

実環境 3D データ処理の実現に向けて

Realize of Real Environment 3D Data Processing

武井 宏将

要約 デバイス技術の進展により、実環境の空間情報をデジタルデータとして取得することが可能になった。このデータを実環境 3D データと呼ぶこととする。日本ユニシスは、実環境 3D データを処理するための技術研究を行っている。

実環境 3D データは、実環境をセンシングするのみで取得できるため、データ取得コストは安価である。一方、実環境 3D データは、内部に様々な要素が混在しており、それらを分離する情報は持たない。そのため、実環境 3D データを処理するためには、データの中から意味のあるまとまりを取得するコンテンツ認識技術を確立する必要がある。

実環境をセンシングするデバイスの特徴の一つは二次元画像と三次元点群の両方が取得できることである。そこで我々は、実環境 3D データ処理におけるコンテンツ認識技術確立のために、二次元画像と三次元点群の両方を利用したアプローチを検証している。本稿において、二次元画像と三次元点群のデータ特性を基に、これらのデータを組み合わせた処理の有効性について考察する。また、両方のデータを組み合わせる有効性について研究事例を挙げて紹介する。

本研究により開発した技術は、IoT や Industry4.0 において、空間センシングデータを処理するための技術として、重要な役割を果たすと考えている。

Abstract Owing to the progress of recent device technology, we can get the space information of real environment as digital data. We call this data as “real environment 3D data”. Nihon Unisys has been making research for data processing technology of real environment 3D data.

We can get real environment 3D data, only by sensing the real environment. As a result, we can get this data at low cost. However real environment 3D data consist of various elements and these elements do not have unique information which can be used to separate them. Thus we need to establish a contents recognition technology which to take out the meaningful gatherings of elements from data to process the real environment 3D data.

The one of the feature of the recent device which senses the real environment is that the device can get both 2D image and 3D point cloud. We validate the approach using both 2D image and 3D point cloud for establishment of content recognition technology. In this paper, we discuss the effective combination of 2D image and 3D point cloud based on feature of these data. Moreover, we introduce the research example of the effective combination of these data.

We believe that our technology plays an important role in processing space sensing data in IoT and Industry 4.0.

1. はじめに

デバイス技術の進展により、実環境の空間情報をデジタルデータ（以降、実環境 3D データと呼ぶ）として取得することが可能になった。例えば、図 1 左は、中長距離レーザ測定機によりプラント施設を測定した三次元点群データである。また、図 1 中は、Mobile Mapping System により道路を走行しながら測定した色付き三次元点群データである。図 1 右は、Kinect により室内を測定した色付き三次元点群データである。このように、実環境をセンシングすることで実環境 3D データを取得できるデバイスが続々と現れてきている。



図 1 実環境から取得したデジタルデータ
(左：プラント施設^[1]，中：道路^[2]，右：室内)

実環境 3D データを取得できるようになると、実環境の空間情報を直接コンピュータで処理することができるようになる。これにより、実環境とコンピュータを結びつけた新しい仕組み作りが可能になると期待できる。

一方で、実環境からセンシングすることで取得したデジタルデータを処理するためには、技術的に解決しなければならない課題もある。特に重要な課題は、実環境 3D データからの「コンテンツ認識」である。ここでコンテンツ認識とは、データの中から意味のあるまとまりを認識する技術を指す。コンテンツ認識技術を有することで、実環境 3D データを利用した様々なアプリケーションが実現できるようになる。日本ユニシスは、このような実環境からセンシングすることで得られた実環境 3D データ処理の実現に向けて技術研究を行っている。

本稿は以下の構成をとる。2 章は、実環境をセンシングして空間情報を取得するデバイスについて紹介する。3 章は、実環境をセンシングすることにより取得した三次元デジタルデータの特徴について述べ、実環境 3D データを扱うための技術的課題について考察する。そして、実環境 3D データ処理に対する我々のアプローチについて述べる。4 章で、我々の研究成果について紹介し、5 章でまとめる。

