

# 遠隔操作ロボットシステムのための複数カメラを用いた半隠消映像の生成

○杉本 和也, 藤井 浩光, 佐藤 貴亮, 山下 淳, 浅間 一 (東京大学)

## Half Diminished Reality Image for Remote Control Robot System Using Multiple Cameras

○Kazuya SUGIMOTO, Hiromitsu FUJII, Takaaki SATO, Atsushi YAMASHITA, and Hajime ASAMA (The University of Tokyo)

**Abstract**— In recent years, remote control robot systems are used for disaster restoration in dangerous areas. In these systems, an operator controls a robot machine remotely viewing image through on-board cameras. However, working efficiency decreases compared with visual operation. In case of using a robot arm, an operator can not see objects on image because the robot arm obstructs them. Thus, it is effective to increase working efficiency, if the operator can see behind the robot arm on single image. In this paper, we propose a method to make half diminished reality image for remote control systems using multiple cameras.

**Key Words:** Half Diminished Reality Image, Remote Control Robot System, Multiple Cameras

### 1. 序論

災害発生時の緊急対応作業の際には、人が直接踏み込むことが危険な場所が数多く存在する。近年、これらの場所では、二次災害を軽減するためにロボットの遠隔操作による作業が行われている。ロボットを遠隔操作する際に達成すべき作業の中には、ロボットアームを用いた瓦礫の除去やバルブの開閉などが存在する。その際、オペレータは、ロボットに搭載されているカメラから得られた映像をもとにロボットアームの操作を行うことが多い。しかし、カメラ映像を用いて遠隔操作を行う場合、目視による操作時と比較して作業効率が低下するといわれている [1]。その原因の一つとして、手前にあるアーム (Fig. 1) 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]。複数のカメラを用いることにより、オペレータは多視点から作業対象物の形状を把握することが可能となる。しかし、実際の作業現場では、作業対象物以外のものが数多く隣接するため、様々な視点からのカメラ映像が必要となる。その際、オペレータは個々の映像を見比べながら作業を行う必要があるため、長時間の作業はオペレータの負担となり作業効率の低下を招く原因になる。そこで、オペレータがカメラ映像をもとにロボットアームを操作する際に、アームによって遮られて見えない作業対象物を、1つの映像上で見えるようにすることは有効であると考えられる。

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。文献 [3] では、映像上で物体を完全に消しきることなく、半透明状態にすることでその物体に隠れている背景が透視できるような表現のことを「半隠消現実感」と呼称している。「半隠消現実感」を用いた研究は、エンターテインメントの分野において数多く報告されている [4–6]。ロボットアームを用いた作業において、「半隠消現実感」を用いることは、アームによって遮られて見えない作業対象物を

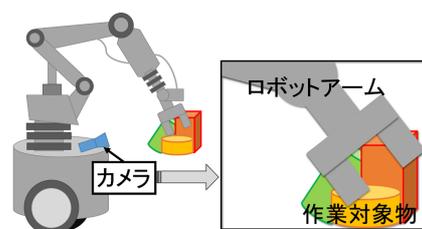


Fig. 1 作業対象物がアームによって遮られている前方映像

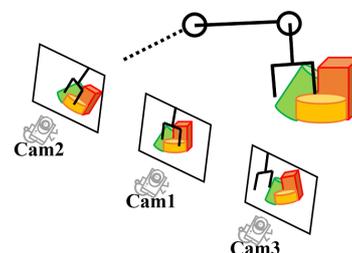


Fig. 2 半隠消映像を生成する環境

1つの映像上で確認可能であると考えられる。そこで本研究では、遠隔操作によるロボットアームを用いた作業の際に、カメラ映像上で映し出されているアーム部分を半透明することで、背景領域が透視可能な半隠消映像生成システムの構築を目的とする。

### 2. 提案手法

ロボットに搭載されている複数台のカメラを用いることで、カメラ映像上に映し出されているアーム部分 (以下、アーム領域) の半隠消映像の作成を行う。使用するカメラは3台であり、それぞれ視差が得られるように設置する (Fig. 2)。目的とする半隠消映像を生成するためには、

