悪環境下での調査・状況確認活動を支援する目標可視化技術の研究開発

代表研究者	陸　慧敏	九州工業大学	工学部　准教授
共同研究者	金　亨燮	九州工業大学	工学部　教授
共同研究者	李　玉潔	福岡大学	工学部　助教
共同研究者	芹川　聖一	九州工業大学	工学部　教授

1 概要

現在，日本では自然災害発生頻度が増加傾向にあり，それに伴い危険な作業環境での人的作業も増加している．そのような環境での作業は，常に事故による傷害や死亡のリスクを伴う．それらの問題を解決するため，危険な環境下で作業を行うロボットの導入が進んでいる．人に代わってロボットが作業を行うことにより，事故のリスクを格段に減らすことが可能となる．

災害現場でのロボットの導入については，様々な状況に対応し作業を行う必要があるため，完全な自動化は行われておらず，ロボットの遠隔操作技術が必須となっている．近年のロボットの遠隔操作技術には 2D モニターからロボットの作業環境を認識し，操作を行うという手法が主流となっている．この方法でロボットの操作を行うには，相応の技術が必要となり，十分な訓練を受けた操縦者が必要となる．そのため，より簡単に操作を行うことができる新たな遠隔操作技術が必要不可欠である．また，災害現場によっては夜間や屋内での作業も想定され，そのような環境下では光源不足により，カメラなどのセンサーによる映像から，作業環境を正しく認識することが難しくなる．本研究では，これらの問題を解消するため，新たな遠隔操作技術の開発として，作業環境の再現方法および低光源環境下における映像改善方法について検討を行っている．

本研究では 様々な分野で注目を集めている Virtual Reality 技術を用いて遠隔地の作業環境の再現を行う．また，近年その有用性について注目されているヘイズ除去アルゴリズムを利用した低光源環境下における映像改善を行う．ロボットの作業環境を，単一の RGB-D センサーを用いてカラー画像および深度画像として取得する．得られたカラー画像に対しては映像改善処理を行い，画像データは WebSocket を介して操縦者のいる遠隔地へ送信される．その情報をもとにロボットの作業環境を点群として Virtual Reality 空間へ再現を行い，その有用性を検討する．

2 従来法

夜間や屋内などの低光源下で撮影された画像（動画を含む）は，通常，視認性およびコントラストが著しく低下し，そのような低品質な画像では対象物の観察や分析に悪影響を及ぼす可能性がある．そこで画像処理技術を用いた低光源下映像の品質向上を図るための技術として，監視カメラ，夜間の運転，医学，映画製作などにおける幅広い用途で注目されている[5]．

視認性の高い画像は，対象シーンの明確な解析を行う上で必要不可欠である．画像における高視認性は，物体検出や追跡などの多くの画像処理技術で求められている．しかし，夜間や屋内などの低光源下で撮影された画像は，視覚的品質を低下させるだけでなく，主に視認性の高い入力用に設計されているアルゴリズムの性能を低下させる可能性がある[6]．

暗い画像に対し，視認性を高めるという目的では，最も直感的に考えられるのが，画像のコントラストを補正するという方法である．しかし，この操作では比較的明るい領域は飽和状態となり，対応する箇所の詳細情報が失われる可能性があり実用的ではない．過去数年にわたり，夜間や屋内などの低光源下で撮影された画像および動画の品質向上のため様々な手法が提案されてきた．それらは主にヒストグラム均等化法，Retinex 理論に基づく方法，ヘイズ除去モデルを利用する方法の 3 つのカテゴリに分類することができる[7,8]．

本研究ではその中でも近年，その有用性について注目を集めている，ヘイズ除去モデルを利用した画質改善を行う．本論文で用いたヘイズ除去モデルを利用する手法については第 3 章で述べる．以下ではその他のヒストグラム均等化法および Retinex 理論を利用する方法の基本的な理論について述べる．

画像のコントラスト改善の代表的な手法として，ヒストグラム均等化法がある[8]．この手法は，図 2.1 に示すような原画像の濃度ヒストグラムを，図 2.2 のように出力画像の濃度ヒストグラムが全画素の濃度値が平坦化するように変換する手法である．多くの画素が同じ濃度値の場合に分配する方法には，乱数で分散させる方法と近傍画素の平均濃度値の高い画素から順に選ぶ方法がある．また，分配を行ず，入力濃度値ごとに出力濃度値を変換する手法もある．濃度ヒストグラムは平坦にはならないが，近似的には同様の効果があり，画像を一括で変換でき処理時間も早い．そして，画素の濃度分布が均等のため情報量が最大となり，非常にコントラスト感が高い画像が得られる．

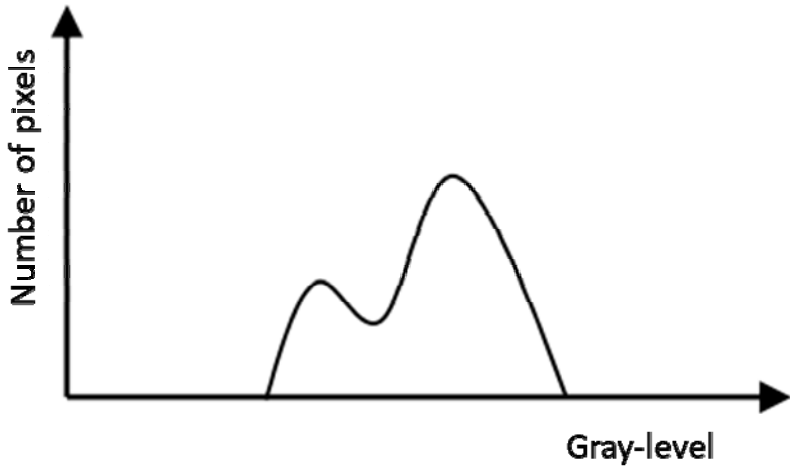


図 2.1　原画像の濃度ヒストグラム[8]

図 2.2　ヒストグラム均等化法[8]

ヒストグラム均等化法を用いてコントラストの改善を行う際にある特定の濃度値が多数存在する場合には，多量のノイズが発生することや過度な強調により不自然な出力画像が得られるという問題点を抱えている．この不自然さを解消するため，ヒストグラム平均化を局所領域に適用する方法などさまざまな手法が提案されている[9-12]．

Retinex 理論は，人間の視覚システムが色をどのように捉えるのかをモデル化した理論である．Retinex 処理は以下の二つの仮定の下で行われる．(1)画素値は照明光と被写体の反射率の積である．(2)照明光は空間的に滑らかに変化している.すなわち,照明成分を取り出した画像における画素値の変化は滑らかである．(2)の仮定を満たす方法で照明光分布を推定し，(1)の仮定に基づいて反射率を推定する．以下に Retinex 処理について簡易的に紹介する．