2. 実環境センシングデバイス

実環境をセンシングすることで空間情報を三次元点群として取得する。三次元点群とは、 (x, y, z) の座標値を持つ点の集まりとして空間情報を表現するデータ形式である。

図 2 左は、室内で撮影した三次元点群データである。このデータの一箇所をズームすると図 2 右のように、多数の点の集まりとして表現されていることがわかる。実環境をセンシングするデバイスの特徴の一つに、三次元点群と合わせて二次元画像も取得できることが挙げられる。そのようなデバイスとして、中長距離レーザ測定機、Mobile Mapping System、コンシューマ向けデバイス（Kinect、Intel RealSense）を以下に紹介する。

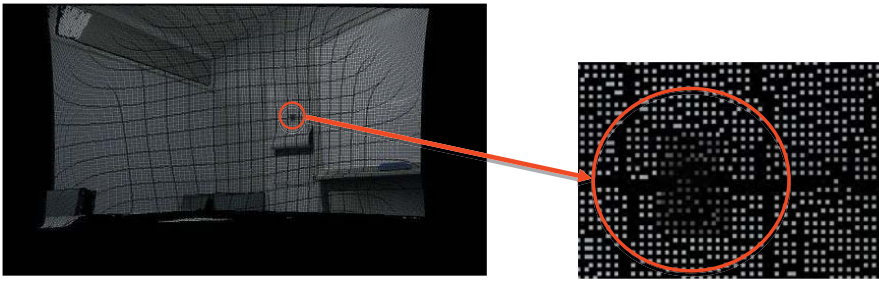


図2 三次元点群データ

1) 中長距離レーザー測定機



図3 中長距離レーザー測定機 (FARO^[31])

中長距離レーザー測定機 (図3) は、測定機からレーザーを放射し、対象物からのレーザーの跳ね返りを受信することで、測定対象までの距離を測定する。中長距離レーザー測定機は、FARO社 Focus 3D X 330 (2013年販売開始) や Leica社 HDS 7000 (2011年販売開始) 等がある。これらのレーザー測定機は一度に広い範囲を測定することが可能であり、FARO社 Focus 3D X 330 は最大 330m、Leica社 HDS 7000 は最大 180m の範囲を測定することができる。そのため、工場やプラント、橋梁等の測定といった用途に利用されている。秒間で 100 万点程度の点群を取得し、精度は 25m 離れた場所で 1mm 程度の誤差と非常に高精度である。中長距離レーザー測定機を利用することで、実環境を非常に高密度かつ高精度に取得できる。

2) Mobile Mapping System



図4 Mobile Mapping System (三菱)^[2]

Mobile Mapping System (図4) は、自動車にデジタルカメラ・3D レーザセンサー・GPS を搭載し、道路を走行しながら周辺環境を計測する装置である。一度の計測で非常に

広範囲のデータを取得できる。精度は数 cm ～数 10cm 程度の誤差といわれている。道路周辺のデジタル化や三次元地図の作成に利用されている。

3) コンシューマ向けデバイス (Kinect, Intel RealSense)



図5 コンシューマ向けデバイス (左: Kinect^[4] 右: Intel RealSense^[5])

従来、三次元測定機は非常に高価なデバイスであった。しかし、近年、低価格なデバイスが登場している。Kinect (図5左) は、Microsoft が2010年にゲーム用に開発したセンサーであり、二次元画像と三次元点群を取得することができる。Intel RealSense (図5右) は、Intel が2015年より提供している3D デプスカメラである。これらのデバイスは、ゲームにおける入力やナチュラルユーザインターフェースの入力として利用されている。今後も様々な低価格三次元測定デバイスが登場することが予想される。