1. アーム領域の抽出と除去
2. アーム領域に隠れている背景領域の抽出
3. アーム領域と背景領域の投影と合成

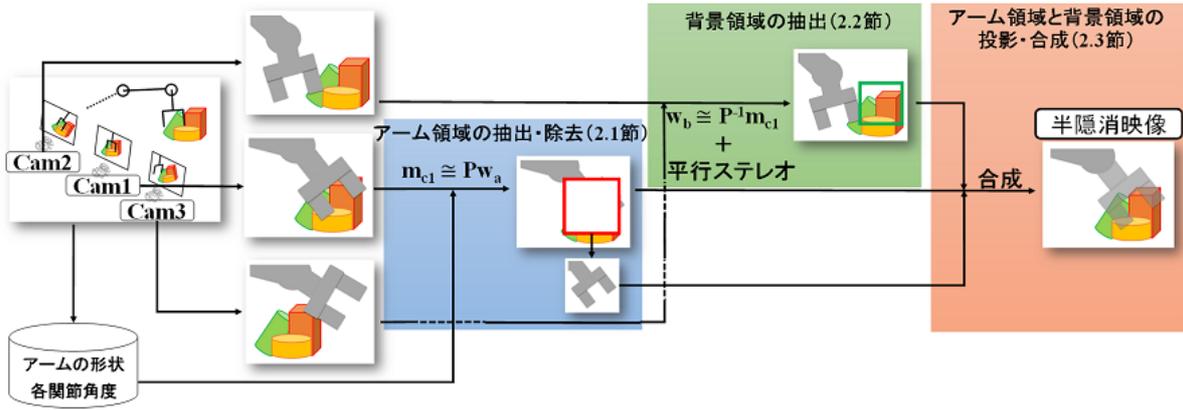


Fig. 3 半隠消映像生成までの流れ

といった処理をそれぞれ行う必要がある。全体的な処理の流れを示した図を Fig. 3 に示す。処理 1~3 の内容を以下に示す。

### 2.1 アーム領域の抽出と除去

Cam1 映像中のアーム領域は、アームの各関節角度を取得することで認識・追跡が可能になる。そこで、あらかじめ用意したロボットアームの形状データから、ロボットアームの 3 次元モデルを作成し、各関節角度を随時取得することで、Cam1 映像中のアーム領域の抽出および除去を行う。

3次元モデル上で、Fig. 4 のように設定したワールド座標系におけるアーム位置座標  $\mathbf{w}_a = [x_a, y_a, z_a, 1]^T$  と、Cam1 映像上の画像座標  $\mathbf{m}_{c1} = [u_{c1}, v_{c1}, f_{c1}, 1]^T$  は、行列  $\mathbf{P}$  を用いると以下のような関係がある (Fig. 4)。ここで  $f_{c1}$  は Cam1 の焦点距離である。また、 $\mathbf{P}$  はカメラの内部パラメータおよびカメラの位置・姿勢を表す回転行列、平行移動行列から成る  $4 \times 4$  の行列である。

$$\mathbf{m}_{c1} \simeq \mathbf{P}\mathbf{w}_a \quad (1)$$

ここで、 $\simeq$  は、左辺は右辺の 0 でない定数倍と等しいということの意味する。上式を用いることで、アーム位置座標に対応する画像座標を求めることにより、Cam1 映像中のアーム領域を取得する。

### 2.2 アーム領域に隠されていた背景領域の抽出

2.1 節で得られた Cam1 映像上のアーム領域を用い、Cam2, Cam3 映像上でこの領域に隠れている背景領域を求める。そこで、式 (1) で用いられた、 $4 \times 4$  の行列  $\mathbf{P}$  の逆行列

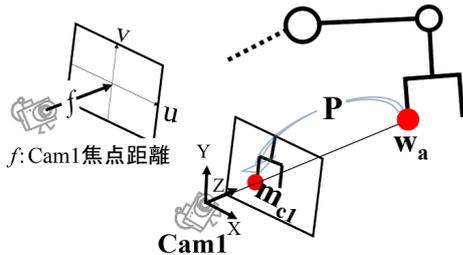


Fig. 4 アーム座標の認識

$\mathbf{P}^{-1}$  を用いる。逆行列  $\mathbf{P}^{-1}$  を用いることで、ワールド座標系におけるアーム領域の画像座標  $\mathbf{m}_{c1} = [u_{c1}, v_{c1}, f_{c1}, 1]^T$  と背景座標  $\mathbf{w}_b = [x_b, y_b, z_b, 1]^T$  は以下の式で表すことができる。