仮定(1)は，

$$I^c(x) = L^c(x)R^c(x) \qquad (2.1)$$

と表現できる．ここで，$I^c(x)$は入力画像，xは画素位置，Cは RGB 成分のいずれかを表す．一般的な Retinex 手法では，まず画像中の照明分布Lを推定し，それを用いて被写体の反射率Rを得る．その後LやRを調整し，視認性の良い画像とする．代表的には Single-Scale Retinex (SSR)[13][14]や Multi-Scale Retinex (MSR)[13]といった手法が存在する．SSR および MSR による処理の結果を図 2.3 に示す．

(a)　入力画像

(b)SSR

(c)MSR

図 2.3　Retinex 処理の例[13]

SSR や MSR の手法では，各色成分を独立に処理し，反射率成分を調整したものを出力としている．これらの処理では，各色成分を独立に処理するため，偽色が発生するという問題がある．また，照明不足による視認性の悪さは改善されるが，照明成分が除去されてしまうことから，出力画像が不自然となってしまう問題点があり，それらを解消するために，様々な手法が提案されている[6, 7, 13, 14]．

2 映像補正

次に，本研究で用いたヘイズ除去モデルによる低光源下における映像改善に関する画像解析手法について述べる．

ヘイズ除去モデル[15]を利用した方法は，暗い画像の RGB 成分を反転したものが，霧やもやなどのヘイズに類似していることに基づき，ヘイズ除去手法を適応させるといったものである．

文献[16]で述べられているように，霧や曇りなどのヘイズがある画像では，背景の画素値はすべてのチャンネルで常に高い．しかし，色の鮮やかさや影などの影響により，住宅，自動車，および人などの主要物体を含む領域では，RGB チャンネルの中で少なくとも 1 つのチャンネルで低い値を取る．それに対し，低光源下で撮影されたビデオについて RGB 成分の最小値を取り，ヘイズ画像との類似性について調査を行った結果，ランダムに選ばれたそれぞれ 30 個のヘイズ動画および RGB 反転した低照明動画について，全画素の RGB 成分

の最小値のヒストグラムが非常に類似しているという研究結果が得られている[15]．図 3.1 に夜間に撮影された画像および RGB 成分反転操作後の画像，霧を含んだヘイズ画像を示す．

図 3.1　低照明度画像とその反転画像およびヘイズ画像例（[17，18]のデータセット）

上記の観察に基づき，低照明度画像に RGB 成分の反転操作を適用後，ヘイズ除去アルゴリズムを適用後，再度 RGB 成分の反転操作を行い，最終的な出力とする低照明度画像強調手法が提案された．本論文では，上記のアルゴリズムを用いた文献[5]の手法を利用する．

まず，低照明入力画像に対し，次のように反転処理を行う．

$$I_{inv}^{c}(x) = 255 - I^{c}(x) \qquad (3.1)$$

ここで，cは RGB カラーチャンネルである．$I^{c}(x)$は低光源下で撮影された入力画像，$I_{inv}^{c}(x)$は入力画像に対し，反転処理を行った画像である．

文献[15]で提案された画像劣化モデルは以下のように表される．

$$I_{inv} = J_{inv}(x)t(x) + A\big(1 - t(x)\big) \qquad (3.2)$$

ここで，$J_{inv}(x)$は情景の輝度を，Aは環境光を表し，透過率$t(x)$よってその比率が決まる．式(3.2)に表された関係より，Aと$t(x)$を求めれば次式のようにI_{inv}から$J_{inv}(x)$を取得することができる．

$$J_{inv}(x) = \frac{I_{inv} - A}{t} + A \qquad (3.3)$$

最後に，$J_{inv}(x)$に対し変換処理を行い，画像を出力する．この処理の際，環境光Aおよび透過率$\iota(x)$を推定する必要がある．以下にこれらのパラメータの推定方法および最終的な改善画像を取得する方法について

述べる．

まず，次式に示すように反転処理後の入力画像に対し RGB 成分の最小値を取る．

$$I_{min}(x) = \min_{c\in\{r,g,b\}}\left(I_{inv}^{c}(x)\right) \qquad (3.4)$$

$I_{mtn}(x)$では，常にノイズが多く含まれているため，ここで中央値フィルタを用いた平滑化処理を行う．次に，平滑化された$I_{mtn}(x)$の中で輝度の最も高い画素を選択し，その位置の画素値を環境光Aとする．

環境光の推定を行った後，透過率の推定を行う．透過率$t(x)$を局所的にできるだけ滑らかに維持するように推定を行う．最初に，反転処理後の入力画像$I_{inv}^{c}(x)$に対し RGB 最小値を取る．このグレイスケール画像をダークチャンネル$\tilde{I}_{dark}(x)$と定める．

$$\tilde{I}_{dark}(x) = \min_{c\in\{r,g,b\}}\left(\frac{I_{inv}^{c}(x)}{A^{c}}\right) \qquad (3.5)$$

ここで，$\tilde{I}_{dark}(x)$の平滑化のために，ガウシアンピラミッド処理を適用する．本論文では，ダウンサンプリングおよびアップサンプリングを一度行う．そして，中央値フィルタを適用する．

$$I_{med}(x) = \mathrm{Med}_S\left(\tilde{I}_{dark}(x)\right) \qquad (3.6)$$

Sは中央値フィルタのスケールである．エッジを保持しながら透過率を滑らかにするため，以下の処理を行う．

$$I_{datail}(x) = \mathrm{Med}_S\left(\left|I_{med}(x) - \tilde{I}_{dark}(x)\right|\right) \qquad (3.7)$$

$$I_{smooth}(x) = I_{med}(x) - I_{datail}(x) \qquad (3.8)$$

さらに，ダークチャンネルを以下のように最適化することができる．

$$I_{dark}(x) = \begin{cases} \mu\tilde{I}_{dark}(x) & \text{if } \mu\tilde{I}_{dark}(x) < I_{smooth}(x) \\ I_{smooth}(x) & \text{others} \end{cases} \qquad (3.9)$$

すなわち，$\mu\tilde{I}_{dark}(x)$が$I_{smooth}(x)$よりも小さい値をとる場合は$\mu\tilde{I}_{dark}(x)$を，その他の場合は$I_{smooth}(x)$をダークチャンネルの値とする．本論文ではμの値を 0.95 としている．