3. 実環境 3D データ処理の実現に向けて

本章は、実環境 3D データ処理実現のための技術的課題とその解決に向けたアプローチについて述べる。2章で述べたとおり、実環境から二次元画像と三次元点群の両方を取得できるデバイスが増えてきている。よって、これらのデータを組み合わせることで、これまでよりも高精度な処理が実現できる可能性がある。3.1節では、実環境 3D データを利用するメリットと技術的課題について考察する。3.2節では、技術的課題の解決に向けた二次元画像と三次元点群の両方を利用したアプローチの有効性についてこれらのデータ特性を基に考察する。

3.1 実環境 3D データを利用するメリットと技術的課題

はじめに、実環境をセンシングして取得した三次元点群データと三次元 CAD でモデリングされた三次元 CAD データを比較することで、実環境 3D データの特性を考察する。三次元 CAD は、デジタル三次元データを作成するツールとして幅広く利用されている (図6)。

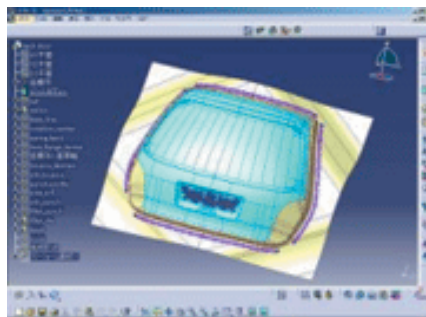


図6 三次元 CAD によりモデリングしたデータ (Dynavista^[6])

まず、データ取得コストについて考察する。三次元 CAD データは、人がモデリングして作成するので、時間と工数がかかり、データ作成のコストは大きい。一方、実環境 3D データは、実環境を一括でセンシングすることで、短時間かつ安価にデータを取得できる。

次にデータへのコンテンツ情報の付与について考察する。ここでコンテンツとは、データ中の意味のあるまとまりを指す。多くの三次元 CAD データは、データ自体がある部品を表している。また、モデリング工程の中で、データ中のある部分に意味を付与することが可能である。つまり、三次元 CAD データはコンテンツ情報の付与が可能である。一方、実環境 3D データは、実環境を一括でセンシングして取得する。そのため、データの中には様々な要素が混在しており、またそれらを分離する情報を付与することは困難である。以上をまとめると表 1 のようになる。

表 1 実環境 3D データと三次元 CAD データから得られる三次元データの特性比較

	実環境 3D データ	三次元 CAD データ
データ作成コスト	実環境から一括取得するため、短時間・安価に取得できる。	三次元 CAD によるモデリングが必要であるため、大きな時間および工数がかかる。
コンテンツ付与	コンテンツの付与は困難	コンテンツの付与は容易

ここで、実環境 3D データ処理に必要な要素を考えるために、実環境 3D データがどのように利用されるかを想定する。我々は、実環境 3D データが利用されるアプリケーションの処理イメージを以下のように想定している (図 7)。