$$\mathbf{w}_b \simeq \mathbf{P}^{-1}\mathbf{m}_{c1} \quad (2)$$

しかしこの式を用いるだけでは、2次元座標から3次元座標を求めているため一般的に解は無数にあることになる (Fig. 5)。

そこで、左右に設置されている Cam2, Cam3 の視差映像を用いて上式における解を一意に決定する。全てのカメラの内部パラメータが等しい場合、光軸の方向をそろえ、視点の高さも等しく設置することにより平行ステレオ法 [7] を用いることで、三角測量の原理により背景座標の取得が可能となる。キャリブレーションされた Cam2, Cam3 において、Fig. 6 のように、ワールド座標系における背景座標  $\mathbf{w}_b = [x_b, y_b, z_b]^T$  が、Cam2, Cam3 の画像座標系においてそれぞれ  $\mathbf{m}_{c2} = [u_{c2}, v_{c2}]^T$ ,  $\mathbf{m}_{c3} = [u_{c3}, v_{c3}]^T$  に像を結んだとき、以下の関係式が成立する。

$$x_b = \frac{b(u_{c2} + u_{c3})}{2(u_{c2} - u_{c3})} \quad (3)$$

$$y_b = \frac{bv}{(u_{c2} - u_{c3})} \quad (4)$$

$$z_b = \frac{bf}{(u_{c2} - u_{c3})} \quad (5)$$

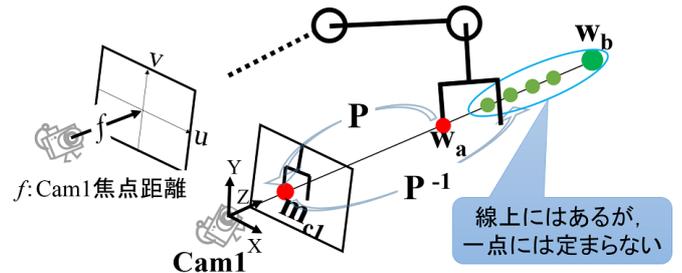


Fig. 5 背景座標の認識

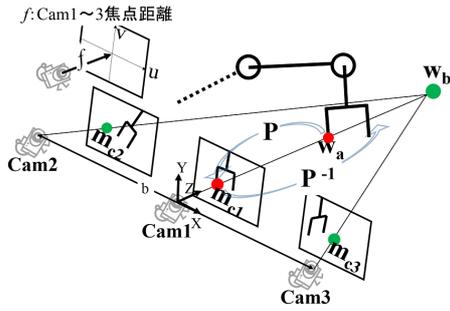


Fig. 6 平行ステレオを用いた背景座標の認識

ここで、 $f$  は Cam1~Cam3 の焦点距離、また  $v = v_{c1} = v_{c2} = v_{c3}$  である。式 (1) で求めたワールド座標系における背景座標  $\mathbf{w}_b = [x_b, y_b, z_b, 1]^T$  を式 (3) ~ (5) にそれぞれ代入する。これにより、左右に設置されている Cam2, Cam3 映像上で背景領域の画像座標  $\mathbf{m}_{c2} = [u_{c2}, v_{c2}]^T$ ,  $\mathbf{m}_{c3} = [u_{c3}, v_{c3}]^T$  をそれぞれ求めることが可能となる (Fig. 6)。これにより、2.1 節で抽出および除去したアーム領域に隠されていた背景領域の Cam2 または Cam3 映像上で対応する領域が取得可能となる。

### 2.3 アーム領域と背景領域の投影と合成

2.1 節で抽出および除去したアーム領域を半透明にし、2.2 節で得られた背景領域と共に Cam1 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 Cam1 映像上のアーム領域とする。本研究では、アーム領域を半透明にし背景領域と合成処理を行う際に以下の式を用いる。 $I, I_s, I_d$  は、それぞれ出力画像、アーム領域画像、背景領域画像の各ピクセルでの色成分である。また、 $\alpha$  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。

$$I = \alpha I_s + (1 - \alpha) I_d \quad (6)$$

## 3. 実験

### 3.1 実験内容

前章で述べた処理により、視差を有する 3 つの視点から得られた映像を用いることで、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、ロボットアームが行う作業としては、前方の瓦礫の移動・撤去などが存在する。そこで本実験では、ロボットアームを用いた瓦礫の移動・撤去作業を想定し、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。提案手法の有効性を示すためには、視差映像を獲得するための最適なカメラ位置を評価する必要がある。そこでカメラ位置を容易に変化可能にするために、ロボットアームの 3 次元モデルおよび作業対象物となる瓦礫モデルを含んだシミュレータを作成し、シミュレーションによる実験を行う。