そして，透過率は以下のように求められる．

$$t(x) = 1.0 - \omega I_{dark}(x) \qquad (3.10)$$

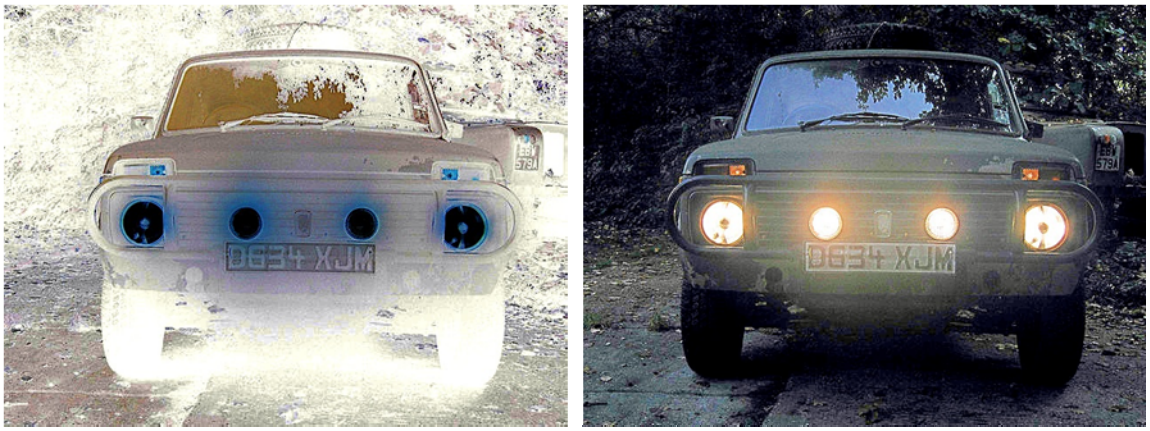
ここで，ωは$t(x)$の値を調整するためのパラメータであり，本論文では 0.98 としている．

推定した透過率$t(x)$および環境光Aを式(3.3)に適応することにより，$J_{inv}(x)$を取得することができる．次式により RGB 成分の反転を行い，最終的な改善画像として出力を行う．

$$J^{c}(x) = 255 - J_{inv}^{c}(x) \qquad (3.11)$$

低光源下で撮影された画像に対し，以上の処理を行った結果を図 3.2 に示す．

(a)入力画像　　(b)RGB 成分反転画像



(c)ヘイズ除去モデル使用後の画像　　(d)出力画像

図 3.2　低光源下映像改善処理の結果（[17]のデータセット）

4. 遠隔操作における遠隔地再現手法

近年，インターネットを介してモニターなどの 2 次元インターフェースによりロボットの操作を行う遠隔操作技術が普及している[19]．これらの遠隔操作技術では，ロボットの制御を行うために 2D モニターおよびキーボードが利用されてきた[20]．しかし，2 次元または 3 次元の情報を 2D モニターから観察し，キーボードなどの入力機器を用いてロボットを操縦することは容易ではなく，遠隔操作を行う際には関連する知識および高い操縦技術が必要となる．したがって，ロボットの遠隔操作を行うためには十分な訓練を受ける必要があり，操作を行うことができるのは限られた技術者のみとなっている．

ロボットの遠隔操作の簡易化を図るための新たな技術は必要不可欠である．そこで，本論文では 2D モニターインターフェースを利用する代わりに，近年様々な分野で目覚ましい活躍を見せている Virtual Reality (VR)技術を利用する方法を考案する．ロボットの作業環境再現に VR デバイスを用いることにより，作業環境の認識を容易に行うことができる．また，操縦技術を持っていない人でも簡単な訓練のみでロボットの遠隔操作を行うことができるようになると考えられる．

しかし，現在の Virtual Reality システムは ROS(Robot Operating System)[21]などのロボットシステムを構成するフレームワークをサポートしておらず,ロボットと VR システムとの結合を行う標準的なインターフェースは存在しない．そこで本論文では，David Whitney らが開発した ROS Reality[22]システムを参考に，ロボットシステムと Virtual Reality システムの結合を行い，遠隔地の作業環境再現を行う．

システムの概要

ロボットの遠隔操作システムを構築するためには，まずロボット作業環境の計測が必要となる．本論文では，RGB-D センサーを用いてカラー画像および深度画像の取得をすることにより，作業環境の計測を行う．また，暗い環境下の場合にはカラー画像に対しては第 3 章で述べた画像解析手法により画質改善を行う．得られた画像データは Rosbridge WebSocket[23]を介して VR デバイスを接続した Unity ゲームエンジンを実行しているコンピュータへ送信する．Unity コンピュータ上で受信したカラー画像および深度画像をもとに点群データを作成し，遠隔地の作業環境の再現を行う．以上で述べた遠隔地作業環境再現システムの概要図を図 4.1 に示す．また，以下にシステムの詳細について述べる．