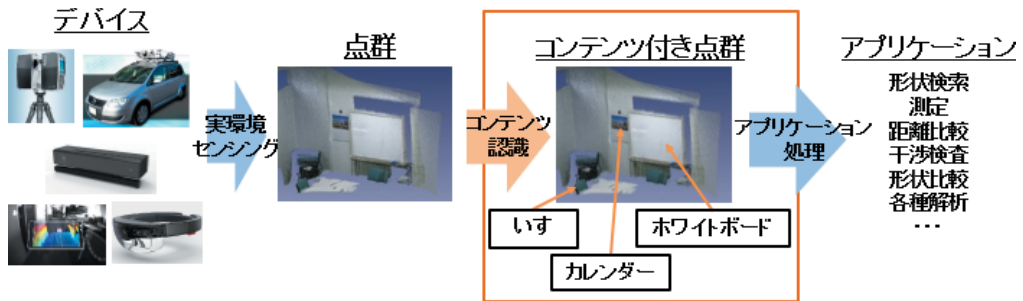


図 7 実環境 3D データを利用したアプリケーション処理イメージ

デバイスから、実環境をセンシングすることで、二次元画像および三次元点群を取得する。二次元画像および三次元点群は、それぞれ色および座標値の集まりに過ぎない。実環境 3D データをアプリケーションで利用するためには、色や座標値の集まりから意味のあるまとまった単位 (コンテンツ) を認識することが必要となる。コンテンツが認識できると、実環境 3D データから意味のあるまとまりを取り出すことができる (図 7 では、いす、カレンダー、ホワイトボード等)。意味のあるまとまりをデータとして取り出すことができると、ある対象に対する測定や距離比較、干渉検査といった処理ができるようになり、様々なアプリケーションを実現することができる。つまり、実環境 3D データをアプリケーションで扱うには、色の集まりおよび点群から意味のあるまとまった単位を認識するコンテンツ認識が必要となる。

3.2 二次元画像と三次元点群を組み合わせたデータ処理の有効性

前節にて、実環境3Dデータを活用するためには、実環境3Dデータに対するコンテンツ認識技術が必要となることを述べた。2章で述べたとおり、実環境をセンシングするデバイスの多くは、三次元点群と合わせて二次元画像も取得することができる(図8)。そこで我々は、二次元画像と三次元点群を組み合わせた実環境3Dデータに対するコンテンツ認識技術の研究を行っている。本節では、二次元画像と三次元点群のデータの特徴に基づき、これらのデータを組み合わせたデータ処理の有効性について考察する。これまで画像処理と三次元形状処理は、それぞれ独立の領域において成長してきた。我々は、この両方の領域で発展してきた技術を混成して利用することで、高精度なコンテンツ認識が実現できると考えている。

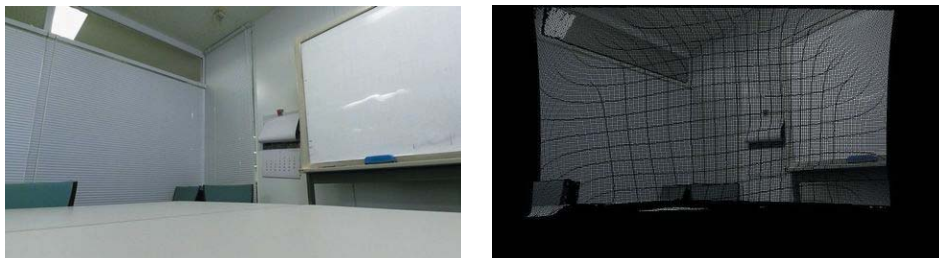


図8 同時に撮影した二次元画像と三次元点群(左:二次元画像, 右:三次元点群)

二次元画像と三次元点群を両方利用する有効性を考察するために、各データの特徴を見ていく。二次元画像は、多くのテクスチャ(模様)を含むが、長さや奥行きといった空間情報を正しく持たない。一方、三次元点群は正確な空間情報を持つが、テクスチャのような情報をほとんど持たない。なぜなら、多くの三次元点群の形状は、十分近傍では平面に近い形をしているためである(表2)。

表2 二次元画像と三次元点群の特徴の違い

	長所	短所
二次元画像	テクスチャ情報が豊富	空間の情報が不正確
三次元点群	空間の情報を正確に持つ	テクスチャ情報が少ない

画像間のマッチングを取る処理やテクスチャに基づく認識処理は、二次元画像の方が三次元点群よりもやりやすい。しかし、二次元画像は空間情報を正確に持たないため、二点間の距離測定や物体の前後関係把握などは困難である。一方、二点間の距離測定や物体の前後関係把握といった処理は、三次元点群の方が二次元画像よりもやりやすい。しかし、三次元点群はテクスチャ情報が少ないため、点群間のマッチングをとる処理や認識処理は困難である(表3)。

