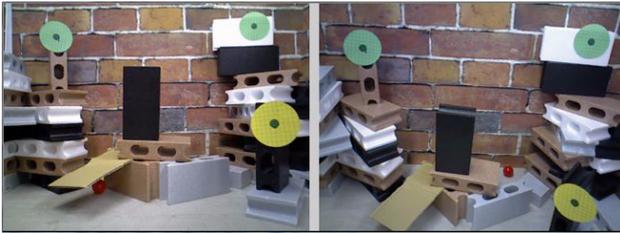
力画像の例を図7に示す。仮想センサの位置・姿勢は、マニピュレータの操作中でも任意に移動することができ、様々な視点から環境を見ることが可能である。

さらに、3.3節で述べた、特定の奥行き領域に注目して映像を生成した結果を図8に示す。ここでは、仮想センサを正面センサの位置に置き、センサからの奥

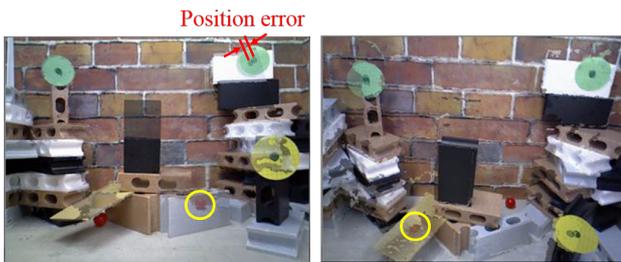


(a) 手先の位置  $X, Y, Z$  (b) 手先の姿勢  $\alpha, \beta, \gamma$

図 5: マニピュレータの手先の軌跡



(a) 正面センサ (左), アームセンサ (右) の入力画像



(b) (a) で仮想センサを正面センサの位置に置いた場合 (左), アームセンサの位置に置いた場合 (右)



(c) 別の瞬間に仮想センサを正面センサの位置に置いた場合 (左), アームセンサの位置に置いた場合 (右)

図 6: 実験結果

行きが 0.8 m 以上である領域を表示しており、遮蔽物の奥にある小球がはっきりと確認できる。

#### 4.3 評価

出力映像のフレームレートは、運動全体における平均が 5.7 fps、標準偏差が 0.19 fps となった。

また、目印の円盤の中心位置のズレ (図 6 (b) 左の図中に例を示す) を基準に、映像統合時の画像のズレを



図 7: 仮想センサを移動した場合の出力画像



図 8: 図 6 (a) で正面センサからの奥行きが 0.8 m 以上である領域の表示

目視で評価した。今回の実験における運動全体において、センサから約 0.6 m 先で 0.05 m (円盤の半径) 以内のズレに収まっていることが確認できた。これは、出力画像上では約 15 pixel 以内のズレに相当する。

また、色の情報が得られず黒く塗りつぶされた画素の割合を評価した。出力画像の全画素数に対する、RGB 情報が得られなかった画素数の割合を、図 6 (a) 右のアームセンサの位置に仮想センサを固定した場合で評価した。結果は、平均 17.7 %、標準偏差 6.4 %となった。すなわち、時間とともに実際のセンサの位置と仮想センサの位置が離れていったにも関わらず、視野の約 82 %の情報が得られたといえる。

#### 5 結論

本研究では、ロボットの遠隔操作において 1 つの映像中で作業対象を把握できるようにすることを想定した、遮蔽物透視システムを提案した。提案手法により、

背景の3次元情報を取得すること、センサが移動する場合でも透視映像を生成すること、実際のセンサとは異なる視点からも環境を見ることが可能となった。また、実機による実験によって提案手法の有効性を確認し、本システムがリアルタイムで実現可能であることも示した。

今後の展望としては、以下の3点が考えられる。

1. 映像統合時の画像のズレの軽減
2. 仮想センサ移動時の、画素の抜けの軽減
3. センサ台数を任意の数に増やした場合への拡張

## 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、本研究で使用したマニピュレータは株式会社安川電機からご提供いただいた。

## 参考文献

- [1] 東京電力株式会社: “2号機原子炉建屋内オペレーティングフロアフェンス撤去およびコアサンプル採取について”, [http://www.meti.go.jp/earthquake/nuclear/pdf/20140327\\_02/140327\\_02\\_020.pdf](http://www.meti.go.jp/earthquake/nuclear/pdf/20140327_02/140327_02_020.pdf) (2014), 閲覧日 2016/03/16
- [2] 山口 崇, 石松 豊, 山元 弘: “無人化施工のマンマシンインターフェイスに関する調査”, 平成16年度建設施工と建設機械シンポジウム論文集, Vol. 18, pp. 145–148 (2005)
- [3] Masaharu Moteki, Kenichi Fujino, Takashi Ohtsuki, and Tsuyoshi 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)
- [4] Akihiko Nishiyama, Masaharu Moteki, Kenichi Fujino and Takeshi Hashimoto: “Research 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)
- [5] 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と諸問題: 現実世界に実在する物体を視覚的に隠蔽・消去・透視する技術について”, 日本バーチャルリアリティ学会論文誌, Vol. 16, No. 2, pp. 239–250 (2011)
- [6] Peter Barnum, Yaser Sheikh, Ankur Datta and Takeo Kanade: “Dynamic Seethroughs: Synthesizing Hidden Views of Moving Objects”, *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality*, pp. 111–114 (2009)
- [7] Yuping Shen, Fei Lu, Xiaochun Cao and Hassan Foroosh: “Video Completion for Perspective Camera under Constrained Motion”, *Proceedings of the 18th International Conference on Pattern Recognition*, Vol. 3, pp. 63–66 (2006)
- [8] Arturo Flores and Serge Belongie: “Removing Pedestrians from Google Street View Images”, *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 53–58 (2010)
- [9] Francesco Cosco, Carlos Garre, Fabio Bruno, Maurizio Muzzupappa and Miguel A. Otaduy: “Augmented Touch without Visual Obtrusion”, *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality*, pp. 99–102 (2009)
- [10] Jan Herling and Wolfgang Broll: “Pixmix: A Real-time Approach to High-quality Diminished Reality”, *Proceedings of the 11th IEEE International Symposium on Mixed and Augmented Reality*, pp. 141–150 (2012)
- [11] Kazuya Sugimoto, Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: “Half Diminished Reality Image Using Three RGB-D Sensors for Remote Control Robots”, *Proceedings of the 12th IEEE International Symposium on Safety, Security, and Rescue Robotics*, 43, pp. 1–6 (2014)
- [12] Songkran Jarusirisawad, Takahide Hosokawa and Hideo Saito: “Diminished Reality Using Plane-sweep Algorithm with Weakly-calibrated Cameras”, *Progress in Informatics*, No. 7, pp. 11–20 (2010)