

遠隔操作ロボットのための複数 RGB-D センサを用いた 半隠消映像のオンライン生成*

藤井浩光** 杉本和也** 山下 淳*** 浅間 一***

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

Hiromitsu FUJII, Kazuya SUGIMOTO, Atsushi YAMASHITA and Hajime ASAMA

This paper presents a methodology to compose half-diminished reality images for operating remote control robots. At sites for disaster response, robots are desired to achieve various tasks. However, operators have problems concerned with camera images shown to them for controlling robots. For example, operators have to understand the environment by comparing many display images from multiple cameras, because the robot arm itself occludes target work objects in main camera image. *Half-Diminished Reality* technique is used for seeing through foreground objects and viewing occluded backgrounds. We have applied the technique to remote operating. In this paper, a fast algorithm to estimate the background image is proposed. Furthermore, an online half-diminished viewing system for remote control is constructed. In the experiment, we confirmed the validity of proposed method with three RGB-D sensors and a robot arm: The proposed method could display online half-diminished reality images to the operator to see through the target work objects occluded by the robot arm.

Key words: half-diminished reality, multiple RGB-D sensors, teleoperation, remote control system

1. 序 論

災害対応の現場など人が立ち入ることが困難な危険環境においては、二次災害を防ぐためにもロボットの導入が強く求められている。近年では、災害現場から離れた遠隔地からオペレータが操縦する遠隔操作型のロボットが現場に導入され、大きな実績を上げてきた¹⁾²⁾。特に、災害現場における瓦礫や土砂の撤去作業、調査対象までの動線確保における障害物の除去作業などにおいては、各種エンドエフェクタを搭載したアームを有するロボットが広く用いられている。

遠隔地からロボットを操作する際に、オペレータに提示する映像は極めて重要である。例えば山口らは、遠隔操作の作業効率に影響を与える主な要因として視覚関連の問題が29%で最大であることを指摘している³⁾。さらに、アームを有するロボットの遠隔操作に関する視覚関連の問題点としては、オペレータに提示する映像中でロボット自体のアームが作業対象物を遮蔽すること(図1)が、作業効率低下を招く原因であると報告されている⁴⁾⁵⁾。今後、遠隔操作ロボットを用いてより多くのタスクを実現する上で、オペレータに提示する映像中のロボット自体による遮蔽の問題を解決する技術が必要である。

無人化施工の現場では、従来からカメラ車と呼ばれるカメラを搭載した移動体を環境に配置し、オペレータの要求に応じた視点からの映像を提供する方式がとられてきた。しかし、災害対応の現場では複数台のカメラ車を適切な位置に配備することは困難であるため、カメラ配置が限定された条件下でロボットの遠隔操作に有効な映像を提示する必要がある。ロボットの遠隔操作でオペレータに提示する映像の視点に関する研究⁶⁾⁷⁾に

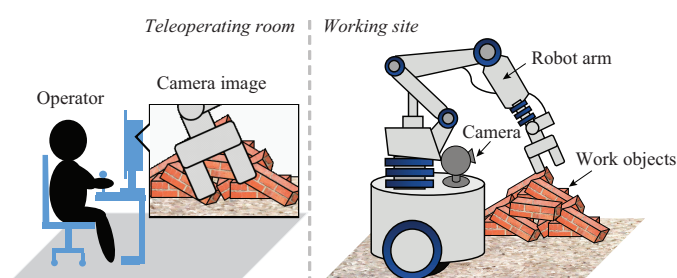


Fig. 1 Schematic view of general teleoperation work: A serious problem is that the robot itself hides target work objects from camera image

ついては、例えばロボットに搭載した複数カメラからの多視点映像をそれぞれ複数のモニタに提示する方式⁶⁾が提案されている。しかし、複数の映像を見比べながらの作業には高度な操作技術が要求され、また作業中には高い集中力が必要とされることからオペレータへの負担が大きいと報告されている⁸⁾。

半隠消映像生成手法(Half Diminished-Reality Imaging)は、複数のカメラ映像を1つに集約する際に、対象物手前の障害物を透過させることで提示映像中から除去する技術であり、エンターテインメントや放映技術などに応用されている⁹⁾。ロボット分野では、例えばTatsumiらが多足歩行型ロボットを遠隔操作するための映像生成に適用しており⁴⁾、各足ごとに搭載された複数カメラからの多視点映像を用いて、周囲の障害物に遮蔽されている背景が透視可能な半隠消映像を生成する手法を提案している。しかし、障害物と背景までの距離が十分に大きな環境を想定しているため背景の環境を2次元平面と仮定している。人の踏み込めないような災害対応現場では周囲環境の形状は未知であり、かつ作業時には対象物に接近する必要があるため、環境を平面と仮定することは困難である。

我々の研究グループでは、ロボットに搭載した複数の RGB-D

* 原稿受付 平成 27 年 5 月 11 日

掲載決定 平成 27 年 7 月 30 日

** 学生会員 東京大学大学院 (東京都文京区本郷 7-3-1)

*** 正 会 員 東京大学大学院

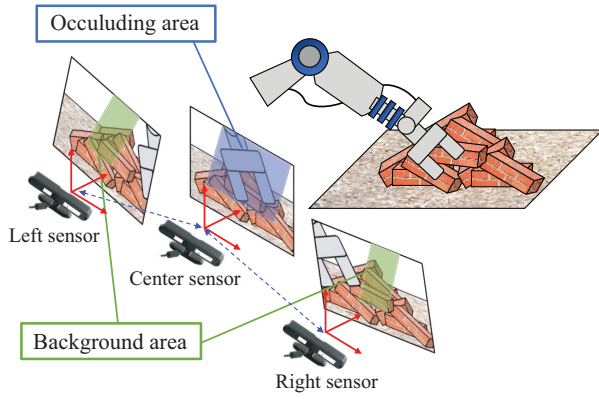


Fig. 2 RGB-D sensor arrangement and camera images obtained from each sensor

センサから多視点の距離情報付き映像を取得し、半隠消映像を生成する手法を提案してきた¹⁰⁾。本研究では、複数の RGB-D センサ映像を用いて半隠消映像を生成するための高速な映像生成アルゴリズムを提案し、オンラインでの操作に適用可能な映像提示システムを構築する。