表 3 二次元画像・三次元点群の得意・不得意な処理

	得意な処理	不得意な処理
二次元画像	マッチング・認識	二点間距離測定・前後関係認識
三次元点群	二点間距離測定・前後関係認識	マッチング・認識

このように、二次元画像と三次元点群はそれぞれ双補完的な特徴を持つ。よって、二次元画像と三次元点群の二つのデータを組み合わせることで、お互いに得意な部分を活かした処理を行うことが可能となる。2010 年位までは、二次元画像と三次元点群を組み合わせたデータを取得できるデバイスがほとんどなかった。そのため、二次元画像と三次元点群を組み合わせた処理技術はあまり発展してこなかった。しかし、2010 年以降、Kinect や中長距離レーザ測定機が点群と合わせて画像も測定できるようになったことで、二次元画像と三次元点群を同時に取得できるデバイスが増えてきた。これにより、画像処理と点群処理を組み合わせ、これまでよりも高度なデータ処理が実現できる可能性が出てきた。例えば、点群のセグメンテーションにおいて、画像を利用することで隣接関係を取得することが容易になり、点群のみを利用する場合と比べて、より高精度なセグメンテーションが可能となる。また、点群の位置合わせにおいて、画像特徴点の対応関係を利用することで、点群のみを利用する場合と比べて、より高精度な位置合わせが可能となる。

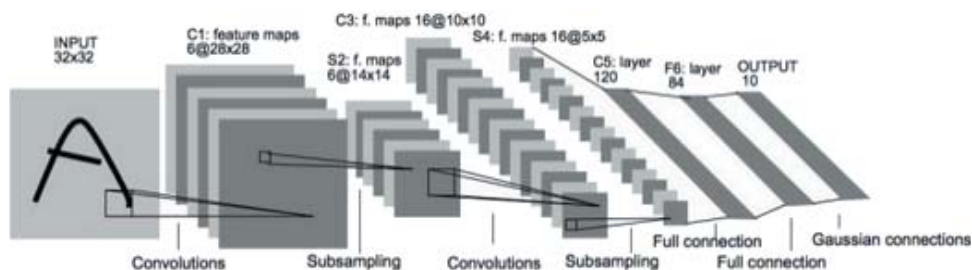
4. 研究事例紹介

本章では、コンテンツ認識に対する我々の取り組みの一つとして、三次元点群による画像補正と二次元画像認識を組み合わせた精度向上手法について紹介する。二次元画像と三次元点群を比較すると、認識処理は二次元画像の方がやりやすい。そこで認識手法として、深層学習を用いた画像認識技術を検証した。深層学習とは、多層のニューラルネットワークを利用した機械学習手法のことであり、画像認識において高い精度がでることが示されている^[7]。図 9 は、深層学習を利用してマークが映っている画像を認識し、マークが示すメーカー名を表示するデモである。深層学習は、音声認識や自然言語処理といった画像処理以外の分野でも有効な結果が示されており、非常に注目を浴びている。



図 9 深層学習による画像認識

本稿では、教師ありの深層学習アルゴリズムの一つである Deep Convolution Neural Network^[8]を用いた。Deep Convolution Neural Network は、入力データとして画像全体を入力し、学習を行うアルゴリズムである (図 10)。

図 10 Deep Convolution Neural Network^[8]

深層学習が利用される以前の画像認識に使われる手法として、Bag of features^[9]等の画像特徴点を利用した手法が知られている。ここで、画像特徴点とは画像中の輝度変化の大きい部分を指し、SIFT^[10]、SURF^[11]、ORB^[12]等多数の画像特徴点抽出手法が知られている（図 11）。