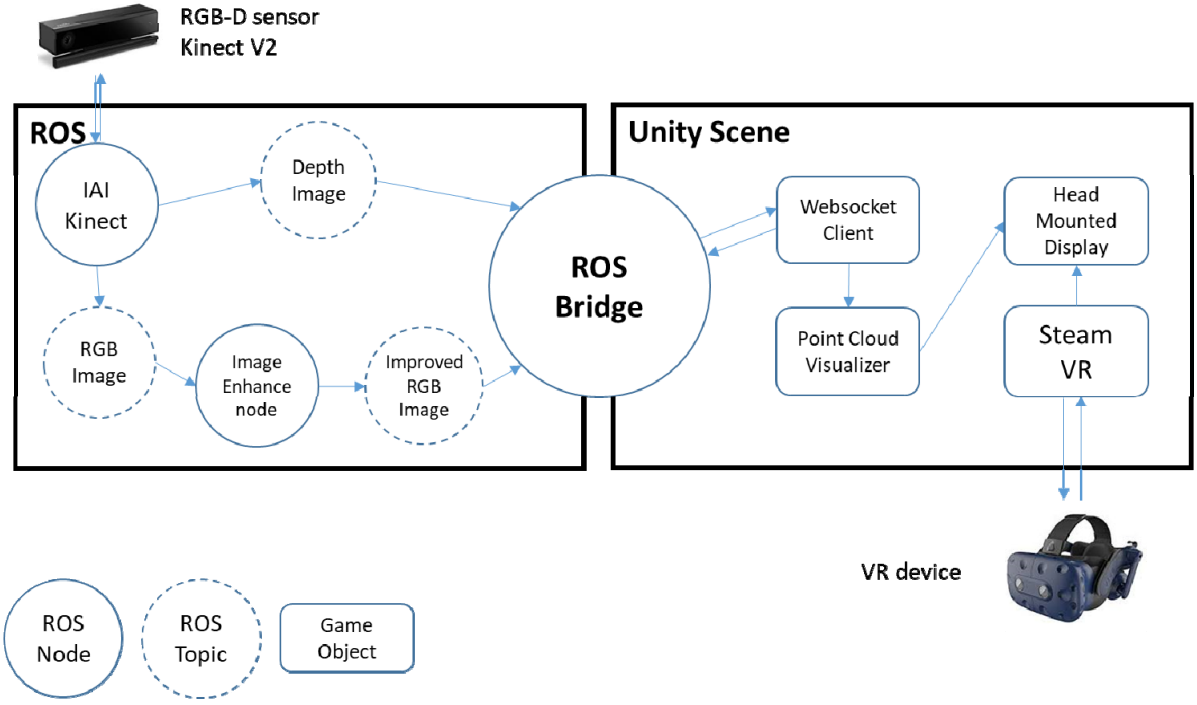


図 4.1　システムの概要図

ROS（Robot Operating System）[21]とは，ロボットアプリケーションプログラムの開発を支援するためのツールとライブラリのセットであるロボットソフトウェアプラットフォームである．ROS はノードと呼ばれる実行可能な最小単位のプロセスにより，全体の機能が実現される．各ノードは独立して並列的に動作し，ROS ネットワークと呼ばれるローカル TCP(Transmission Control Protocol)ネットワーク上で双方向のデータ送受信が可能である．ノードはネットワークを介してトピックとしてデータを配信するためのパブリッシャ，またはデータの購読を行うためのサブスクライバを作成する．

本研究では，ロボット作業環境下のコンピュータで Kinect v2 ROS ノード[24]，画像改善処理ノード，Rosbridge WebSocket サーバーを起動し，カラー画像および深度画像の取得，処理，データの通信を行う．

Unity

Unity は 2D，3D，および Virtual/Augmented/Mixed Reality アプリケーション開発に使用される人気の高いゲームエンジンである．接触力学や材料力学シミュレーションを処理できる物理エンジンを内蔵している．一般販売されている VR ハードウェアはいくつかのゲームエンジンと物理エンジンをサポートしているが，最もよくサポートされている開発プラットフォームは Unity である．Unity はカスタムシェーダーを記述するためのシェーダー言語を提供している．Unity のシーンには，基本的な構成要素であるゲームオブジェクトがあり，ゲームやオブジェクトの振る舞いに関する心臓部としての役割を果たすコンポーネントを所持している．コンポーネントはすべてのゲームオブジェクトの関数の一部であり，数十と多くの種類が存在するがその中でも最も重要な役割を果たすのがスクリプトである．スクリプトは，レンダリングフレームごとに

実行される小さな C#プログラムであり，提案するシステムではこれらの Unity スクリプトを用いて実装される．

スクリプトの構成

ROS Reality[22]システムを参考にして，ロボット作業環境再現システムは以下に説明する一連の C#スクリプトで構成される．

(1) WebSocket Client

このスクリプトは，Rosbridge クライアントとしてデフォルトで用いられる roslibjs[25]を C#スクリプトで実装したものである．これは，ROS トピックの公開，購読，配信をサポートしている．すべてのメッセージは JSON 形式で送受信され，その際画像データは Rosbridge の形式に従い，base64 でエンコードされる．

(2) Kinect PointCloud Visualizer

これは，Kinect からの RGB 画像と深度画像から点群を構築するためのスクリプトである． ROS 側のコンピュータから配信されている RGB 画像と深度画像トピックを購読し，RGB 画像と深度画像の各画素に対応する Quad を作成する．Quad の色は対応する RGB 画像の各画素値に対応させる．次に，カメラに対する Quad の位置を算出する必要がある．深度画像の各画素は，カメラからの距離をミリメートル単位で表したものである．したがって，最初にミリメートルからメートル単位に変換し，カメラに対する Quad の位置を計算する．次にその位置に Unity シーンの Kinect の変換行列を乗算し，Quad のワールド空間の位置を取得する．ワールド空間の位置は，最終的にビュー行列とプロジェクション行列によって乗算され，頂点シェーダーに渡される．

作業環境再現結果

図 4.2 に取得した RGB 画像および深度画像例を示す．また得られた画像情報をもとに Unity シーン上に再現を行った結果を図 4.3 に示す．

(a) RGB 画像　　(b) 深度画像

図 4.2　Kinect からの画像情報

図 4.3　Unity シーンへの再現結果

本研究では，ロボットの作業環境を計測するため，Kinect for Windows v2(Kinect v2)を使用した．Kinect v2 はカラー画像および深度画像の計測を目的としたマイクロソフト社が 2014 年にリリースした Kinect センサー及び SDK の総称である．深度情報については，Time of Flight の計測方法を使用する．Kinect v2 の外観を図 4.4 に，動作仕様を表 4.1 に示す[26]．

また，本論文では作業環境認識のためのインターフェースとして，VIVE Pro Head Mounted Display(HMD)を使用した．VIVE Pro HMD は，HTC Corporation がリリースした Virtual Reality Head Mounted Display 製品の一つである．VIVE Pro HMD の仕様を表 4.2 に，外観を図 4.5 に示す[27]．

表 4.1　Kinect v2 の動作仕様

カラー画像の解像度	最大 1920×1080[pixel]
カラー画像のフレームレート	30[fps]
深度画像の解像度	512×424[pixel]
深度画像のフレームレート	30[fps]
深度情報計測方式	ToF(Time of Flight)
深度情報認識範囲	500~ 8000[mm]
水平視野角	70 度
垂直視野角	60 度

図 4.4　Kinect v2 の外観

表 4.2　VIVE Pro HMD の仕様

スクリーン	デュアル AMOLED3.5 インチ（対角）
解像度	片目 1440×1660[pixel] 合計 2880×1600[pixel]
リフレッシュレート	90[Hz]
視野角	110 度
オーディオ	ハイレゾ対応ヘッドセットおよびヘッドフォン
入力	内臓マイク
接続	USB-C3.0，DP1.2，Bluetooth
センサー	SteamVR トラッキング，G センサー， ジャイロスコープ， 近接センサー，IPD センサー
人間工学	レンズ距離調整による瞳距離調整 調整可能な IPD，ヘッドフォン，ヘッドストラップ

図 4.5　VIVE Pro HMD の外観

開発環境と画像情報

本研究ではロボットの作業環境の計測，遠隔地で作業環境を再現するために 2 台の PC を使用した．それぞれの PC のスペックを表 4.3 に示す．また，図 4.6 に示すように，PC に接続し，Baxter Research Robot の頭部に固定した Kinect v2 センサーを用い，リアルタイムで処理を行う．その際の画像情報を表 4．4 に示す．