2. 半隠消映像生成手法

2.1 半隠消映像生成手法の概要

本研究では、距離情報付きテクスチャ映像を取得可能な RGB-D センサを複数台ロボットに搭載し、ステレオ計測の原理に基づき半隠消映像を生成する。本研究における複数の RGB-D センサを用いた計測系を図 2 に示す。搭載する複数の RGB-D センサは、オペレータへの提示映像において基準となる視点を持つセンサ（中央センサ）1 台と、中央センサと異なる視点からのセンサ群（周囲センサ）から構成される。提案手法では、1 台の周囲センサのみでは取得できない死角の補完を考慮するため 2 台の周囲センサを用いる。以降は図 2 に示す通り、説明のため中央センサを挟んでロボットの正面方向に左右に位置する周囲センサを、それぞれ左センサおよび右センサと呼ぶこととする。

提案手法の処理の概要を図 3 に示す。ここで、中央センサ映像中で作業対象物などの背景がロボット自体に遮蔽されて見えない領域を遮蔽領域（Occluding area）とする。また、周囲センサ映像からは取得可能な遮蔽領域に対応する背景のテクスチャ情報を含む領域を背景領域（Background area）とする。半隠消映像の生成手法は以下の 3 つの処理から構成される。

- (1) 遮蔽領域の算出
- (2) 背景領域の推定
- (3) 遮蔽領域と背景領域の合成

(1) の遮蔽領域はロボットの関節角を取得することで中央センサ映像から算出し、(2) の背景領域は周囲センサから得られる距離情報付きテクスチャ映像とステレオの原理を用いて周囲センサ映像から推定する。(3) では、(1)、(2) でそれぞれ得られた遮蔽領域と背景領域を、中央センサ映像上で合成する。以下、詳細に説明する。

2.2 ロボットの形状モデルを用いた遮蔽領域の算出

遮蔽領域は中央センサ映像中で背景がロボット自体に遮蔽されている領域であり、最終的にこの遮蔽領域を透過表現することで背景を透視可能とする。本研究では、ロボットの形状および RGB-D センサの取り付け位置は既知であるため、ロボット

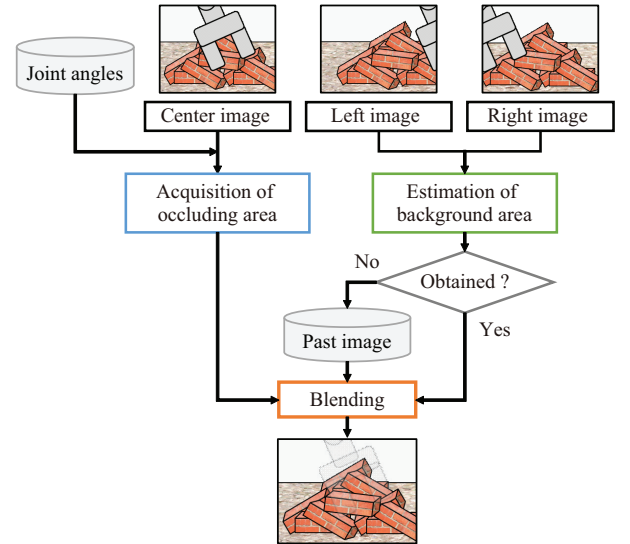


Fig. 3 Processing flow of proposed method: processing flow to compose half-diminished image from output of three RGB-D sensors

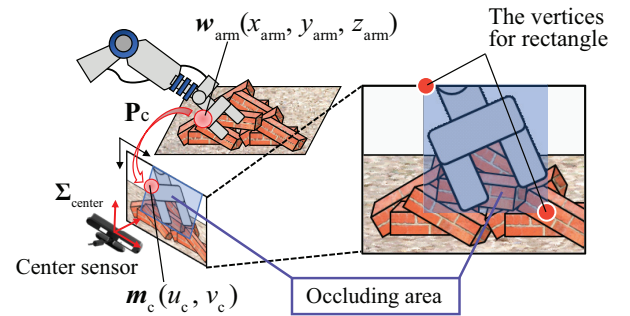


Fig. 4 Calculation of occluding area from center image

の関節角を随時取得することで、順運動学計算および透視投影変換によりアームの中央センサ映像中における位置を求めることができる。具体的には、同時表現で表したロボットアーム上の 3 次元位置 $\mathbf{w}_{\text{arm}} = [x_{\text{arm}} \ y_{\text{arm}} \ z_{\text{arm}} \ 1]^T$ と中央センサ映像中の同次座標 $\mathbf{m}_c = [u_c \ v_c \ 1]^T$ の関係は、以下のように表される。

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

ここで、透視投影行列 $\mathbf{P}_c$ はカメラの内部パラメータおよび世界座標系に対するセンサの位置・姿勢を表す行列であり、事前にカメラキャリブレーションを行うことで求める。記号 $\simeq$ は定数倍による同値を表す。

遮蔽領域の算出処理の概要を図 4 に示す。先行研究⁹⁾¹¹⁾では視野を遮蔽する物体の輪郭を正確に求めるよりも、遮蔽物を包含する大まかな領域に対して処理を行う方が、半隠消映像を生成した際に発生する合成画像に特有の不自然さが緩和されるという報告がなされている。本研究でも、アームの輪郭を厳密に求めるのではなく、形状モデルを用いてアームを直方体の集合で近似し、式 (1) に従い映像上に投影されるアームの各頂点を包含する最小の矩形領域を遮蔽領域として求める。さらに、この手法はアームの輪郭を厳密に求めるより計算量の観点からも有利であり、ロボットの遠隔操作においてオンラインの画像生成を行う上で適している。