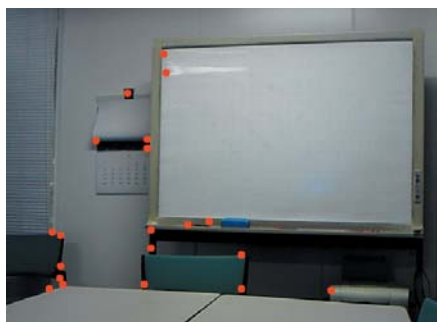


図 11 画像特徴点

画像特徴点を利用した認識手法と比較して、深層学習を利用した認識手法が優れている点は以下の二点である。

1) 画像特徴点を利用するには、テクスチャが豊富にある必要があるため、単純なマークのような形状の認識には不向きである。また、画像内でテクスチャの疎密がある場合、テクスチャの密な部分からは画像特徴点が大量に抽出されるが、疎な部分からはほとんど画像特徴点が抽出されない。図 11 を見ると、いすやホワイトボードの枠付近には画像特徴点が見られるが、ホワイトボードの面の部分や壁面には画像特徴点が見られない。そのため、認識結果がテクスチャの密な部分に引きずられる傾向がある。Deep Convolution Neural Network を利用した認識手法は、画像全体を入力として、学習および判定を行う。そのため、画像特徴点の抽出状況の影響はほとんど受けない。

2) 類似のテクスチャが繰り返されるような画像においては、類似の画像特徴点が多数抽出され、それらは区別することができない。そのような画像には、画像特徴点を利用した認識手法は不向きである。Deep Convolution Neural Network を利用した手法は、画像全体を利用するため、このような問題による影響はほとんど受けない。

我々は、図 9 のような自動車会社のロゴを使った画像認識の実験により二つの手法の認識精度の検証を行った。ロゴは 7 種類用意した。検証の結果、Bag of features を用いた認識手法が 25% 程度の認識精度であったのに対して、Deep Convolution Neural Network を用いた認

識手法は 98% 程度の認識精度を達成した。本検証で用いた画像は、シンプルな形状であり、画像認識としては難しい分類に入ることから、Deep Convolution Neural Network を用いた認識手法の精度の良さを示している。

一方で、多くの画像処理手法と同様に、Deep Convolution Neural Network による認識手法も、認識対象に大きな角度変化がある場合、認識精度が低下することがある。そこで、前処理として三次元点群を利用して画像補正を行い、画像認識の精度を向上させる手法を検証した。画像に写った認識対象付近の三次元点群を用いて、認識対象付近の法線方向を算出する。この法線が、カメラのビュー方向を向くように変換した画像を作成することで、認識対象が正面を向いた補正画像を作成することができる。

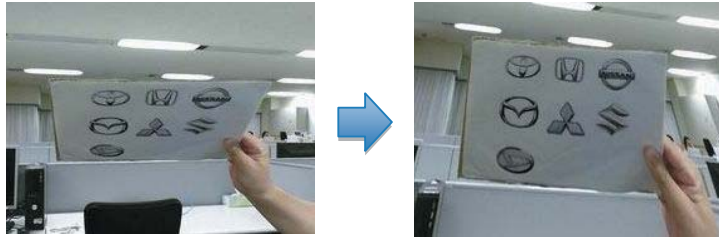


図 12 三次元点群を利用した画像補正

実際の画像補正を行った画像が図 12 である。図 12 左の、カメラに対して大きな角度変化をした画像が、画像補正により正面を向いていることがわかる (図 12 右)。

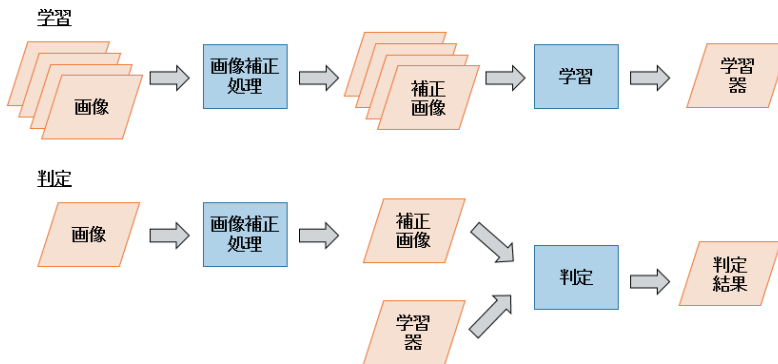


図 13 本処理の処理フロー

図 13 は、本処理のフローを表している。画像認識は、学習と判定の二つのステップに分かれる。学習ステップでは、はじめに、入力画像に対して図 12 に示した三次元点群を使った画像補正を行う。次に、補正画像を用いて学習を行う。判定ステップでは、はじめに、入力画像に対して点群を使った画像補正を行う。次に、補正画像を学習器に入力して判定を行う。この補正画像を用いて、深層学習により画像認識を行うことで、補正を行わない場合と比較して、認識精度を向上することが可能となる。

