

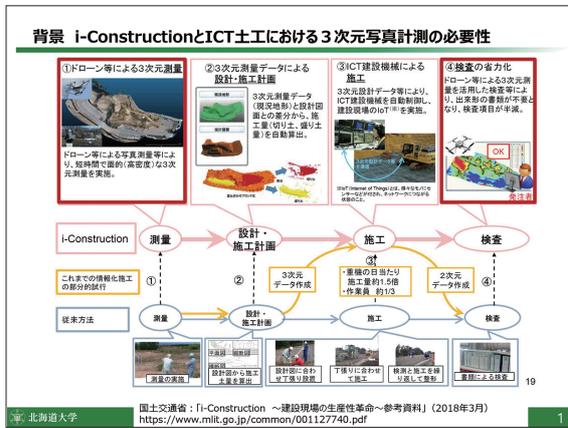
# 社会・技術動向講演会 2023 in 札幌

## SfM/MVS を用いた大規模環境の 3次元計測とデジタルツイン化

北海道大学 大学院情報科学研究院 特任教授 金井 理

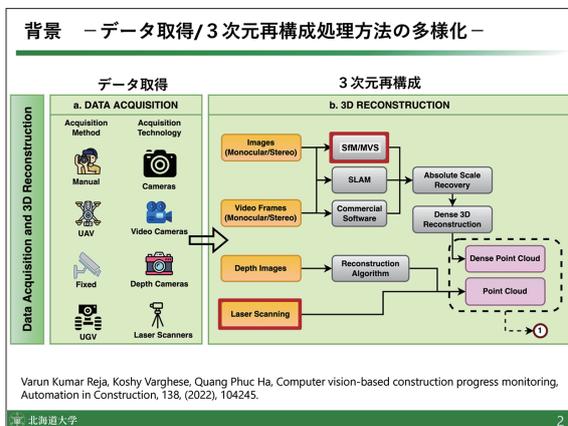
皆さま、おはようございます。今日はSfM/MVS、これは写真計測技術の1つですが、これを用いたわれわれの研究事例、研究動向をかいつまんでお話ししたいと思います。

### 1. 背景



それではまず背景から説明します。昨今、この業界の方にi-Constructionの話をするのは釈迦に説法ですが、i-ConstructionとかICT土工、ICT施工においては、3次元の測量や、3次元での竣工検査が求められ、このために非常に大規模な環境を3次元で非接触計測するということが必要になっております。

群も含めたデータ化の技術としては、昨今いろいろなデータ取得技術が進んできております。主要なところでは、画像で撮って、今日お話しするSfM/MVSでメッシュ化するとか点群化する方法、もしくはこれもあとでお話しますがSLAMという技術で3次元化する方法などがあります。一方、先ほどの講演にもあったように、従来から使われているレーザ計測で3次元データを生成する手法もあります。その他に最近ですと、深き情報を直接取れる小型のLiDAR付機器で点群化するというようにいろいろなモードで3次元計測することが可能になってきています。



この大規模環境を対象とした3次元計測や、計測点

また昨今、実環境を3次元計測するデバイスやプラットフォームもいろいろなものが出てきています。従来は地上型レーザスキャナー、もしくは高価な一眼レフを用いた写真計測で3次元化するといったことが普通だったのですが、それ以外にバックパック型とか、四足歩行ロボットの上にレーザスキャナー載せるとか、自動運転台車にレーザを載せるなど。データ取得のデバイスもスマホを使ったり、小型のアクションカメラ、例えばGoProのようなものを使ったりとか、360度カメラを使ったり、それから先ほどご紹介した深さを直接測れるLiDAR機能ももち深度を直接計測できるDepthカメラなど。こういうものが、例えばiPhone13Proとかに載っているのをご存じだと思いますけれど、私たち