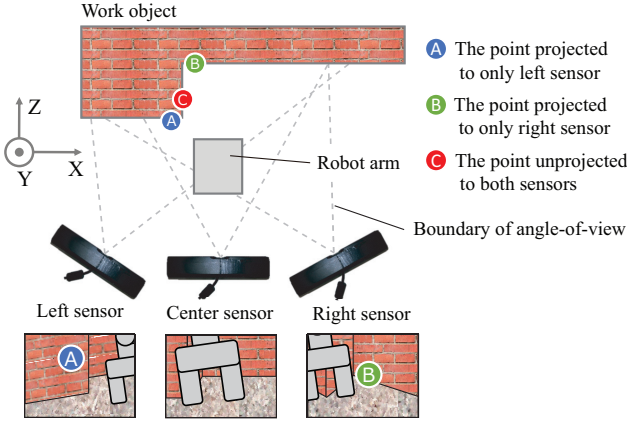


Fig. 5 Schematic view of dead area against each sensor

2.3 ステレオの原理に基づく背景領域の推定

2.3.1 距離情報を用いた周囲センサ画素の評価

中央センサ映像からは見えない遮蔽領域の背後にあるテクスチャ情報を、周囲センサからステレオ計測の原理を用いて推定する。例えば、中央センサと左センサを、互いに光軸が平行かつ高さが等しい平行ステレオシステムとする。その時、計測対象である背景の3次元座標 $\mathbf{w}_{\text{back}} = [x_{\text{back}} \ y_{\text{back}} \ z_{\text{back}}]^T$ と、 $\mathbf{w}_{\text{back}}$ が中央センサ映像の遮蔽領域中で結ぶ像の画像座標 $\mathbf{m}_c = [u_c \ v_c]^T$ 、および $\mathbf{w}_{\text{back}}$ が左センサ映像の背景領域中で結ぶ像の画像座標 $\mathbf{m}_\ell = [u_\ell \ v_\ell]^T$ には以下の関係が成立する。

$$z_{\text{back}} - \frac{b_{\text{LC}} \cdot f}{|u_\ell - u_c|} = 0, \quad (2)$$

ここで、 f は全センサで共通の焦点距離の値であり、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_{LC} は中央センサと左センサのセンサ間距離である。式(2)は中央センサと右センサに関しても同様に成り立つ式であるが、いずれの場合も中央センサと周囲センサは互いに光軸が平行であり高さが等しい位置に設置する必要がある。ロボットの形状によってはセンサの搭載位置に制限があるため、実際には厳密な平行ステレオの关系到複数センサを配置することは困難である。そこで、本研究では画像処理によりセンサ間のずれを補正する。具体的には、センサ設置後にキャリブレーションを行いセンサ間の相対位置・姿勢を計測し、ホモグラフィ行列を求めることでエピポーラ線の平行化処理を行う。

本研究では、RGB-D センサからテクスチャ情報と同時に得られる距離情報を利用して背景領域の探索を行う。遮蔽領域中の画像座標 $\mathbf{m}_c = [u_c \ v_c]^T$ について、対応する背景のテクスチャ情報 I_{back} を左センサ映像中から求める問題を、左センサの画像座標 $\mathbf{m}_\ell = [u_\ell \ v_\ell]^T$ に関する式(2)に基づく評価値 $J_{\text{LC}}(\mathbf{m}_\ell)$ を最小とする探索問題として、以下のように定式化する。

$$J_{\text{LC}}(\mathbf{m}_\ell) = \left| z_\ell(\mathbf{m}_\ell) - \frac{b_{\text{LC}} \cdot f}{|u_\ell - u_c|} \right|, \quad (3)$$

$$\mathbf{m}_\ell^* = \arg \min_{\mathbf{m}_\ell} J_{\text{LC}}(\mathbf{m}_\ell), \quad (4)$$

$$I_{\text{back}} = I_\ell(\mathbf{m}_\ell^*), \quad (5)$$

ここで、 $I_\ell(\mathbf{m}_\ell)$ 、 $z_\ell(\mathbf{m}_\ell)$ はそれぞれ $\mathbf{m}_\ell$ における左センサのテクスチャ情報および距離情報である。両センサは平行化処理済みであり、両映像におけるエピポーラ線の高さは等しいため、探索においては $v_\ell = v_c$ となる u 軸方向の1ライン上を探索す

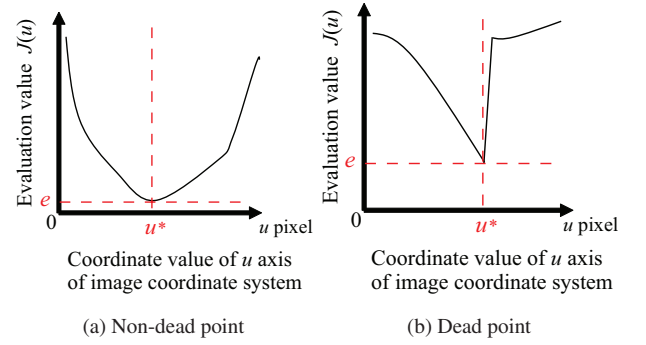


Fig. 6 Schematic view of difference in evaluation value $J(\mathbf{m})$ between non-dead point and dead point

ばよい。探索の高速化に関しては 2.3.3 項で詳説する。

2.3.2 隠れの位置の判定

実際の作業環境には複雑な形状をした物体が数多く存在し、周囲センサが1台のみでは観測できない状況が発生する。本研究では、そのような状況を隠れの位置と呼び、2台の周囲センサを用いることで隠れの位置にある背景のテクスチャ情報を取得する。隠れの位置が発生する例を図5に模式的に示す。図5に示した点Aは左センサからは観測できるが右センサからは観測できない隠れの位置にあり、同様に点Bは右センサからは観測できるが左センサからは観測できない隠れの位置にある。点Cに関してはいずれのセンサからも観測できない位置にある。