表 4.3　PC の主なスペック

OS	Ubuntu14.04LTS(Linux 64bit)	Microsoft Windows 10
CPU	Intel®Core™i7-3770	Intel®Xeon®W-2102
GPU	GeForce GTX 1070	GeForce GTX 1050 Ti
メモリ	16.0[GB]	16.0[GB]
開発環境	ROS Indigo	Unity2017 4.5f

表 4.4　画像データ

カラー画像	512×424[pixel]
深度画像	512×424[pixel]

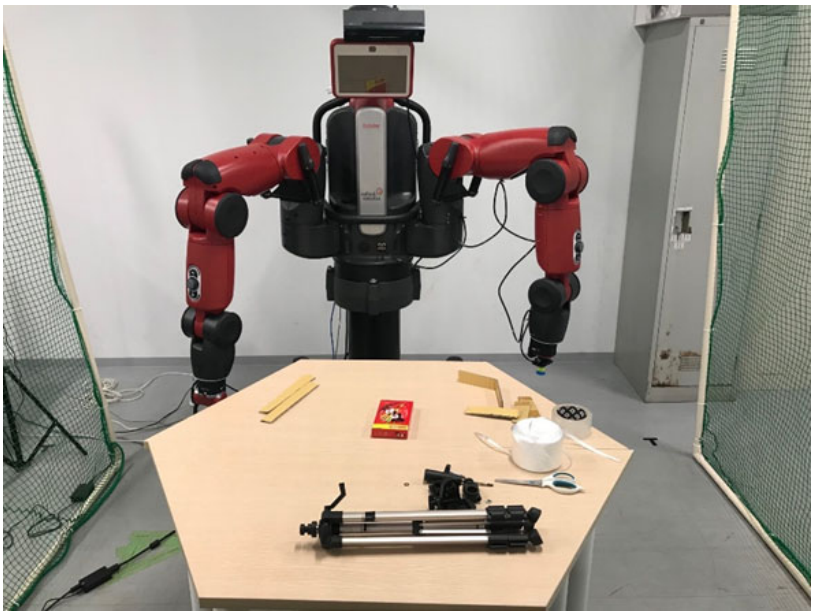
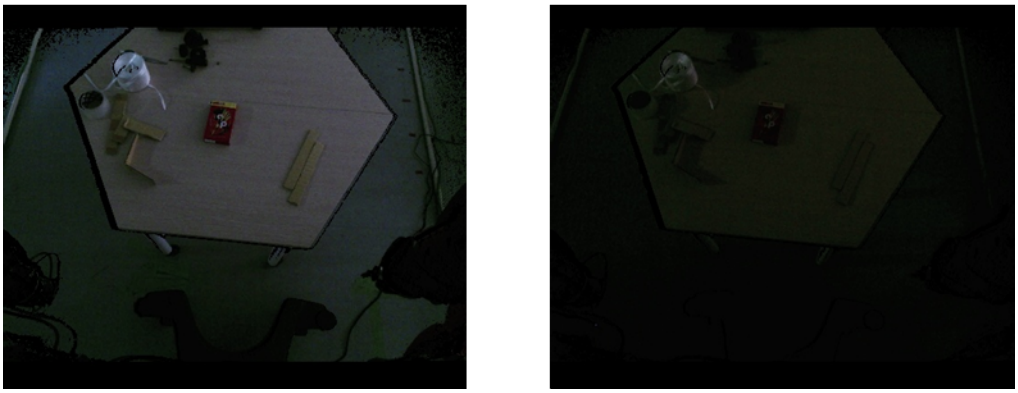


図 4.6　Kinect センサーの位置

実験内容

本研究で提案した Virtual Reality 技術および低光源下における映像改善処理を用いた遠隔操作システムの有用性を確認する．本研究では，2 つの異なる低光源下での作業環境下において作業環境の計測を行い，Virtual Reality 空間への再現を行った．その際，得られた画像データをもとにそのまま再現した場合と，映像改善処理を行い再現した場合で視覚的な比較を行った．2 つの異なる暗い環境下における実験環境の様子を図 4.7 に示す．この 2 パターンの実験環境は，部屋の明かりを全て消し，作業環境の光源を外からの環境光のみとした状態である．外光の時間的変化により 2 つのパターンの撮影を行った．また，2 パターンの作業環境下において，映像改善処理を行わずに Virtual Reality 空間へ再現を行った結果を，図 4.8，図 4.9 に示す．