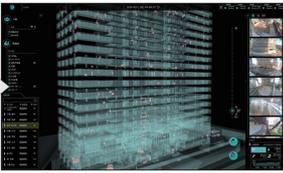
はいろいろなモードのものを使えるようになってきました。こういうデータ取得デバイスとして、特にカメラ関係の高解像度化、小型化、それから撮影範囲の拡大とか、低価格化というのが顕著だと言えるのではないかと思います。iPhoneの中に確かにLiDARのセンサーも付いているのですが、5メートル先ぐらいいましか撮れないとか、点群の取得精度があまりよろしくないということで、たぶん測量グレードでの計測というのはなかなか難しいと言われていたのですが、それに比べてカメラのほうは、かなり高解像度化しているというのが特徴かなと考えます。

**背景 - デジタルTwin構築技術の高度化 -**

実現場



仮想空間 (デジタルTwin)



- ・ 作業員や重機の実時間位置情報、フォークリフトなど重機の稼働率分析結果、作業員の移動履歴等の「過去位置再生」などのデータを遠隔地で表示可能
- ・ 稼働率の低い余剰機材の削減、作業員の滞留を回避する計画の立案などの効果が期待

■ 様々なセンサを利用した、実世界の人員・資機材の位置情報のリアルタイム取得・3D可視化・分析の可能性 (特に施工中)

2020/04/27 建設現場にIoTセンサーを導入 建設現場のIoT化で効率化 様々な分野での活用が期待 [Sponsored]  
 産 北海道大学 産 北海道大学 (senkei.com) 4

もう一つ、こうして撮った3次元の点群データですが、従来は例えば工事が始まる前に1回撮って、それから工事が終わった後に1回撮って、検査などに使うというような利用が多かったのですが、最近ではもう、建設現場でも状況をリアルタイムに3次元で撮って、その中から人員や資材等の位置とか重機の位置、稼働率などを分析して、それをコンピュータの中に再現するいわゆるデジタルツインを作って、重機がちゃんと思っただように稼働しているのかとか、さぼっているのはど

**写真計測の特長 - レーザ計測と写真計測(SfM/MVS)の比較 -**

	レーザー計測 (LiDAR)	写真計測 (SfM/MVS)
計測システムコスト	× (高価なスキャナ・ソフト)	○ (カメラ+比較的安価なソフト)
計測時間	○ (数min/スキャン)	△ (多数の画像撮影が必要)
計測データ形式	3次元点群+反射輝度 (+点のRGB情報)	3角形メッシュ + 表面テクスチャ画像
計測から計測データ出力までの処理	○ (計測後、直ちにデータ出力可能)	× (モデル生成 数10min~数h, 要スケール調整)
計測精度	○ (1mm~数mm)	△ (数cm)
インフラ維持管理で使える情報	<ul style="list-style-type: none"> <li>・ 3DCADモデル生成</li> <li>・ 構造体の経年変形</li> <li>・ スケール、剥落</li> </ul>	<ul style="list-style-type: none"> <li>・ 3DCADモデル生成</li> <li>・ 表面クラック、遊離石灰、錆、剥落などの変状</li> </ul> <p>中小建設業への導入も容易</p>

北海道大学 5

こかなど、そういうことを即時に分析することも可能になってきました。

ということで、3次元の計測技術とその用途も非常にいろいろな面で有効活用できる可能性が出てきたのですが、一方、建設業界では昔から言われていますが、中小建設業のところにこうした技術を導入する時に、レーザはやはりシステムの価格が高い、デバイスが高い、それから場合によっては作業を外注しないといけないということで、コスト的になかなか許容できないという問題がありまして、精度は確かにいいのですが、やはりちょっと中小建設業への導入というのは難しい状況です。

一方、写真計測のほうは普通のカメラさえあれば、あとは10万円ぐらいの比較的安価なソフトで3次元化ができるということで、精度はレーザにこそ及びませんが、非常にリアリティーの高い3Dのモデルが作れるので、中小建設業への導入も、どちらかという写真計測のほうが少し進んでいて、今後も進む可能性が高いのではないかなと考えています。

それでわれわれの研究室も、何年前まではレーザ計測でのデータ処理を専門に研究してきたのですが、数年前から「やはり少し写真計測のほうの研究もしよう」というようになってきています。今日はその中の代表的なものを少しご紹介させていただきます。