多眼ステレオで計測における隠れ位置の問題について、菅谷らが隠れ位置の判定法を提案している¹²⁾。この手法では、矩形領域同士のテクスチャパターンの相違度を評価し、最小値を探索することで視差を推定する。探索の際に相違度の最小値が閾値よりも大きい場合はマッチングに不整合が発生したと判断し、その計測対象は隠れの位置として判定する。本研究では、菅谷らの手法を参考に以下の手順で隠れの位置を判定する。

背景領域の探索においては、先述のように u 軸方向の1ライン上を探索すればよい。画像座標 $\mathbf{m}$ に関する評価値 $J(\mathbf{m})$ を画像座標系におけるエピポーラ線上の座標値 u の関数と見なすことができる。以下では必要に応じて $J(u)$ と表記する。式(3)に示したステレオの原理に基づく評価値 $J(u)$ の変動の概要を、対象点が観測可能な場合と隠れの位置にある場合について、図6(a)および図6(b)にそれぞれ示す。観測可能な場合はステレオの関係が成立するために評価値は0に近い小さな値で最小値をとる。一方で、周囲センサから隠れの位置にある場合には、センサの距離情報は背景を隠している手前の障害物までの距離を示すために、ステレオの関係が成立せずに評価値は全体的に大きな値をとり、かつ最小値付近で急激な変化をする傾向がある。本研究では、その性質を利用して隠れの位置の判定を行う。具体的には、菅谷らの手法¹²⁾でテクスチャパターンの相違度を評価する指標として提案されたものと同様の形式の以下の2つの特徴量を用いる。

(1) 評価値 $J(u)$ の最小値 e

(2) 評価値 $J(u)$ の最小値近傍での変化率 e'

ただし、本研究の距離情報を用いたピクセル単位での隠れ位置の判定においては、式(3)の評価関数はステレオ計測の原理に基づく幾何的整合度である点で異なっており、図6に示した評

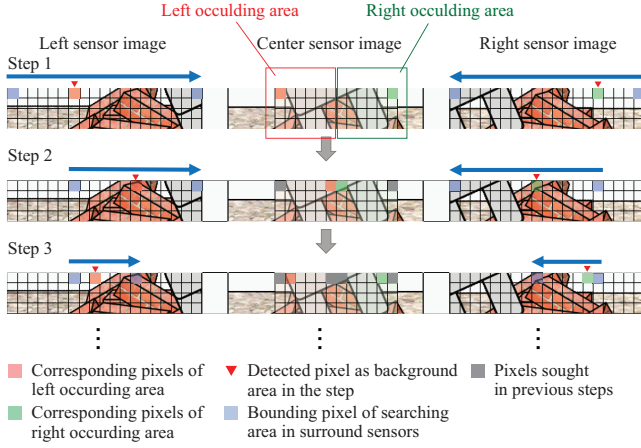


Fig. 7 Schematic view of RGB-D sensor arrangement and camera images obtained from each sensor

価関数の形状の分類に関しても、菅谷らの手法¹²⁾で提案されたテクスチャ情報の有無による細分類と比較して単純化されている。ここで、 $J(u)$ の最小値近傍での変化率 e' は、最小値を与える u 座標の値 u^* に関して以下のように定義する。

$$e' = \max(J(u^* - 1) - J(u^*), J(u^* + 1) - J(u^*)), \quad (6)$$

e' は常に正値である。 e および e' に関しては、隠れの位置の判定のための閾値をあらかじめ試行錯誤的にそれぞれ設定し、各ピクセルに対する e および e' が閾値より大きい場合を隠れの位置と判定する。

2.3.3 周囲センサの選択と探索の高速化

遠隔操作のオペレータにオンラインで映像を提示するために、本研究では背景領域の探索においてステレオ計測の幾何的な性質を利用した高速化処理を提案する。まず、隠れの位置が存在しない場合、中央センサと周囲センサ間の視差を考慮すると、中央センサ映像における遮蔽領域の左側および右側の領域は、それぞれ主に左センサおよび右センサから取得可能である。そこで、遮蔽領域を左右で2分割した左遮蔽領域および右遮蔽領域に関して、それぞれ対応する側の周囲センサから探索を開始し、隠れの位置にあると判定された場合にはもう一方の周囲センサで探索を行う。図5で示したC点のように、左右いずれのセンサでも隠れの位置と判定され観測できない場合には、中央センサから取得した過去映像中のテクスチャ情報を用いる。中央センサ映像上の遮蔽領域は、2.2節で述べた方法で既知情報から取得可能であるため、過去映像には中央センサ映像の該当ピクセルが遮蔽領域に含まれていなかった最新の時刻におけるテクスチャ情報を用いる。過去映像は中央センサ映像と同じサイズのフレームバッファとしてメモリ上に用意し、その更新は遮蔽領域以外の部分に対して毎フレーム行う。

先述のように背景領域の探索に関しては、エピソード線上の1ラインのみを探索すればよいが、本研究では以下のアルゴリズムで探索範囲を縮小更新することで探索のための計算量を削減する。アルゴリズムの概要を図7に示す。例として、左遮蔽領域で説明を行う。まず、左遮蔽領域中の左端一列目の画素を探索の参照点とする。その後、左センサ映像を用いて参照点と同じ高さにあるエピソード線上の領域を探索範囲として、基準点より右方向（画像座標 $+u$ 方向）に評価値式(3)を最小とする座標値の探索および隠れの位置の判定を行う（図7中 Step