シミュレータ上で任意の 3 つの視点からの映像を作成し、それぞれモデル Cam1 映像、モデル Cam2 映像、モデル Cam3 映像と呼ぶことにする。作成した各モデル映像を用いて半隠消映像を生成する。Fig. 7 のような位置・姿勢に指定したロボットアームに対し XYZ 軸を設ける。まず、Z 軸方向に 200mm、Y 軸方向に 200mm、X 軸方向に -100mm の位置にモデル Cam1 映像の視点を設置した。

さらに、Cam1 映像の視点を中心に各視点同士が X 軸方向に 200mm~400mm ずつ離れるようにモデル Cam2、モデル Cam3 映像の視点をそれぞれ設置した (Fig. 7)。取得映像のサイズは 640×480pixel とした。

瓦礫モデルとしては、実際の作業現場で散在した瓦礫の移動・撤去作業を行うことを想定し、瓦礫を擬似的に再現した異なる 2 つの位置・形状の物体を用いた。1 つ目の物体としては、テクスチャ (Fig. 8) が貼り付けられている幅 640mm、高さ 360mm の板状のボードとした。2 つ目の物体としては、テクスチャの手前に置かれた、幅 100mm、高さ 200mm、奥行き 100mm の直方体形状のものとした。作成したロボットアームモデルに対し、板状のボードおよび直方体形状の物体の設置位置を Fig. 7 に示す。本実験では、設置した各視点から背景物体までの距離は既知とした。以上の条件より、本実験では、アーム領域に隠れて見えない背景領域に異なる位置・形状の物体を含む場合、物体ごとに即した領域を視差映像から抽出する必要がある。

次に、作成したロボットアームモデルに対し、Fig. 9 のように各関節角  $\theta_1 \sim \theta_3$  を設定する。ロボットアームが前方の物体にアクセスする動作 (Fig. 9) を想定し、 $\theta_1 \sim \theta_3$  が、 $0.73\text{rad} \leq \theta_1 \leq 0.93\text{rad}$ ,  $-1.92\text{rad} \leq \theta_2 \leq -0.86\text{rad}$ ,  $-1.39\text{rad} \leq \theta_3 \leq -0.35\text{rad}$  の時の、各視点からのモデル映像を入力映像とした。本実験では、入力映像中でロボットアームの 3 次元モデルのエンドエフェクタ (Fig. 9) を半透明にすることで半隠消映像を生成した。また、(6) 式における、透過値を表す係数  $\alpha$  は、半透明にする全ての領域で 0.5 とした。

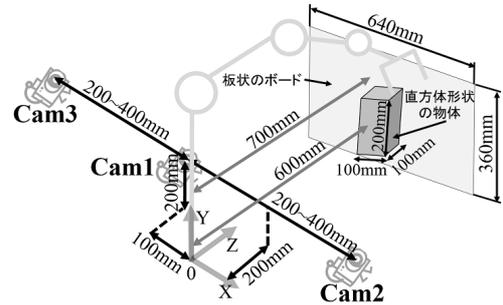


Fig. 7 各カメラと背景領域となる物体の位置



Fig. 8 背景とするテクスチャ

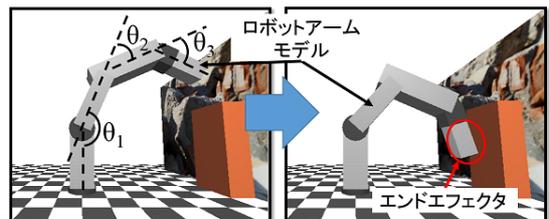


Fig. 9 ロボットアームモデルと各関節角度

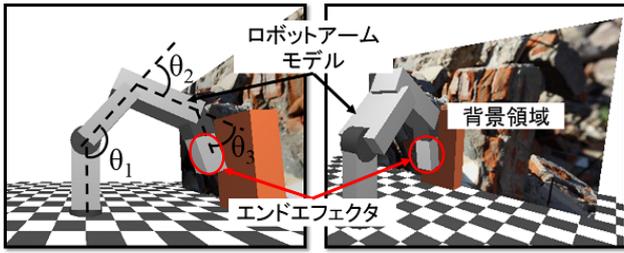
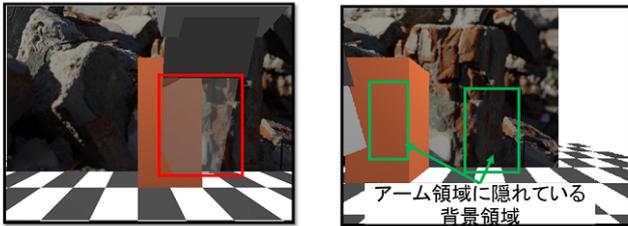