(a) パターン 1　　(b) パターン 2

図 4.7　実験環境の様子

図 4.8　パターン 1 における作業環境再現結果

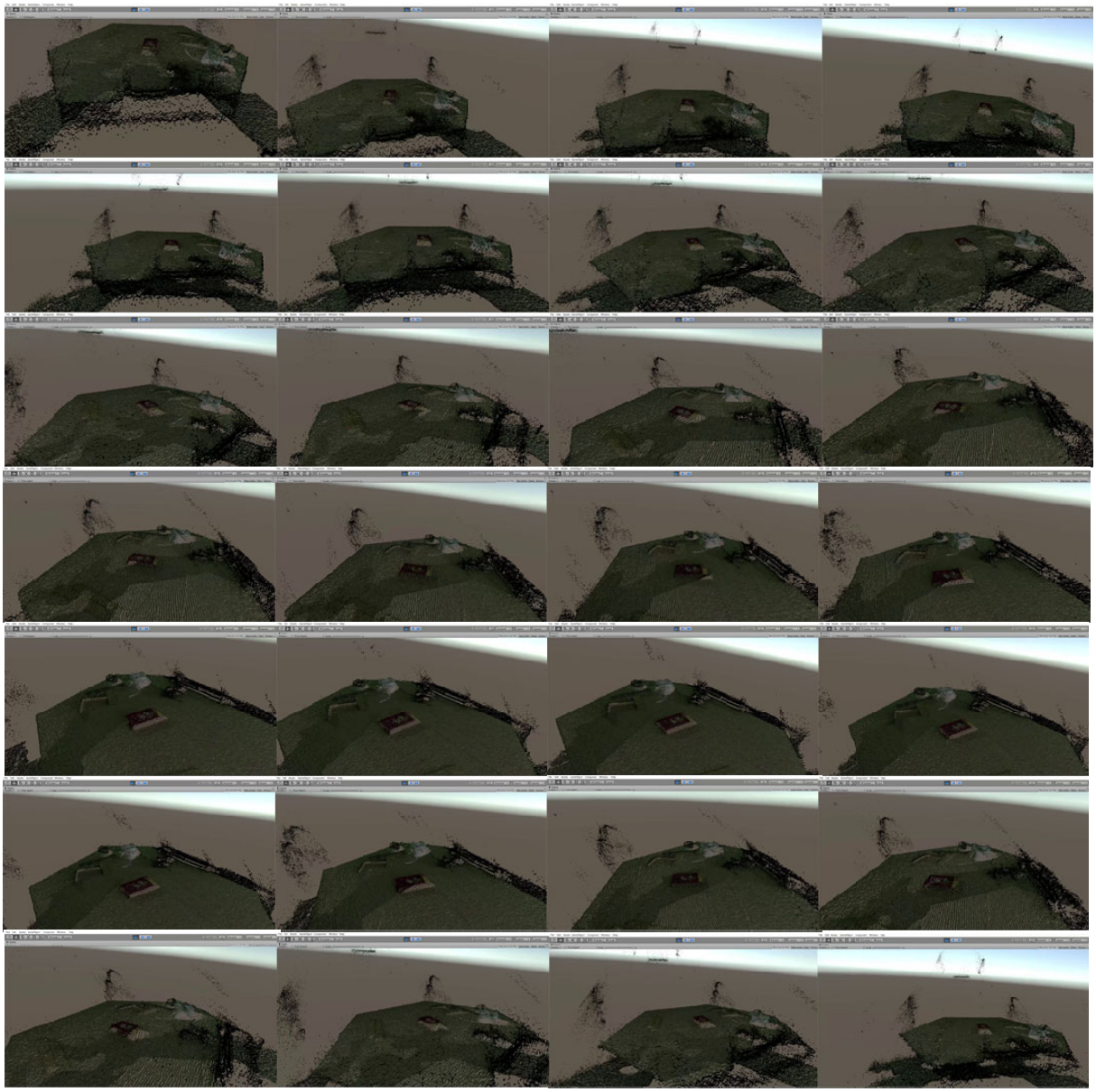
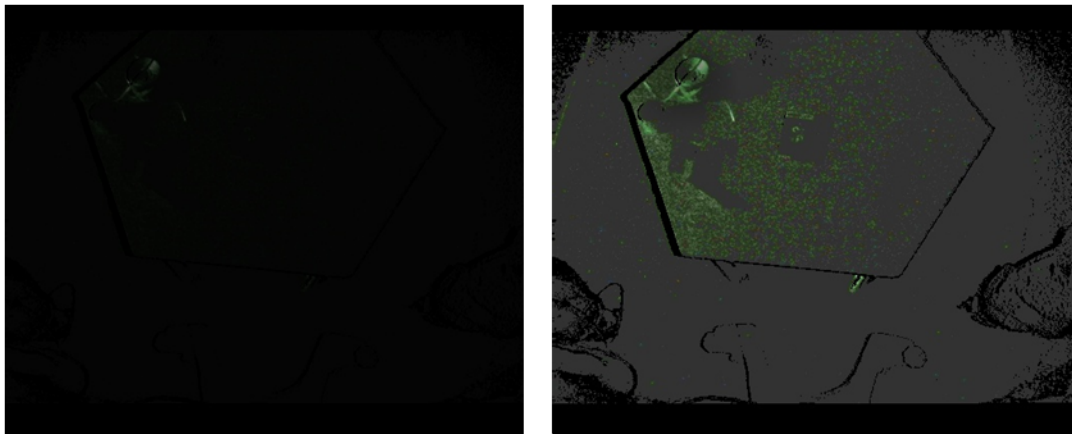


図 4.9　パターン 2 における作業環境再現結果

本研究では，Virtual Reality 技術および低光源下における映像改善処理を用いた遠隔操作システムにおける作業環境再現方法を提案した．前記で述べた映像改善処理手法については処理速度が 30[fps]を超えていたため，遠隔操作において重要なリアルタイム性能は保たれる．本実験では，2 パターンの暗い環境下において映像改善処理を行い，Virtual Reality 空間への再現を行った．パターン 1 の場合，暗い環境下ではあるが視認性はある程度保たれている．そのような環境下で改善処理を行った場合，明るい環境下での結果に近い再現結果が得られた．しかし，パターン 2 の場合は暗い環境下で視認性が低い．その場合，視認性の向上は見られるが，明るい環境下での結果とは色情報が異なる再現結果が得られた．これは，画像解析手法における環境光の推定の際に生じる問題であると考えられる．本研究では入力画像の平滑化処理を行った後，輝度値が最も高い画素の RGB 値を環境光と定めている．そのため，光源のほとんどない環境下における出力画像は，入力画像の色情報に左右されてしまい，結果として明るい環境下とは大きく異なる出力結果となったと考えられる．図 4.10 に示すように光源がほとんどない環境下で映像改善処理を行っても，視認性の向上はそれほど見込まれない．

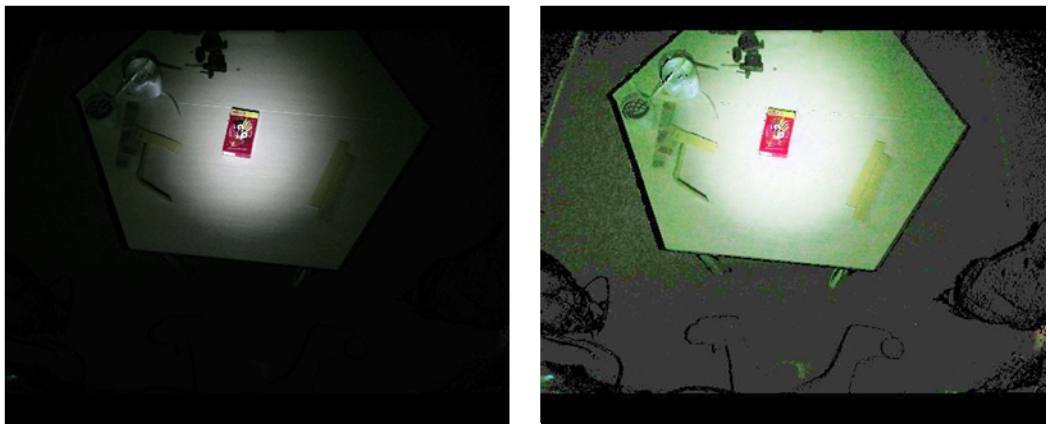


(a)入力画像　　(b)出力画像

図 4.10　光源がほとんどない環境下の結果

しかし，実際にロボットの遠隔操作を行う際は，作業環境周辺を照らすことが可能なライトをロボットに取り付けることを想定している．したがって，ライトの電池がなくなる，または，故障する場合以外に光源が全くないという状況は想定されにくい．そこで，ロボットに光源を取り付けることを想定し，図 4.10 と同じ環境下で簡易的なライトにより作業環境を照らした環境下で映像改善処理を行った結果,図 4.11 に示すように，視認性の向上が見られた．