1）。初期の探索 Step 1 においてはエピソード線上の探索範囲は左右のセンサの画像幅のサイズと等しくなる。

次に Step 2 の処理を行う場合は、左遮蔽領域の右から第1列目の画素を参照点に変え同様の探索をする。この際、中央センサ映像と左センサ映像の視差を考慮すると、左遮蔽領域中において未探索である参照点の中で最右端画素に対応する背景領域は、左センサ映像中において必ず1つ前の Step の探索結果よりも右方向（画像座標 $+u$ 方向）に存在する。そのため、Step 1 における探索結果である画素を探索の起点に設定することで探索範囲を縮小することが可能である。次の Step 3 でも同様に、探索の参照点を左遮蔽領域において未探索である最左端画素（左から第2列目）に変えて探索を行う。このように、遮蔽領域上の参照点を左右交互に変更すると同時に、直前の Step での探索結果を探索の起点とすることで背景領域の探索範囲を逐次的に狭めていき、探索処理を高速化する。探索を行う総ステップ数は遮蔽領域の列数であり、右遮蔽領域に関しても左右を入れ替えた同様のアルゴリズムで探索を行う。

2.4 遮蔽領域と背景領域の合成

遮蔽領域中の画像座標 m_c における中央センサ映像のテクスチャ情報を $I_c(m_c)$ とし、周囲センサの背景領域中から探索した対応する画像座標におけるテクスチャ情報を I_{back} とすると、中央映像上で合成される半隠消映像のテクスチャ情報 $I_{\text{hdr}}(m_c)$ は以下のようにアルファ合成される。

$$I_{\text{hdr}}(m_c) = \alpha I_c(m_c) + (1 - \alpha) I_{\text{back}}, \quad (7)$$

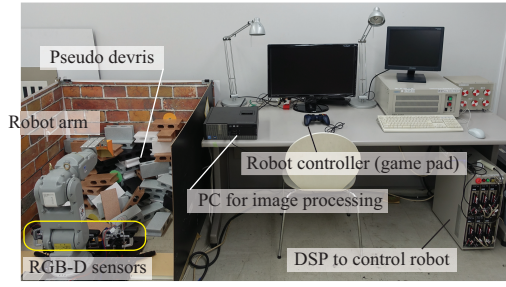
ここで、 $0 \leq \alpha < 1$ は遮蔽領域の不透明度であり、遠隔操作システムのオペレータが任意に変更可能なパラメータである。例えば $\alpha = 0$ の場合、中央センサ映像の遮蔽領域における映像合成では、中央センサ映像のテクスチャ情報は用いられず探索された背景部分のみが表示される。値域外の $\alpha = 1$ に関しては、周囲センサ映像のテクスチャ情報は用いられず中央センサ映像そのものを表示することに相当するため、半隠消映像の生成は行われない。アルファ合成に要する処理時間は背景領域の探索に要する時間と比較して小さく、式(7)からも確認できるように遮蔽領域の大きさにのみ依存し α の値によって変動しない。

3. 半隠消映像生成実験

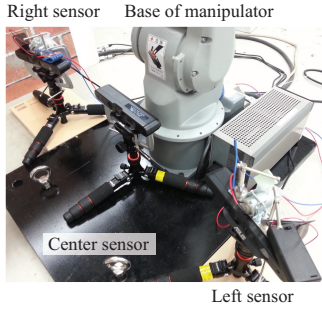
3.1 オンライン半隠消映像生成システムの構築

本研究では、第2章で提案した半隠消映像生成手法を実装し、オンラインでの半隠消映像提示が可能なシステムを構築した。構築したシステムを図8(a)に示す。システムは、3台のRGB-Dセンサと映像合成を行いオペレータに提示する映像処理用PC、ロボットアームとその位置・姿勢制御を行うためのDSPから構成される。3台のRGB-Dセンサは、図8(b)に示すように中央センサに対して左右センサをほぼ対称に配置した。本システムにおいて設定した世界座標系を図8(c)に示す。世界座標系はロボットのベース位置に固定しており、座標原点が床面上に存在しZ軸方向をロボット正面方向、Y軸をロボットのベース軸と一致させた右手座標系である。

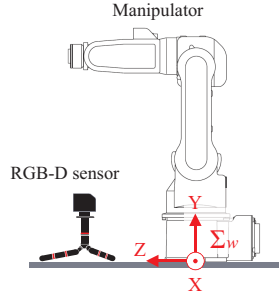
映像処理用PCにはコントローラを接続しており、オペレータはコントローラを操作することでロボットアームの位置・姿勢を変化させることが可能である。映像処理用PCとロボット制御用DSPの間ではシリアルケーブルによる通信を行い、オペレータがコントローラを用いて入力した動作指令によりロボットアームは位置姿勢を変化させる。ロボット制御用DSP



(a) Experimental environment



(b) Three RGB-D sensors



(c) World coordinate at the base of manipulator

Fig. 8 Experimental environment and equipment

ではロボットアームの関節角度を計算し、同じシリアルケーブルを介して映像処理用 PC へ送信する。映像処理用 PC には、複数の RGB-D センサから得られた RGB のテクスチャ情報と対象物までの距離情報、およびロボットアームの各関節角度情報が入力され、これらの情報を統合して処理することで半隠消映像が生成される。映像処理用 PC には、Intel 社製 Core™ i7-4770 CPU (3.40GHz), および 16GB のメモリを搭載した市販の汎用計算機を用いた。

各 RGB-D センサについては、フレームレートは 30fps, テクスチャ映像および距離映像の解像度は 320×240 とし、ロボットアームからの各関節角の出力レートは 100Hz とした。RGB-D センサとしては、ASUS 社製 Xtion Pro Live を使い、ロボットアームには安川電機社製の 6 軸のマニピュレータ MOTOMAN-HP3J を用いた。また、オペレータが用いるコントローラには市販のゲーム用パッドを用いた。



Fig. 10 A result of half-diminished reality imaging ($\alpha=0.5$)