2. SfM/MVS技術と課題

**SfM-MVSによる3次元as-isモデル生成プロセス**

- 現状の3次元写真計測 = SfM/MVS処理による3次元化
- SfM (Structure-from-Motion) : タイポイント\*とカメラポーズ\*\*の推定 (短時間 e.g., 6min/1125枚)
- MVS (Multi-View-Stereo) : 高密度な3次元モデルの生成 (長時間 e.g., 45min/1125枚)

\* 同一点を異なる画像上で計測した3次元対応点  
 \*\* 物体に対するカメラの3次元の撮影位置・姿勢

入力

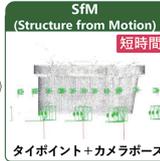
多視点からの複数画像



画像集合

SfM (Structure from Motion)

短時間



タイポイント+カメラポーズ

MVS (Multi-View Stereo)

長時間



高密度モデル

金井理：「SfM-MVS技術の動向 (小特集 フォトグラマトリ(その1))」, 写真測量とリモートセンシング, 60(3), pp.95-99, (2021)  
 北海道大学 6

さて、写真計測ですが、まずSfM/MVSって何だ？ということなのですが、これは3次元の写真計測で、皆さんも日常の業務でお使いになっていると思います。SfMというのはStructure from Motion、MVSというのはMulti-View Stereo、それを合体させた言葉です。

普通は写真計測のソフトの中にSfMとMVSの両方の機能が入っていて、写真から3次元化ができるようになっています。ご存じだと思いますが、写真計測の場合、対象を重複して撮った多数の視点からの画像をこのソフトに入ると、まず初めにSfMが、物体の表面上の疎な特徴点であるタイポイントと、それからその時のカメラの位置・姿勢、この緑色の部分（スライド中央）ですけれども、これを推定します。その後、多視点のステレオマッチングを行って画素単位での深さの推定を行って、こういう（スライド右）高密度の点群もしくはメッシュを生成します。

SfMの部分の処理は非常に短時間で、最近のコンピュータですと1,000枚ぐらい入れても数分ぐらいで答えを出してくれますが、それに比べるとMVSの方は、高密度モデルを作るのに画素単位での一つ一つの深さの推定が必要になりますので、10倍、場合によっては100倍ぐらいの時間がかかるという課題があります。

**SfM-MVSの代表的研究プロジェクト**  
**-Building a Rome in a day [Agarwal 2010] -**

- ウェブ上から収集した大量の画像から都市全体の3Dモデルを、SfM-MVSを用いて自動で再構築する並列分散システムの開発
- 特に著名観光地（ローマ、ベニス等）で、異なるカメラ/撮影位置/撮影日で撮影された画像データセットを用いた3D化
- 米国ワシントン大学GRAILラボのコミュニティ画像コレクションプロジェクトの一部として実施
- 大規模画像間での特徴点マッチング、バンドル調整を効率よく行う処理系を構築

北海道大学 | Agarwal 2010 "Reconstructing Rome", Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Brian Curless, Steven M. Seitz and Richard Szeliski, IEEE Computer, pp. 40-47, June, 2010

この技術は、実は2010年代から実用化が進んできていて、元々はワシントン大学でやっていたグループが開発した技術です。ウェブ上のインスタグラムのようなサイトから、観光客が撮った、例えばローマのコロッセウムのような有名な観光地のいろいろな写真のデータ、カメラも違えば撮っている位置も違うのですが、それを収集してきて、そこから自動で3Dのモデルを生成するために開発された技術です。ですから、今使われている写真計測の技術のほとんどは、このプロジェクトの中で使われていたということになります。これはその時の3Dのモデルです。そして、ここ（スライド右下）にある黒い点が、観光客が撮った一つ一つの写真の位置を自動で同定した結果になります。これが現在、建設や土木で使われているSfM/MVSの原理になって