Fig. 10 シミュレータ全体図

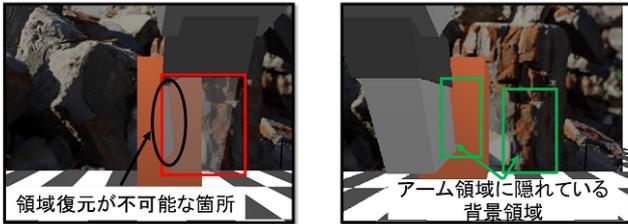


Fig. 11 入力画像



(a) (b)

Fig. 12 出力画像：各視点間の距離 400mm の場合



(a) (b)

Fig. 13 出力画像：各視点間の距離 200mm の場合

### 3.2 結果と考察

ロボットアームモデルが前方の物体に対して Fig. 10 のようにアームを振り下ろした時 ( $\theta_1 = 0.73\text{rad}$ ,  $\theta_2 = -1.39\text{rad}$ ,  $\theta_3 = -0.87\text{rad}$ ) 入力画像と出力画像を Fig. 11~13 に示す。

Fig. 11 は、ロボットアームモデルが上記の位置・形状の時の、モデル Cam1 映像である。本実験では、この入力画像においてエンドエフェクタ領域 (Fig. 11) を半透明にすることで半隠消映像を生成した。また、モデル Cam1 映像でアーム領域に隠れて見えない領域に異なる位置・形状の物体を含むため、物体ごとに即した背景領域を視差映像から抽出した。Fig. 12(a) は、各視点間の距離が 400mm の場合にエンドエフェクタ領域を半透明にし、この領域に

隠れている背景を復元した結果である。また、Fig. 12(b) は、その際のモデル Cam2 映像と、物体ごとに抽出した背景領域を示した図であるこの結果、エンドエフェクタ領域全体で、この領域に隠れて見えない背景領域を透視可能な半隠消映像の生成が可能となった。

一方、Fig. 13(a) は、各視点間の距離が 200mm の場合に、エンドエフェクタ領域を半透明にし、この領域に隠れている背景を復元した結果である。また、Fig. 13(b) は、その際のモデル Cam2 映像と、物体ごとに抽出した背景領域を示した図である。この結果、一部のエンドエフェクタ領域で、背景領域が復元不可能な領域が存在した。これは、各視点間の距離が小さく、十分な視差を得られなかったためである。

以上より、カメラ映像上でどの位置にアーム領域がきたとしても、安定した背景領域の復元を行うためには、各視点間の距離を十分に広くとる必要があることが分かる。しかし本研究では、ロボットに搭載されている複数台のカメラを用いることを前提としているため、ロボットの機構によりカメラの設置位置が制約される可能性が大いにある。そこでロボットの機構を考慮した際、本研究で提案したように、視差が得られるように設置した 3 台のカメラのみでは背景領域の復元が不可能な状況が考えられる。そのため、ロボットアームが前方の作業対象物にアクセスする前に計測された映像を用いるといった手法の検討が必要である。

### 4. 結論

本稿では、3 台のカメラを使用することで、ロボットアームのアーム領域を半透明にし、背景領域が透視することが可能な半隠消映像の生成システムを提案した。また本手法を用いることで、モデル上で任意に設定したカメラ位置において、半隠消映像が生成可能なことを確認した。今後は、実機で取得した映像からリアルタイムに半隠消映像を生成するシステムを構築していく。

### 参考文献

- [1] 林 宏樹, 田村 隆幸, 野末 晃: “油圧ショベル遠隔操作における通信遅延の影響分析と改善提案”, 建設機械, Vol. 46, No. 10, pp. 32-37, 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] 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.
- [4] 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.
- [5] 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.
- [6] 清水 直樹, 橋本 章宗, 植松 裕子, 斎藤 英雄: “デブ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.
- [7] 出口 光一郎: ロボットビジョンの基礎, コロナ社, 2000.