3.2 実験環境

ロボットアームの動作中にオンラインで半隠消映像を生成する実験を行った。実験では、図 8(a) に示すような瓦礫を模したプラスチック製のブロックを作業対象物として環境に散在させた状態で作業を行った。3 台の RGB-D センサは図 8(b) に示すように作業対象物に対して十分な視差が取れるように配置した。具体的には、図 8(c) に示したロボットのベース位置に設定した世界座標系 Σ_w に対して Z 軸方向に約 150mm, Y 軸方向に約 250mm の位置に中央センサを設置し、同じく世界座標系で中央センサから X 軸方向に約 200mm 離れている位置に左センサを設置し、X 軸方向に約 200mm 離れている位置に右センサを設置した。さらに、左右センサは Y 軸周り内側（中央センサ方向）へ約 10deg ずつ回転させることで輻輳ステレオシステムとした。左右センサの回転角度については、センサ間の距離およびセンサの画角を考慮した上で、作業対象物が左右センサで撮影可能なように作業目的に応じて設定する必要がある。本研究では、実際の現場におけるロボットへの搭載を考慮して、各センサの設置および相対姿勢については精緻な位置合わせを行わず、事前のキャリブレーションを行うことでロボットと各 RGB-D センサ間、および各 RGB-D センサ同士の相対位置・姿勢を求めセンサ映像同士の平行化を行った。

隠れの位置の判定についての特徴量 e および e' に関する閾値は、事前実験によって試行錯誤的に決定し、それぞれ 50mm および 5mm/pixel と設定した。

3.3 半隠消映像の生成結果

アームを持ち上げた遮蔽領域が存在しない状態での中央センサ映像を図 9(a) に示し、ロボット前方の作業対象物へアプロー

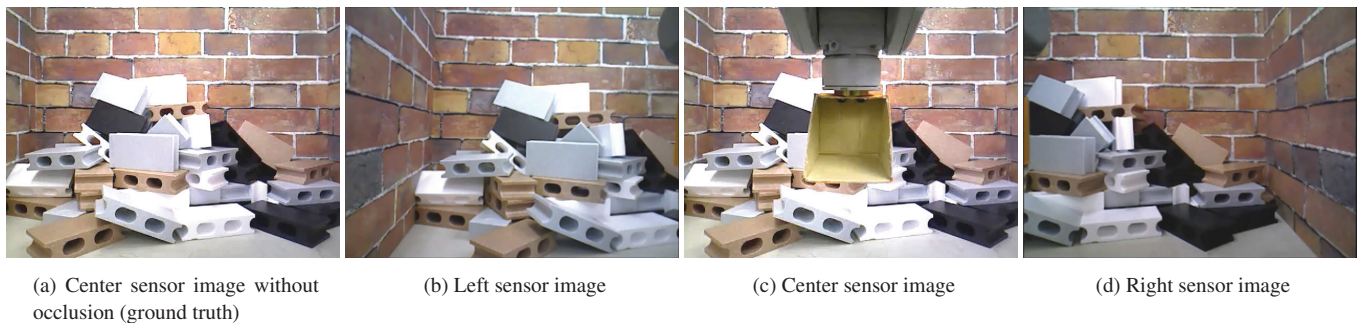


Fig. 9 An example of each camera image and ground truth of half-diminished imaging

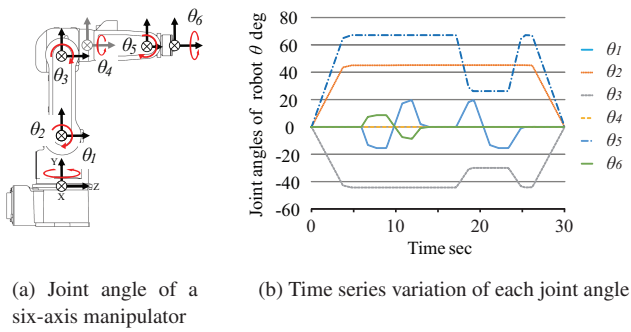


Fig. 11 Orbit of robot motion imitating digging work

チした時の RGB-D センサのテクスチャ情報を、図 9(b), (c), (d) に示す。図 9(c) から、中央センサ映像においてロボットのアーム部分に背景の作業対象物が遮蔽されていることが確認できる。一方、図 9(b), (d) から確認できるように、周囲センサからはロボットアームを視認することができない。この例のように、中央センサ映像上でロボット自体による遮蔽が生じ、異なる視点の周囲センサ映像上ではエンドエフェクタの位置が確認できない場合に、ロボットと作業対象物との相対位置を把握することは極めて困難である。

同時刻における半隠消映像の生成結果を図 10 に示す。式 (7) における合成の不透明度 α の値は 0.5 とした。図 9(a) に示した背景の真値と比較して、中央センサで遮蔽されていた背景の作業対象物が透視できていることが確認できる。

半隠消映像においては、中央センサ映像上に合成した作業対象物の位置が、遮蔽が存在しない場合に実際に投影される位置に対して正確であることが重要である。そこで、位置精度の評価値としてテクスチャパターンの類似度の指標として広く用いられる正規化相互相関係数を用いる。図 9(a) に示したようなアームによる遮蔽がない場合の映像を真値として、真値と不透明度 $\alpha = 0$ で合成した半隠消映像の遮蔽領域における正規化相互相関係数を求めた結果、図 9(c) の位置から約 2.7s 間の左右への旋回動作中における 83 フレームの平均で 0.78 と高い相関が得られた。この時、正規化相互相関係数の標準偏差は 0.07 と画質の変動は小さく、最大値は 0.87、最小値は 0.53 であっ

た。正規化相互相関係数の値が最小値をとったフレームにおいては、遮蔽領域の大きさが 36×43 pixel と極めて小さく、遮蔽された背景領域がブロック表面など画素値のダイナミックレンジが小さい平坦な部分であったことが、テクスチャパターンの相関が低下した原因であると考えられる。以上より、実際の背景映像と半隠消映像のテクスチャパターン間に高い類似性が確認できた。

3.4 連続動作中の映像生成結果

アームを有するロボットによる瓦礫の撤去作業を想定し、掘削を模した動作をロボットアームで行った。実験に使用した 6 軸マニピュレータの各関節角を図 11(a) に示し、実験で行った動作中に取得された各関節角の時間変化を図 11(b) に示す。作業時間は約 30s であった。