いるわけです。

**Structure-from-Motion(SfM)の原理**

- 各画像上の自然特徴点検出 (○印)
- 特徴点間の対応関係発見 (同一色○印が応、RANSAC+エビポラ方程式)
- 疎な対応点 (タイポイント) の3次元座標とカメラ位置の推定
- タイポイントの再投影誤差を最小化するようタイポイント位置およびカメラポーズを微調整 (バンドル調整) (カメラ上の赤いタイポイントの再投影誤差は大きいので、カメラ3のポーズと赤のタイポイント位置を最適化で微調整)

出力：バンドル調整後のカメラース、タイポイント

北海道大学

もう少し詳しく見てみると、SfMというのは、例えばこういう同一の物体を異なる方向から撮った写真が3枚あった時に、まずこの画像の上から特徴点というのを検出して、対応している同一の特徴量を持っている点同士を探索してあげて、それで、例えば異なる画像上のこの赤い点が同じ特徴点だとすると、その特徴点ペアの位置からこれを撮った写真のカメラの位置はこういう位置ですねということを行って、2枚、3枚、複数枚の写真から同定するということを行っています。

**SfM処理内での自然特徴点間の対応関係推定方法**

1. 画像からの自然特徴点抽出
  - ・ テクスチャ、色、輝度などが周辺と異なる点
  - ・ 画像の回転・拡大縮小・射影変換に不変
  - ・ SIFT, ORB, AKAZE 特徴点等
  - ・ 各特徴点に多次元特徴ベクトルが付随
2. RANSACによる異なる画像上の尤もらしい自然特徴点間の対応関係の推定
3. 自然特徴点間の対応関係を用いたカメラ位置 (カメラポーズ) と3次元特徴点 (タイポイント) 位置を計算

北海道大学 | <https://cmssc426.github.io/sfm/#featmatch>

このスライドは特徴点間の対応関係を推定した例ですが、これ（左上の画像）は1枚目のものです。これ（右上の画像）は2枚目の画像で視点が違ってきます。この緑色の線の両端がそれぞれ同じ特徴量を持っている特徴点のペアになります。ここにはいっぱい特徴点の対応が出てきていますが、実はちゃんと正しい対応を示している特徴点というのは、この中の3分の1ぐらいしかありません。（下の画像を指して）この赤い線の対応点は、本当は対応してないところを誤判定しているものです。ですからソフトの中で大量に出てくる対応

の中から正しいものだけを見つけないという処理をSfM  
ではやっています。

### Multi-View-Stereo(MVS)の原理 -Local Matching手法-

- Camera 1上の画素pと、その周辺窓wを設定
- 画素pの深さ方向に、深さ・法線候補値と、そこにwを逆投影した対応窓(赤、緑、青)を有限個仮定
- 各深さ候補に対応した対応窓を、他のCamera 2,3上に投影
- w内のテクスチャと、投影窓内のテクスチャの類似性(consistency)を評価
- 最大の類似性を与える深さ(緑)を選択

→ 出力: 各画像の各画素における  
深さ(もしくは視差)  
・法線マップ

[Agarwal 2010] "Reconstructing Rome", Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Brian Curless, Steven M. Seitz and Richard Szeliski, IEEE Computer, pp. 46-47, June, 2010

北海道大学 10

こうやってカメラの位置が求まりますと、次はステレオマッチングを使って、例えば(中央Camera1の画像を指して)この1枚目の画像のこの画素に写っているこの点の深さについて、(左右の2つの画像を指して)他の画像から画素周辺のパターンが似たものを探し出して深さを推定します。ここなのか、もしくはここなのかというのを推定して、全ての画素の画素単位で、一つ一つ深さ推定を行うということになりますので非常に時間がかかることとなりますが、その結果、非常に高密度な点群を得ることができるようになります。

### SfM/MVSによる消波ブロック3Dモデル化結果

-豊浦工区 護岸ブロック、画像数 68枚-

北海道大学 11

こうやって作られた3Dモデルですが、これは、われわれが海岸の護岸工事のところで68枚の写真を撮って3次元化したものです。工事の進捗状況を捉えるようなモデルを作ることができます。これは実はただ撮ったのではなくて、コンピュータの指示に従ってモデルがきれいにできる位置に撮影者が行って「ここで撮りなさい」という指示を受けながら撮っていました。このどこで撮ったらいいのかという位置をコンピュータのほ

