

遠隔操作ロボットのための3台のRGB-Dセンサを用いた半隠消映像の生成

○杉本和也（東京大学） 藤井浩光（東京大学） 山下淳（東京大学） 浅間一（東京大学）

1. 序論

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。

近年、災害発生時の対応作業の際には、二次災害を軽減するために、人間の代わりにロボットによる作業が期待されている。その中でも瓦礫の撤去、バルブの開閉といった作業では、図1に示すような車体にアームがついたロボットが用いられている。

ロボットを遠隔操作する際、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もとに遠隔地から操作を行う。しかし、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て作業対象物の映像が十分に得られないことが挙げられる。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し、作業対象物の位置・形状がより確認しやすくなる。しかしオペレータは多視点の映像を見比べながら作業を行う必要があるため、操作に熟練を要するばかりか、作業に集中を要するため、長時間の操作はオペレータへの負担となる。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。

近年、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・除去する技術が注目されている。特にエンターテインメントの分野において数多くの研究事例が報告されている[3, 4, 5]。文献[6]では、映像上の対象物体を完全に消しきることなく、半透明の状態にすることでその物体に隠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。そこでロボットに搭載されたカメラの映像に「半隠消現実感」を用いることで、アームによって遮られて見えない作業対象物が1つの映像上で認識可能になると考えられる。

筆者らはこれまでに2台のRGB-Dセンサとロボットアームを用いて、センサ映像上のロボットアーム部分を半透明にすることで背景が透視可能な半隠消映像の生成を提案した[7]。しかしこれまでのシステムでは、一方のセンサ映像上でロボットアームに遮蔽されている背景は、もう一方のセンサ映像から取得するしかなく、取得不可能な死角領域が存在していた。そこで本稿では使用するRGB-Dセンサの台数をもう1台増やすことで、死角となる領域にロバストなシステムの構築を提案する。

2. 提案手法

本研究では、目的とする半隠消映像を生成する際、環境の3次元計測を行う必要がある。そこで、RGB-Dセ

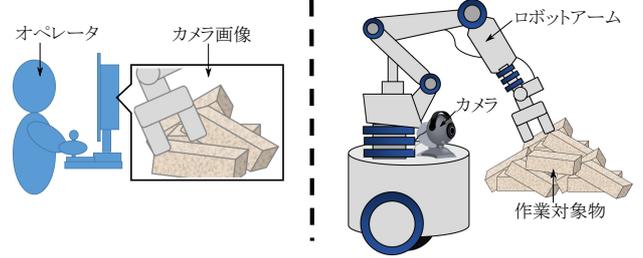


図1 ロボットを用いた遠隔操作システム

ンサを用いる。RGB-Dセンサとは、センサ1台でRGBの色情報に加えてセンサから対象物までの距離情報が取得可能なセンサである。本研究ではRGB-Dセンサを合計で3台使用する。本稿では簡単のために、各RGB-Dセンサをそれぞれ左カメラ、中央カメラ、右カメラと呼ぶ。また、各カメラから得られる映像をそれぞれ左カメラ映像、中央カメラ映像、右カメラ映像と呼ぶ。

ここで、実験環境の概念図および各カメラの位置関係を図2に示す。各カメラは必ず視差映像が得られるように設置する。さらに各カメラ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アーム領域とし、アーム領域に隠れて見えない背景物体に対応する領域を被遮蔽領域とする(図3)。

以上の環境において、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) アーム領域の取得
- (2) 被遮蔽領域の取得
- (3) 取得した映像の合成

本稿では、中央カメラ映像のアーム領域に対応する被遮蔽領域を左右のカメラ映像より取得することを考える。各処理の内容を次節から述べる。

2.1 アーム領域の取得

図2のように中央カメラのカメラ中心に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 $\mathbf{w}_{arm} = [x_{arm}, y_{arm}, z_{arm}, 1]^T$ が、中央カメラ映像の画像座標系において $\mathbf{m}_{camC} = [u_{camC}, v_{camC}, f_{camC}, 1]^T$ 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。

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

\mathbf{P} は中央カメラ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、中央カメラ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、中央カメラ映像のアーム領域がリアルタイムに取得可能となる。