掘削動作中の各時刻における結果を、図 12 に示す。図 12 中の上段にはロボット頭上から撮影した動作時の姿勢を示しており、中段には中央センサ映像を示している。下段が提案手法で生成した半隠消映像であり、合成の不透明度は $\alpha = 0.5$ とした。ロボットの姿勢変化時にも適切に遮蔽領域を算出し、対応する背景領域が得られていることが確認できる。作業中で、アームによる遮蔽が中央センサ映像上で発生したのは 998 フレームであり、その際の遮蔽領域の画素数の平均は 1 フレームあたり 10,630 pixel であった。

2.3 節で述べた通り背景領域の探索では、左右の周囲センサ映像または中央センサの過去映像のいずれかが用いられている。アームによる遮蔽が生じた 998 フレームで探索された背景領域における各映像の構成比の時間変化を図 13 に示す。横軸は時間であり、縦軸は背景領域を構成する各映像の比率を百分率で示している。

例えば、約 5s から約 10s の間、および約 20s から約 25s の間で左センサ映像からの割合が増加した直後に右センサ映像からの割合が増加している。これらの時間帯には、ロボットは右に旋回しており、まず右センサ映像中の一部がアームによって一時的に遮蔽された。この時、右センサ映像から背景を取得できない右遮蔽領域について左センサ映像から探索を行い取得したため、左センサ映像からの構成比が増加した。さらに旋回を続けることで、右センサ映像の遮蔽が解消されるとともに、左

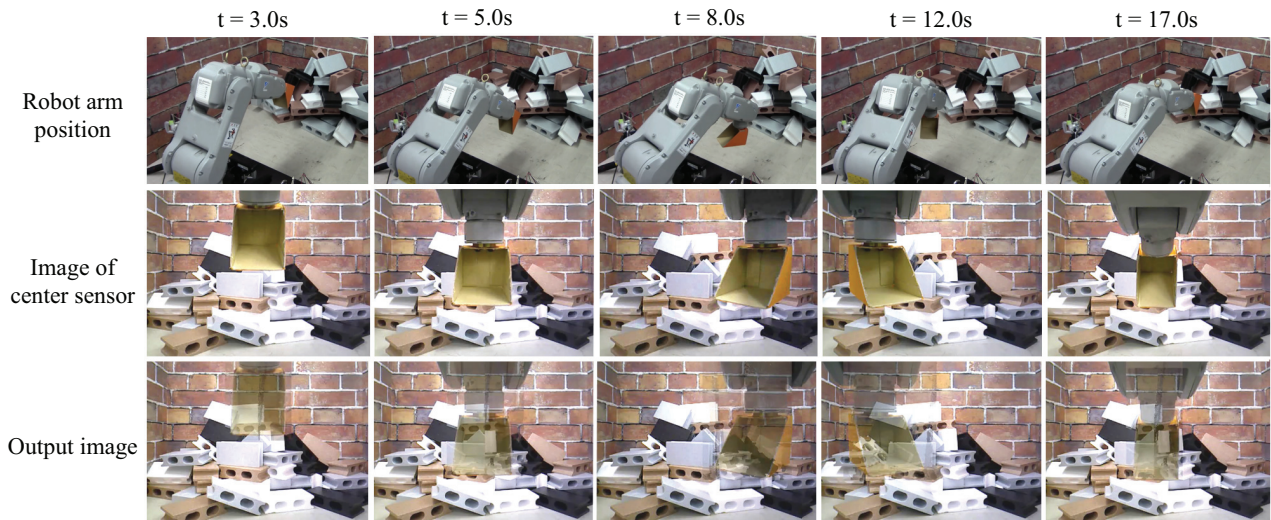


Fig. 12 Online composing results of half-diminished imaging during a digging motion ($\alpha = 0.5$)

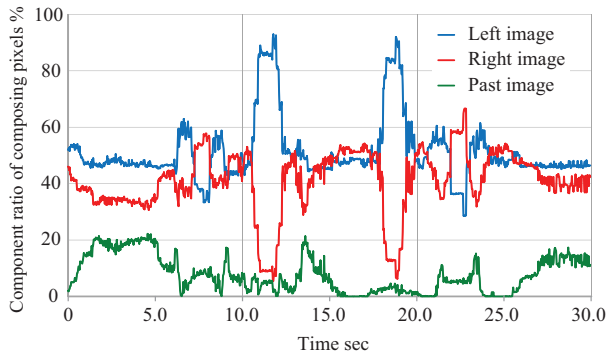


Fig. 13 Component ratio change of pixels composing background area

Table 1 Statistics of composing pixels from each sensor: The number of frames with occlusion by arm was 998

	Left image	Right image	Past image
Average ratio	51.3%	41.0%	7.7%
Maximum ratio	93.0%	66.7%	22.2%
Minimum ratio	28.6%	5.5%	0.0%

センサ映像からは観測できない左遮蔽領域が生じたため、左右センサの構成比が逆転した。

また、約 11s から約 12s の間、および約 18s から約 19s の間の時間帯には左センサ映像から用いた画素の割合が増加している。図 11(b) および図 12 で示したように、これらの時間帯でロボットは左に旋回した後に初期位置に復帰する動作を行った。例えば約 11s から約 12s の間、旋回によって遮蔽された中央センサ映像の背景を左センサから探索し取得したため左センサ映像からの割合が増加した。このように、視差の異なる 2 台の周囲センサで探索範囲を補い合うことで、異なる姿勢においても半隠消映像生成が可能となる。

時刻約 1s から約 5s の間などに過去映像の割合が増加しているのは、不規則に配置され手前にせり出した対象物にロボットが近づいた時間帯である。エンドエフェクタの手先まで至近距離から対象物が隠れることで、いずれの周囲センサからも隠れの位置となる観測できない遮蔽領域が発生したことが原因である。この例のようにアームと対象物が急接近した場合には、過去映像として作業直前の映像を用いることが可能であるため、オペレータは半隠消映像上でも作業環境を適切に把握できると考えられる。