うで推定してあげるという技術も、われわれのところ  
で開発しています。

### 鉄筋部の動画撮影例

DJI Pocket 2 /動画

DJI Pocket 2 /動画

動画再生は、上記画面内をクリック

北海道大学 12

次の例は、鉄筋の動画からの3Dモデル化です。これも大規模構造物の工事の進捗状況をリアルに記録しデジタルツイン化したいということで、高速道路の高架橋の橋脚の鉄筋工部分の、足場の狭い空間の数十cmぐらい離れたところから写真を撮って、それを3次元モデル化して作ったものです。見てお分かりのように、実は一眼レフのカメラ(スライド右)で撮るよりも、携帯型のジンバル付きアクションカメラ(スライド左)で動画撮影し、その後はSfM/MVSのソフトで3次元化すると、撮影時間も少なくかつ結構きれいにモデル化できるということも分かりました。

### モデル上での鉄筋径の計測結果

	DJI Pocket 2/動画/マーカ無	SONYα7C/動画/マーカ無	SONYα7C/静止画/マーカ無
主筋 D41 基形 部径	38.9mm	38.8mm	38.7mm
基準値	39.0mm		
帯筋 D22 公称 直径	19.8mm	19.8mm	20.7mm
基準値	22.2mm		

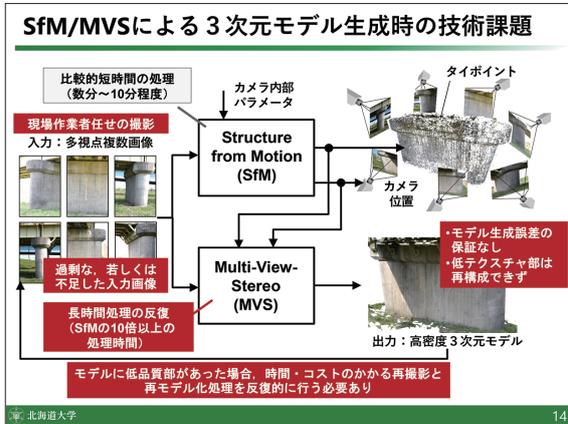
■ Φ40mm程度の主筋: いずれの計測方法でもモデルの誤差レベルは同程度鉄筋の規格寸法を特定可能な誤差で計測可能

■ Φ20mm程度の帯筋: 鉄筋径の精密測定には、注意が必要

北海道大学 13

こうやって撮ったデータの上で、例えば主筋に使われていた鉄筋の径や、帯筋の径は結構それなりの精度で計測でき、特に主筋についてはたぶん鉄筋の規格寸法の特定が可能なぐらいの精度で計測できると分かりました。工事現場の場合、狭い空間がありますよね。なかなかあの大きなカメラを持ってパシャパシャ写真を撮るのも、鉄筋工の方の作業を邪魔するの

で嫌がられたりしますけれども、こういう小さいカメラでサッと撮っていくぐらいだったら、鉄筋工の方が昼休みの間に足場をぐるっと1周するだけで済みますので楽に行うことができます。



このようにいろいろ工夫はしているのですが、やはりまだ写真計測による3次元モデルの生成には、いくつか技術課題があることも分かってきています。

先ほどのように、そもそも対象物に対して現場の撮影者任せで撮る位置を決めているので、場合によっては写真を撮り過ぎるとか、もしくはここ撮ってほしかったのに画像が不足しているというのが後になって分かるといった問題があります。また、先ほどご紹介したように、MVSでステレオマッチングするところの処理に非常に時間がかかるので、現場ですぐモデルの生成結果の品質を確認できないというような問題もあります。