また，比較結果を見ると，(a)の HE アルゴリズムを用いた方法では，入力画像の色情報に左右されることなく強調できている．しかし，画像によっては過度な強調が見られ，ノイズによる画質の劣化も生じている．一方，同図(c)，(d)のヘイズ除去アルゴリズムを用いた手法では，画像強調性能は HE アルゴリズムには劣っているように見えるが，過度な強調およびノイズによる画質劣化は抑えられている．また，本論文で用いた画像解析手法による結果と，画像の RGB 反転に Dark Channel Prior[16]を適用した結果を比べると，詳細部分の視認性に関しては同等であるが，画像の明るさは改善されていることがわかる．



(a)入力画像　　(b)出力画像

図 4.11　ライトのある環境下での結果

以上により，本研究で用いた画像解析手法の有用性を示したが，問題点も生じている．改善処理後の複数の画像において，不自然な色に強調されている画像が存在する．また，ロボットの操作の際に遠隔地の環境を認識しやすくするためにも，より明るい環境下の色情報に近づける必要がある．以上で述べた問題を解決するため，新たな画像解析手法を開発することが今後の課題である．

5. 結論と展望

本研究では，新たなロボットの遠隔操作システムの開発支援を目的としたロボットの作業環境再現方法を提案した．暗い環境下における視認性の向上のため，ヘイズ除去モデルを用いた映像改善処理を行い，Virtual Reality 技術を用いた遠隔地の作業環境再現を行った．

異なる複数の暗い環境下において映像改善処理を行い，Virtual Reality 空間への再現を行った場合と映像改善処理を行わなかった場合との再現結果との比較を行い，改善処理による視認性の向上を確認することができた．

今後の課題としては，暗い環境下における映像改善品質の向上を図るための画像解析手法の改善および新たな手法の開発[28-74]，また，本論文で提案したシステムを用いたロボット操作の実装およびその有用性の実証である．

謝辞

本研究を進めるにあたり，数多くの助言と，実験の実装をして頂きました上田　晃君に深く感謝いたします．

【参考文献】

1) 国土技術研究センター，“自然災害の多い国　日本”，http://www.jice.or.jp/knowledge/japan/commentary09#jump_02（アクセス: 2018.12.30）
2) 淺間一，“災害時に活用可能なロボット技術の研究開発と運用システムの構築”，日本ロボット学会誌，Vol. 32, No. 1, pp. 37-41, 2014.
3) 林鍾勳他，“暗所調査ロボットのための複数照明制御を用いた白とび，黒つぶれ画像の補正”，若葉研究者の集い５サマーセミナー，2015.
4) Kaiqi Huang et al., “A real-time object detecting and tracking system for outdoor night surveillance”, Pattern Recognition, Vol. 41, pp. 432-444, 2018.
5) Xuesong Jiang et al., “Night Video Enhancement Using Improved Dark Channel Prior”, in IEEE International Conference on Image Processing, pp. 553-557, 2013.
6) Xiaojie Guo, “LIME: A Method for Low-light Image Enhancement”, Proceedings of the 24th ACM international conference on Multimedia, pp. 87-91, 2016.
7) Liang Shen et al., “MSR-net: Low-light Image Enhancement Using Deep Convolutional Network”, arXiv:1711.02488, 2017.
8) Li Tao et al., “Low-light image enhancement using CNN and bright channel prior”, IEEE International Conference on Image Processing(ICIP), pp. 3215-3219, 2017.
9) 村瀬他，“ヒストグラム均等化法に基づく自然なコントラスト改善”，The Journal of the Institute of Image Electronics of Japan, Vol. 41, No. 2, pp. 124-130, 2012.
10) J. Alex Stark, “Adaptive Image Contrast Enhancement Using Generalizations of Histogram Equalization”, IEEE Transactions on Image Processing, Vol. 9, No. 5, pp. 889-896, 2000.
11) M. Abdullah-Al-Wadud et al., “A Dynamic Histogram Equalization for Image Contrast Enhancement”, IEEE Transactions on Consumer Electronics, Vol. 53, No. 2, pp. 593-600, 2007.
12) Tae Keun Kim et al., “Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering”, IEEE Transactions on Consumer Electronics, Vol. 44, No. 1, pp. 82-87, 1998.
13) 義 如，“悪条件下で撮影された画像の視認性改善に関する研究”，名古屋市立大学　博士学位論文，2017.
14) Z. Rahman et al. “Multi-Scale retinex for color image enhancement”, IEEE International Conference on Image Processing, Vol. 3, pp. 1003-1006, 1996.

15) Xuan Dong et al, “Fast efficient algorithm for enhancement of low lighting video”, IEEE International Conference on Multimedia and Expo, pp. 1-6, 2011.
16) K. He et al, “Single Image Haze Removal Using Dark Channel Prior”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, pp. 2341-2353, 2011.
17) Yuen Peng Loh et al., “Getting to know Low-light with The Exclusively Dark Dataset”, Computer Vision and Image Understanding, Vol. 178, pp. 30-42, 2019.
18) K. Ma et al., “Perceptual evalutin of single image dehazing algorithms”, IEEE International Conference on Image Processing, pp. 3600-3604, 2015.
19) K. Goldberg et al., “Desktop teleoperation via the world wide web”, IEEE International Conference on Robotics and Automation, Vol. 1, pp. 654–659, 1995.
20) Jean Vertut, Teleoperation and Robotics: Applications and Technology, Springer Science & Business Media, 2013.
21) 倉爪亮他, ROS ロボットプログラミングバイブル, オーム社, 2018.
22) David Whitney et al., “ROS Reality: A Virtual Reality Framework Using Consumer-Grade Hardware for ROS-Enabled Robots”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1-9, 2018.
23) Christopher Crick et al., “Rosbridge: ROS for Non-ROS Users”, Robotics Research, pp. 493-504, 2017.
24) Thiemo Wiedemeyer, “IAI Kinect2” https:// github.com/code-iai/iai_kinect2 (アクセス: 2019.1.30)
25) Russell Toris et al., “Robot web tools: Efficient messaging for cloud robotics”, IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4530-4537, 2015.
26) 中村他, KINECT for Windows SDK プログラミング Kinect for Windows v2 センサー対応版, 秀和システム, pp. 2-6, 2015.
27) HTC, “Vive pro” https://www.vive.com/jp/product/vive-pro/ (アクセス: 2019.2.5)
28) M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3213-3223, 2016.
29) M. Schwarz, A. Milan, C. Lenz, A. Munoz, A. S. Periyasamy, M. Schreiber, S. Schüller, and S. Behnke, “Nimbro picking: Versatile part handling for warehouse automation,” Proc. of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3032-3029, 2017.
30) O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” The International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science, Vol. 9351, pp. 234-241, 2015.
31) R. T. Azuma, “A survey of augmented reality,” Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pp. 355-385, 1997.
32) J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, 2015.
33) L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” The IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol.40, No.4, pp.834-848, 2017.
34) S. Zagoruyko and N. Komodakis, “Wide residual networks,” Proc. of the British Machine Vision Conference (BMVC), pp. 87.1-87.12, 2016.
35) K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Proc. of the International Conference on Learning Representation (ICLR), pp.1-14, 2015.
36) K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