5. おわりに

本稿は、実環境 3D データを取得するデバイスの紹介と実環境 3D データにより取得されるデータの特性および技術的課題について述べた。さらに、二次元画像と三次元点群を両方利用したデータ処理の有効性について、それぞれのデータ特性を基に考察した。我々は、二次元画像と三次元点群を両方利用したデータ処理は有効な技術と捉えており、引き続きこの方向の研究を進めていく予定である。

最後に、本研究と IoT および Industry4.0 との関係について述べる。IoT は、モノのインターネットと呼ばれ、様々なモノがネットワークを介して繋がる世界を指す。Industry4.0 は、工場におけるモノとモノがネットワークを介して繋がる世界を指す。この両者において、実環境をセンシングするデバイスを用いることで、ネットワークを介して実環境とコンピュータを繋ぐことができる。本研究は、デバイスにより取得した実環境の空間情報を処理するための技術研究であり、IoT や Industry4.0 においても、有効な役割を果たす技術の一つになると考えている。

近年、デバイスの進展はさらに進んでおり、実環境 3D データが手軽に手に入る環境は、より身近になってきている。実環境 3D データの利用が一般的になった時に、本研究の成果が幅広い領域で活用できるよう技術研究を進めていきたいと考えている。

最後に、本稿執筆にあたりご指導頂きました松林 毅殿に心より感謝申し上げます。また、本研究を進めるにあたりご議論頂きました多くの方々に、心より感謝申し上げます。

-
- 参考文献 [1] 3D レーザー計測サービス, 東芝プラントシステム株式会社
http://www.toshiba-tpsc.co.jp/services/3d_laser_04.htm
 [2] 三菱モービルマッピングシステム, 三菱電機
<http://www.mitsubishielectric.co.jp/mms/tokucho.html>
 [3] FARO Laser Scanner Focus3D, FARO ジャパン
<http://www.faro.com/ja-jp/products/3d-surveying/laserscanner-faro-focus-3d/overview>
 [4] Kinect, 日本マイクロソフト株式会社
<http://www.xbox.com/ja-JP/kinect>
 [5] Intel RealSense, インテル株式会社
<http://www.intel.co.jp/content/www/jp/ja/architecture-and-technology/realsense-overview.html>
 [6] Dynavista, 日本ユニシス株式会社
<http://dynavista.unisys.co.jp/>
 [7] A. Krizhevsky I. Sutskever G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Proc. of NIPS*, p.p.1097-1105, 2012.
 [8] Y. LeCun L. Bottou Y. Bengio P. Haffner, "Gradient-based learning applied to document recognition", *Proc. of the IEEE* 86 (11), p.p.2278-2324, 1998.
 [9] G. Csurka C. R. Dance L. Fan C. Bray, "Visual categorization with bags of key-points", *Proc. of ECCV*, p.p.1-22, 2004.
 [10] D. G. Lowe, "Object recognition from local scaleinvariant features", *Proc. of ICCV*, p.p.1150-1157, 1999.
 [11] H. Bay T. Tuytelaars L. Van Gool, "Surf: Speeded up robust features", *Proc. of ECCV*, p.p.404-417, 2006.
 [12] E. Rublee V. Rabaud K. Konolige G. Bradski, "ORB: an efficient alternative to SIFT or SURF", *Proc. of ICCV*, p.p.2564-2571, 2011.
 (上記参考文献中の URL は、2015 年 7 月 9 日時点での存在を確認。)

執筆者紹介 武井 宏 将 (Hiromasa Takei)

2004 年 日本ユニシス(株)入社.

入社時より CAD/CAM 分野のシステム開発業務に従事. 2013 年より画像処理・三次元形状処理の研究開発を担当.