それから、作られたモデルのどの部分がどれぐらいの精度でできているのかという目安が、ソフトによっては出力されないという問題。さらに、照明が暗い部分や、規則的な模様が反復されているような面、こういうのを低テクスチャ部と言うのですが、そういうとこ

ろではステレオマッチングが不安定になって深さの推定誤差が大きくなってしまおうというような問題もあります。こういったいろいろな技術課題が存在していて、これはまだ残念ながら完全に解決されているわけではありません。

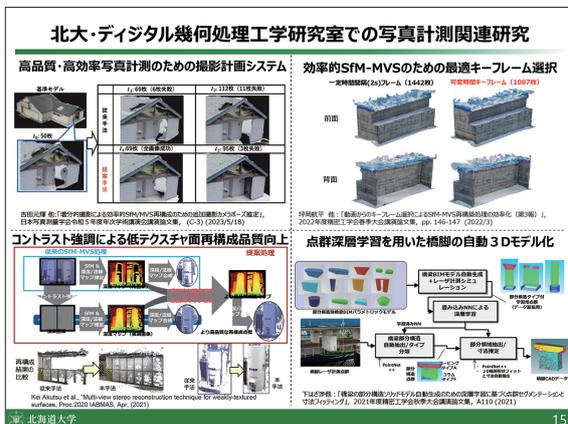
今日は時間の都合上この話はしませんが、われわれの研究室では、こういった問題を解決するために、例えばコンピュータで撮影をガイドするシステムとか、動画から可変時間で静止画を抜いてくる技術。あるいはコントラスト強調を行って3次元化できなかった部分も深さを出せるようにする技術。さらに、こういう定型形状であれば、CADのモデルを点群から完全自動で立ち上げるといったような研究を行っています。

### 3. 大規模環境の3次元計測とデジタルツイン化事例

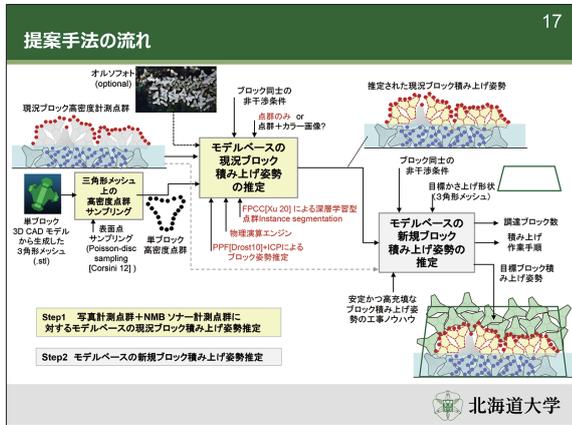
#### ① 消波ブロック認識と補充作業計画への応用

その中で今日は、写真計測の一つの応用として、消波ブロック認識というものと、それから北海道ですの除雪への応用、この2つを簡単にご紹介したいと思います。

まずは消波ブロック認識ですが、これは消波ブロックの一個一個をUAVで撮って写真計測した3Dの点群とソナーの点群から、その位置と姿勢を全部同定して再現するという技術です。消波ブロックのことをお分かりの方、ご経験のある方はご存知かと思いますが、消波ブロックは波の力で沈下しますので定期的にかさ上げ工事というのをしなければなりません。その時にいったい何個のブロックを補充したらいいのかというのを、今は結構アバウトに見積もりされているようです。アバウトと言うと怒られるかもしれませんが、実際の手順としては、まずドローンで現況の点群を撮って、海