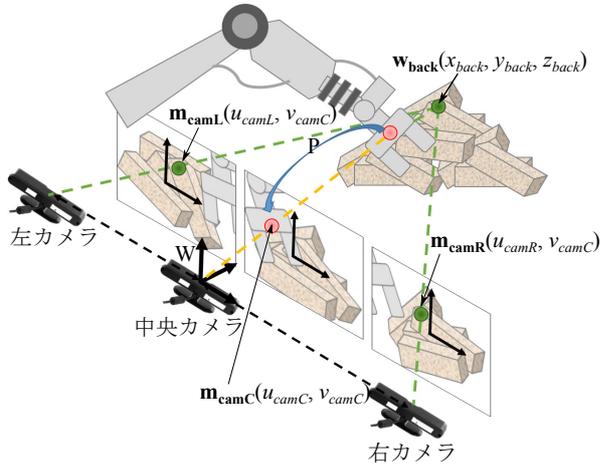


図2 実験環境の概念図

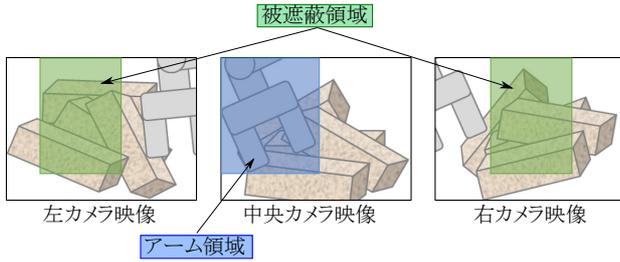


図3 カメラ映像上の各領域の定義

2.2 被遮蔽領域の取得

被遮蔽領域を左右のカメラ映像より求める。まず本研究では、2.1節で求めたアーム領域を縦に2分割し、それぞれ左アーム領域、右アーム領域と定義する。アーム領域に隠れて見えない作業対象物の表面形状から、被遮蔽領域を以下の3通りに分類できる(図4)。

- (A) 左カメラのみ取得可能な領域
- (B) 右カメラのみ取得可能な領域
- (C) どちらのカメラからも取得不可能な領域

ここで(A)～(C)の取得手法の概要を以下に示す。また処理の流れを図5に示す。まず(A)、(B)の領域を求めるために、2.2.1項に示す手法により、左アーム領域に対応する被遮蔽領域を左カメラ映像から、右アーム領域に対応する被遮蔽領域を右カメラ映像から探索する。次に直前の処理で取得不可能であった左右のアーム領域内の点を、探索するカメラ映像を変更し再探索を行う。最後に以上の処理で取得不可能であった領域を(C)とし、この領域を時系列映像で補間することで目的とする被遮蔽領域を取得する。

2.2.1 左右のカメラ映像を用いた探索

左右のカメラ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(図2)。そのため2.1節で取得したアーム領域の画像座標 $\mathbf{m}_{\text{cam}} = [u_{\text{cam}}, v_{\text{cam}}, f_{\text{cam}}, 1]$ に隠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、左右のカメラ映像上で $\mathbf{m}_{\text{camL}} = [u_{\text{camL},i}, v_{\text{camL},i}, f_{\text{camL}}, 1]$ 、 $\mathbf{m}_{\text{camR}} = [u_{\text{camR},i}, v_{\text{camR},i}, f_{\text{camR}}, 1]$ にそれぞれ像を結ぶとき以下の関係式が成立する。