37) F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251-1258, 2017.
38) F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” Proc. of the International Conference on Learning Representations (ICLR), 2016.
39) S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987-5995, 2017.
40) W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1874-1883, 2016.
41) G. J. Brostow and R. Cipolla, “Semantic object classes in video: A high-definition ground truth database,” Pattern Recognition Letters, Vol. 30, No. 2, pp. 88-97, 2009.
42) H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” Proc. of the IEEE International Conference on Computer Vision (ICCV), pp. 1520-1528, 2015.
43) V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” The IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 39, No. 12, pp. 2481-2495, 2017.
44) L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” Proc. of the International Conference on Learning Representations (ICLR), 2015.
45) P. Krähenbühl and V. Koltun, “Efficient inference in fully connected crfs with gaussian edge potentials,” Advances in Neural Information Processing Systems (NIPS), pp. 109-117, 2011.
46) S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, P. H. S. Torr, “Conditional random fields as recurrent neural networks,” Proc. of the IEEE International Conference Computer Vision (ICCV), pp. 1529-1537, 2015.
47) Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, “Semantic image segmentation via deep parsing network,” Proc. of the IEEE International Conference on Computer Vision (ICCV), pp. 1377-1385, 2015.
48) W. Liu, A. Rabinovich, and A. C. Berg, “Parsenet: Looking wider to see better,” Proc. of the International Conference on Learning Representations Workshop (ICLRW), 2016.
49) H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2881-2890, 2017.
50) L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, “Rethinking atrous convolution for semantic image segmentation,” arXiv preprint arXiv:1706.05587, 2017.
51) L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” Proc. of the European Conference on Computer Vision (ECCV), pp. 833-851, 2018.
52) D. Han, J. Kim, and J. Kim, “Deep pyramidal residual networks,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6307-6315, 2017.
53) S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” Proc. of the International Conference on Machine Learning (ICML), Vol. 32, pp. 448-456, 2015.
54) V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” Proc. of the International Conference on Machine Learning (ICML), pp. 807-814, 2010.
55) J. Redmon and A. Farhadi, “Yolo9000: Better, faster, stronger,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517-6525, 2017.
56) N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” Journal of Machine Learning Research, Vol. 15,

pp. 1929-1958, 2014.
57) W. Shi, J. Caballero, L. Theis, F. Huszar, A. Aitken, C. Ledig, and Z. Wang, “Is the deconvolution layer the same as a convolutional layer?” arXiv preprint arXiv:1609.07009, 2016.
58) A. P. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, “Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize,” arXiv preprint arXiv:1707.02937, 2017.
59) S. Tokui, K. Oono, S. Hido, and J. Clayton, “Chainer: A next-generation open source framework for deep learning,” Workshop on Machine Learning Systems at Neural Information Processing Systems (NIPS), 2015.
60) I. Loshchilov, and F. Hutter, “Sgdr: Stochastic gradient descent with warm restarts,” Proc. of the International Conference on Learning Representations (ICLR), 2017.
61) A. Kendall, V. Badrinarayanan, and R. Cipolla, “Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding,” arXiv preprint arXiv:1511.02680, 2015.
62) A. Kundu, V. Vineet, and V. Koltun, “Feature space optimization for semantic video segmentation,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3168-3175, 2016.
63) S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,” the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1175-1183, 2017.
64) M. A. Islam, M. Rochan, S. Naha, N. D. B. Bruce, and Y. Wang, “Gated feedback refinement network for dense image labeling,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4877-4885, 2017.
65) P. Bilinski and V. Prisacariu, “Dense Decoder Shortcut Connections for Single-Pass Semantic Segmentation,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6596-6605, 2018.
66) TEDx Talks (28. 11. 2017). How Self-Driving Cars Work | David Silver | TEDxWilmingtonSalon [Video File]. Retrieved from https://www.youtube.com/watch?v=Ly92UcnoEMY.
67) D. A. Pomerleau, “Alvinn: An autonomous land vehicle in a neural network,” Advances in Neural Information Processing Systems (NIPS), pp. 305-313, 1989.
68) M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, “End to end learning for self-driving cars,” arXiv preprint arXiv:1604.07316, 2016.
69) H. Xu, Y. Gao, F. Yu, and T. Darrell, “End-to-end learning of driving models from large-scale video datasets,” arXiv preprint arXiv:1612.01079, 2016.
70) Y. Chen, P. Palanisamy, P. Mudalige, K. Muelling, and J. M. Dolan, “Learning on-road visual control for self-driving vehicles with auxiliary tasks,” arXiv preprint arXiv:1812.07760, 2018.
71) A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, “Flownet: Learning optical flow with convolutional networks,” Proc. of the IEEE International Conference on Computer Vision (ICCV), pp. 2758-2766, 2015.
72) S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, Vol. 9, No. 8, pp.1735-1780, 1997.
73) Udacity self-driving car simulator, https://github.com/udacity/self-driving-car-sim (accessed 31. 1. 2019).
74) E. Santana and G. Hotz, “Learning a driving simulator,” arXiv preprint arXiv:1608.01230, 2016.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Output Bounded and RBFNN-based Position Tracking and Adaptive Force Control for Security Tele-surgery	ACM Transactions on Multimedia Computing Communications and Applications	2020.05
Deep-sea Organisms Tracking Using Dehazing and Deep Learning	Mobile Networks and Applications	2020.04
Low Illumination Underwater Light Field Images Reconstruction Using Deep Convolutional Neural Networks	Future Generation Computer Systems	2018.01
Extreme ROS Reality: A Representation Framework for Robots 9 Using Image Dehazing and VR	3rd EAI International Conference on Robotic Sensor Networks	2019.08