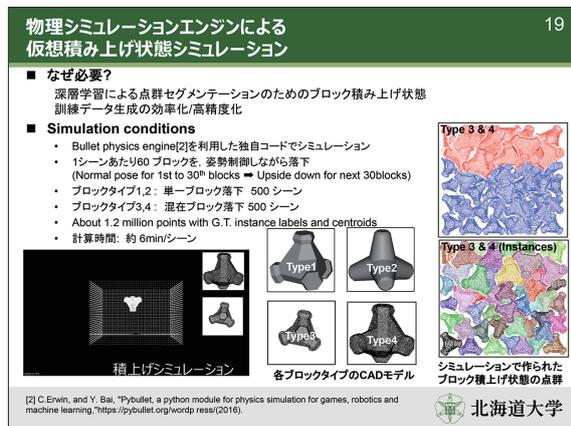
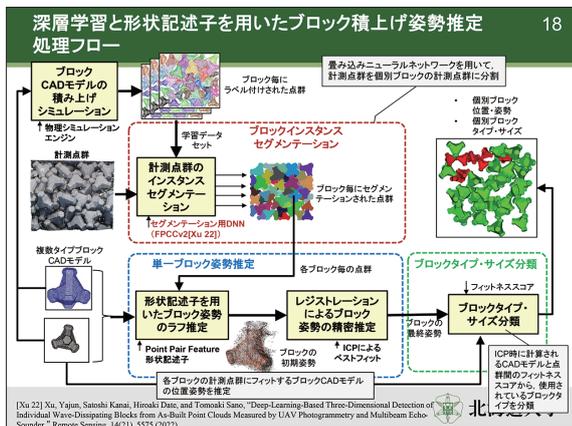
面下のところはソナーで撮って、それをCADの中に入れます。一方、かさ上げ工事には、補充してここまで高くしなければいけないとかかさ上げ目標形状があり、それと比較して、この間の不足部分の体積を計って、充填(じゅうてん)に必要な補充ブロックの個数を出すということをやっています。これで必要ブロックの個数はある程度出てくるのですが、結構個数の見積りに誤差があって、足りないというのは許されないのでブロックいっぱい作り過ぎてしまったとか、多めに作っても不足してしまうこともままあるようで、例えば、施工長さ70mぐらいのところ、この方法で見積ると710個程度必要だったのが、実際には150個程度のブロックが不足したというようなケースもあるようです。



こういう課題を解決するために、われわれのところでは、海面から下はソナーで測って、上はUAV写真計測で測った高密度の点群から、まず現況のブロックはどういうふうに積み上げられているのかというのを一個一個CADデータと比較して、そのシーンを再現し、そのシーンに対して今度はブロック積み上げの目標となる形状を与えて、この緑色の補充ブロックを上からばらばら

ばらと落としていくシミュレーションやって、必要な個数を推定するというをやっています。

まず積み上げ姿勢の推定ですが、これは結構難しいです。例えば計測点群、上から見るとこのような形になっていて(スライド左端、中央上部)、人間が見ると大体分かるのですが、この点群とここに積み上げられているブロックの3DのCADデータ(スライド左端、中央下部)を使って、この点群にびったりフィットするような、ブロックの位置と姿勢(スライド右端、中央上部)をロボット分野で使われているPoint Pair Featureと呼ばれる手法を使って推定して出しています。ここの点群(スライド左端、中央上部)からここのブロックの部分の領域の点群を一つ一つ切り出す(スライド中央上部)、これをインスタンスセグメンテーションと言いますが、ここに実はディープラーニングの技術を使っています。このように切り出してしまえば、あとはこの部分点群(スライド中央下部)に対して、このCADデータ(スライド左端、中央下部)をフィッティングするのは、ロボットで使われているような技術をうまく活用すれば可能となります。ブロックのタイプとかサイズの違いもちゃんと分類できるようになっています。時間があまりないので細かい話は割愛させていただきますが、興味のある方は、資料に論文の情報が載っていますので、そちらをご覧くださいになっていただければと思います。(論文情報: Xu, Yajun, Satoshi Kanai, Hiroaki Date, and Tomoaki Sano, Deep-Learning-Based Three-Dimensional Detection of Individual Wave-Dissipating Blocks from As-Built Point Clouds Measured by UAV Photogrammetry and Multibeam Echo-Sounder, Remote Sensing, 14(21), 5575 (2022))