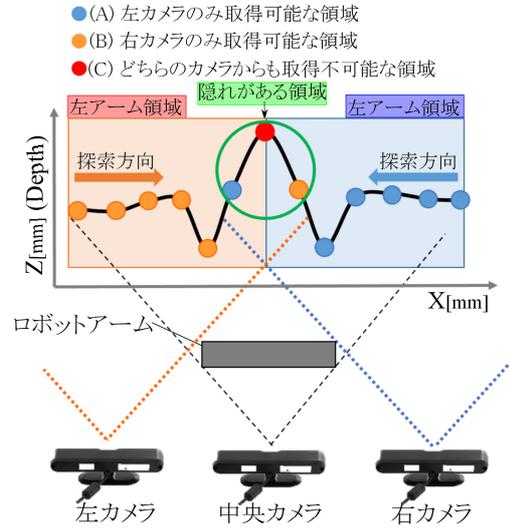


図4 各領域に対応するカメラ映像

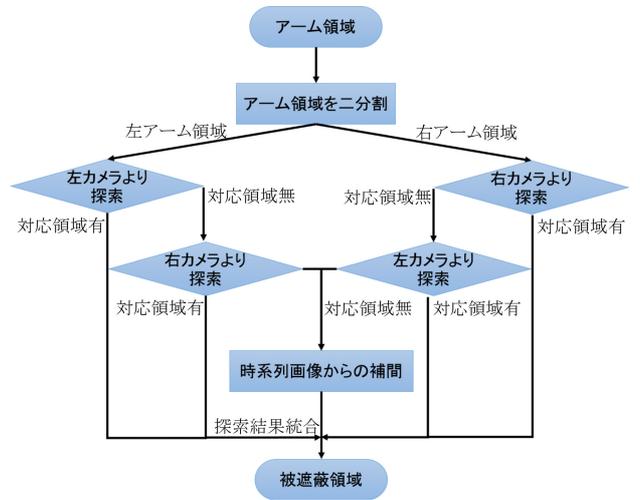


図5 被遮蔽領域取得までの流れ

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

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

f は、中央、左右の3台のカメラで共通の焦点距離であり、既知の値である。また b_{camLC} は左カメラと中央カメラのカメラ間の距離であり、 b_{camCR} は中央カメラと右カメラのカメラ間の距離である。添え字の i は左右の映像上の画像座標において $u_{\text{camR}} = u_{\text{camL}} = u_{\text{camC}}$ 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、各カメラ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式(4)および式(5)を最小にする i の探索を行うことでアーム領域に隠れている被遮蔽領域の取得が可能となる。

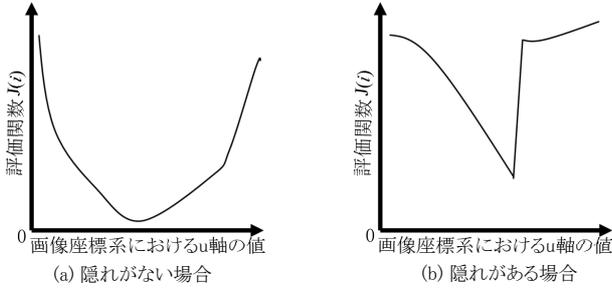


図 6 隠れがある場合と無い場合の評価値

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

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

図 4 に示すように、左アーム領域はアーム領域の左端を探索出発点とし、画像座標における u 軸の正方向に探索を行う。反対に右アーム領域はアーム領域の右端を探索出発点とし、画像座標における u 軸の負方向に探索を行う。以上の処理で各アーム領域に対応する遮蔽領域が取得可能になる一方で、図 4 に示すように、各カメラ映像に対して隠れ領域が存在する。そのため、このような領域は 1 回目の探索では取得が不可能である。そこで本研究では、隠れ領域を判定する方法として以下の手法を用いる。

本研究で用いる評価関数式 (4) および式 (5) は、仮に左アーム座標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。以上のことから、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 $d-1$ 、 $d+1$ のうち大きいほうの値との差とし、以下のように定義をする。

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

e および e' が任意の閾値より小さい場合を隠れがない領域とし、任意の閾値より大きい場合を隠れがあり、対応領域が取得不可能な領域とする。一度の処理で取得が不可能であった領域に対しては、探索に使用するカメラ映像を変更して再探索を行う。以上の処理を行った結果、取得不能であったアーム領域に関しては以下に示す手法により補間を行う。



(a) 使用機材と配置図 (b) 実験環境全体図

図 7 実験環境図

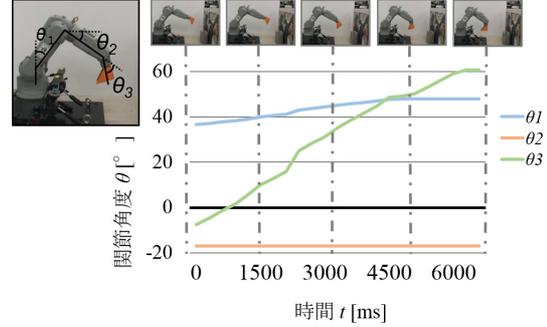


図 8 ロボットの関節角度

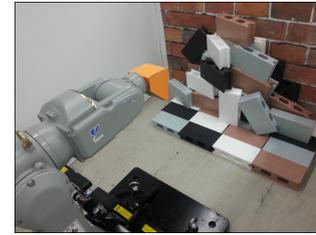


図 9 背景環境図

2.2.2 時系列映像を用いた補間

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、基本的に $t = T$ の状態の各カメラ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被遮蔽領域（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して何か作業を行う際には、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ままでは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、2.2.1 項の処理で対応する被遮蔽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。