各フレームにおける構成比の平均値、最大値、最小値を表 1 に示す。左センサ映像からの画素が占める割合の最大値が右センサ映像と比較して大きいのは、本実験で行った動作において左方向への旋回が右方向への旋回と比較して旋回量が大きく、右センサからは観測できない右遮蔽領域が中央センサ映像上の左側に多く発生したことが原因である。提案手法を適用することで、背景領域の大部分が実際に観測している周囲センサ映像から構成されており、過去映像の利用されている割合は極めて少ないことが確認できる。

図 13 で用いた遮蔽が生じたフレーム群に関して、周囲センサから各画素の背景領域を探索するのに要した処理時間の時系列変化を図 14 に示す。処理時間が増加している約 1s から約 5s の間や約 11s から約 12s の間、および約 18s から約 19s の間の時間帯などは、図 13 における過去映像の構成比が増加している時間帯や、周囲センサの一方の構成比が増加している時間

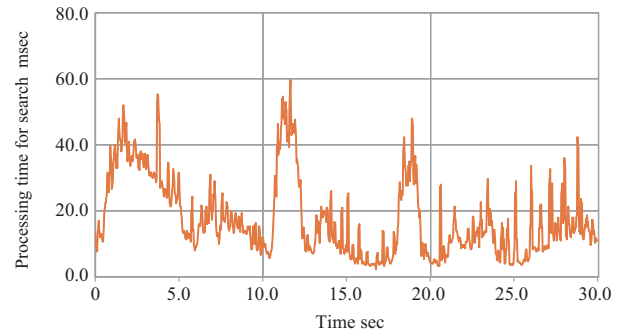


Fig. 14 Processing time for search of background area

帯と一致している。これらは周囲センサ映像上で隠れの位置が生じた場合であり、各ピクセルに対して左右センサ映像の双方で探索を行ったことが処理コストの増大の原因である。実験において背景領域の探索に要した時間は 1 フレームあたりの平均で約 17.9ms であり、最大でも約 59.7ms であった。すなわち平均で約 55.7fps のフレームレートに相当し、最も処理時間を要する場合においても約 16.8fps を実現可能である。

提案手法において探索以外に要する処理時間は微小であり、その処理時間の変動も小さい。例えば、対象物が左右センサのいずれに対しても隠れの位置にある場合は過去映像を用いた補完処理を行うが、2.3.3 項で述べた通り過去映像の更新はメモリ上にあるフレームバッファの遮蔽領域以外に対して行う処理であるため非常に高速に実行可能である。また、過去映像の更新は隠れの有無に関わらず毎フレーム行う処理であるため、隠れの位置の有無による補完処理に要する時間の増減も極めて小さい。

以上より、提案手法を適用することでロボット操作中にオンラインで半隠消映像提示が可能であることを確認した。

4. 結 論

本研究では、複数の RGB-D センサ映像を用いて半隠消映像を生成するための高速な映像生成アルゴリズムを提案し、ロボットの操作中にオンラインで半隠消映像を提示可能なシステムを構築した。提案手法によりステレオ計測の原理に基づき探索範囲を限定していくことで、ロボット自体に遮蔽された背景領域を高速に獲得することが可能である。実験では、3 台の RGB-D センサとロボットアームを用いた遠隔操作システムを構築し、ロボットの操作中にオンラインで映像提示が可能であることを確認した。

ロボットの形状によってはカメラ間に十分な視差を設けることが困難な場合や、ロボット自体以外にもフェンスのような遮蔽物が存在する場合が想定される。そのような場合、ロボットのアームなど可動部にカメラを取り付けることで十分な視差を獲得することが可能であると考えられる。カメラ位置を制御することで動的に変化する視差を用いた半隠消映像生成は、今後の興味深い課題である。

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」および JSPS 特別研究員奨励費 269039 の助成を受けた。また、株式会社安川電機の関係者各位からはマニピュレータを提供して頂いた。

参 考 文 献

- 1) Satoshi Tadokoro: Rescue Robotics Challenge, Proceedings of the 2010 IEEE Workshop on Advanced Robotics and its Social Impacts, (2010), 92.
- 2) 浅間一: 福島原発事故および廃炉対策におけるロボット技術の活用, 日本機械学会誌, **117**, (2014), 2.
- 3) 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, **18**, (2005), 145.
- 4) Haruhiko Tatsumi, Yasushi Mae, Tatsuo Arai, Kenji Inoue: Translucent View for Robot Tele-operation, Proceedings of IEEE International Workshop on Robot and Human Interactive Communication, (2003), 7.
- 5) Hironao Yamada, Tatsuya Doi: Teleoperation of Hydraulic Construction Robot Using Virtual Reality, Proceedings of the 7th JFPS International Symposium on Fluid Power, (2008), 109.
- 6) 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, **76**, (2012), 110.
- 7) Fumio Okura, Yuko Ueda, Tomokazu Sato, Naokazu Yokoya: Teleoperation of Mobile Robots by Generating Augmented Free-Viewpoint Images: Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, (2013), 665.
- 8) Akihiko Nishiyama, Masaharu Moteki, Kenichi Fujino, Takeshi 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, (2013), 772.
- 9) 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, **16**, **2**, (2011), 239.
- 10) 杉本和也, 藤井浩光, 山下淳, 浅間一: 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, 第 20 回ロボティクスシンポジウム講演予稿集, (2015), 315.
- 11) Francesco I. Cosco, Carlos Garre, Fabio Bruno, Maurizio Muzzupappa, Miguel A. Otaduy: Augmented Touch without Visual Obtrusion, Proceedings of the IEEE International Symposium on Mixed and Augmented Reality 2009, (2009), 99.
- 12) 菅谷保之, 大田友一: 隠れ処理を含まない方式と含む方式の連結統合による多眼ステレオ法; 電子情報通信学会 技術研究報告, パターン認識・メディア理解, **99**, 574, (2000), 33.