2.3 取得した映像の合成

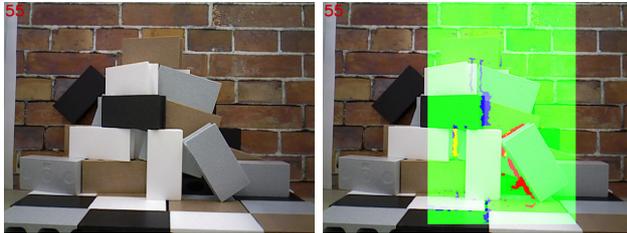
2.1 節で取得したアーム領域を半透明にし、2.2 節で得られた被遮蔽領域と共に中央カメラ映像に投影および合成処理を行う。投影位置は 2.1 節で求めた中央カメラ映像上のアーム領域とする。本研究では、2.1 節で取得したアーム領域を半透明にし 2.2 節で取得した被遮蔽領域と合成処理を行う際に以下の式を用いる。 I は各映像の各ピクセルにおける RGB の色成分である。また α は不透明度を表す係数である。 $\alpha = 0$ で完全な透明となり、 $\alpha = 1$ で不透明な状態を示す。

$$I_{\text{output}} = \alpha I_{\text{arm}} + (1 - \alpha) I_{\text{cover}} \quad (7)$$



(a) 左カメラ映像 (b) 中央映像 (c) 右カメラ映像

図 10 入力画像



(a) 背景画像 1

(b) 背景画像 2



(c) 半隠消映像

図 11 出力画像

3. 実験

3.1 実験概要

前章で述べた処理を用いて、視差のある3つ映像からアーム領域の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、図 7(a) のように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、前方の瓦礫にアクセスする際にロボットアームに隠れて見えない瓦礫が透視可能な映像の生成を行う。前章で述べた処理と同様、中央カメラ映像のアーム領域に対応する被遮蔽領域を左右のカメラ映像より取得する。

実験環境の全体図を図 7(b) に示す。また実験環境の座標系および機器の位置・姿勢を図 7(a) に示す。Z軸方向に150mm、Y軸方向に250mmの位置に中央カメラを設置した。さらに中央カメラからX軸方向に200mm離れている位置に左カメラを設置し、X軸方向に-200mm離れている位置に右カメラを設置した。取得映像サイズは640×480pixelとした。

背景物体の図を図 9 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つ目の物体としては画像が貼り付けられている幅1000mm、高さ800mmの板状のボードとした。2つ目の物体としてはボードの手前に散在させた幅200mm、高さ200mm、奥行き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 8 のように各関節角 $\theta_1 \sim \theta_3$ 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、 $\theta_1 \sim \theta_3$ を図 8 のように変動させた際の各カメラ映像を入力映像とした。また (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。

3.2 結果と考察

ロボットアームを図 8 のように変動させた際の $t = 4500\text{ms}$ における、各カメラからの入力映像を図 10 に示す。図 10(a) が左カメラ、図 10(b) が中央カメラ、図 10(c) が右カメラより得られた映像である。また探索により取得不可能であった欠損領域を補間する際に用いた画像を図 11(a) に示す。

以上の画像群を用いた際の出力結果を図 11(b) および図 11(c) に示す。図 11(b) は、図中の緑色部分が2.2.1項の処理において一度目の探索で取得が可能であった領域、赤色および青色部分が探索を行うカメラ映像を変更し、再探索した際に取得が可能になった領域である。また、黄色部分は2.2.2項の処理を用いて欠損領域の補間を行った領域である。図 11(a) および図 11(c) の画像を比較しても分かるとおり、背景が複雑な形状の物体であっても半隠消映像が生成可能であることが分かる。

4. 結論

本稿では、3台のRGB-Dセンサを使用し、中央に設置したカメラ映像中のアーム領域を半透明にすることでアームに隠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、目的とする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。

参考文献

- [1] 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.
- [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. Enomoto and H. Saito: “Diminished Reality Using Multiple Handheld Cameras”, Proceedings of the 8th Asian Conference on Computer Vision 2007, pp. 130-150, 2007.
- [4] 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.
- [5] 清水 直樹, 橋本 章宗, 植松 裕子, 斎藤 英雄: “デプスカメラを用いたリアルタイム領域抽出による隠消現実感映像生成”, 映像情報メディア学会誌, Vol. 66, No. 12, pp. J549-J552, 2012.
- [6] 森 尚平, 一川 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.
- [7] 杉本 和也, 藤井 浩光, 山下 淳, 浅間 一: “無人化施工のためのカメラとRGB-Dセンサを用いた半隠消映像の生成”, 第20回画像センシングシンポジウム講演論文集 (SSII2014), IS2-28-1-IS2-28-4, 2014.