点群をディープラーニングで分割する場合、学習データを作らなくてはなりません。でも学習データを他の港に積んであるデータから、一生懸命人間が一個一個ちまちまラベル付けて作るのは大変なので、ここではゲームエンジンを使って、ゲームエンジンの中で上からブロックのモデルをこのように落下させて(スライド左下)、海岸に積んでいるのと同じようなブロックの積み上げ状況を作ります。このようにすると、個々のブロックは全部それぞれ違うブロックだとゲームエンジン内では分かっていますから、上から見た時にどういふような積み上げのパターンになっているのかの正解が分かり、個々のラベルが全部自動的に付けられます。このように、シミュレーションで類似のシーンを作って学習し、実シーンの物体認識にそれを利用するという点の特徴です。

20

### テストサイトと実計測点群一覧

Site	砂原港(Sawara)		楯法華港(Todohokke)		江良港(Era)	
	UAV	MBES	UAV	MBES	UAV	MBES
点群計測方法	UAV	MBES	UAV	MBES	UAV	MBES
テスト領域サイズ[m]	~10 × 10	~10 × 10	~15 × 15	~15 × 15	~25 × 25	~25 × 25
総テスト領域数	14	15	20	10	6	7
計測点密度(点/m <sup>2</sup> )	~600	~260	~370	~250	~580	~800
使用ブロックタイプ	A1	A1	B1	B1+B2	A2+A3	A2+A3
1テスト領域内の計測点数概算(×10 <sup>5</sup> )	100~120	30~35	70~100	65~80	400~700	700~1,000

北海道大学

北海道にある砂原(さわら)、楯法華(とどほっけ)、江良(えら) という3つの港で、この消波ブロックの認識をやってみました。使われているブロックのタイプやサイズが結構異なっていて、いろいろなものが混在している港もあります。

21

### ブロックインスタンスセグメンテーション精度

	Sawara port (砂原港)		Todohokke port (楯法華港)		Era (江良港)	
	Precision	Recall	Precision	Recall	Precision	Recall
UAV	80.1	55.3	82.7	74.2	88.6	49.5
MBES	90.4	69.7	72.7	75.8	86.3	68.2

[%], IoU 閾値: 0.5

北海道大学

これは計測点群をブロックごとの部分点群に分けるインスタンスセグメンテーションの精度ですが、まずこういう点群(スライド左)を与えて、一個一個のブロックとして点群をちゃんと分離できているかどうか、その精度を検証したものです。認識した点群の中に間違いがどの程度含まれているのかという評価の一つの指標を「プレジジョン」といいます。100%が一番いい状態を表しているのですが、この指標で見ると80%程度ですから、結構いいということが分かるかと思えます。

22

### ブロック位置・姿勢推定精度 (実計測点群を使用した場合)

Sawara port (砂原港)				Todohokke port (楯法華港)				Era (江良港)			
UAV		MBES		UAV		MBES		UAV		MBES	
$E_f$ [mm]	$R$ [°]	$E_f$ [mm]	$R$ [°]	$E_f$ [mm]	$R$ [°]	$E_f$ [mm]	$R$ [°]	$E_f$ [mm]	$R$ [°]	$E_f$ [mm]	$R$ [°]
33	71.32	39	67.28	34	66.17	38	60.54	48	75.12	58	71.52

●  $h_{th} = 100$  mm for Sawara and Todohokke,  $h_{th} = 200$  mm for Era

北海道大学

さらに積み上げられたブロックの位置・姿勢の精度については、実計測データとCADで配置されたブロックのモデルとの間の平均距離を算出してみると、大体数cm程度でした。ブロックの大きさというのは大体数mぐらいですので、数mぐらいのブロックに対し数cmぐらいのオーダーでフィッティングできているとことが分かりました。海面下のソナーのデータ(各港の表の右側、MBES)は、結構(海が)荒れていたもので少し悪くなりますけれども、それでもそれなりにフィッティングできていることが分かります。

23

### ブロックタイプ分類精度 (実計測点群)

	Todohokke port (楯法華港)		Era (江良港)	
	B1	B2	A2	A3
使用ブロックタイプ	B1	B2	A2	A3
実ブロック個数	77	58	111	70
誤分類ブロック数	4	5	3	3
分類精度	94%		96%	

北海道大学

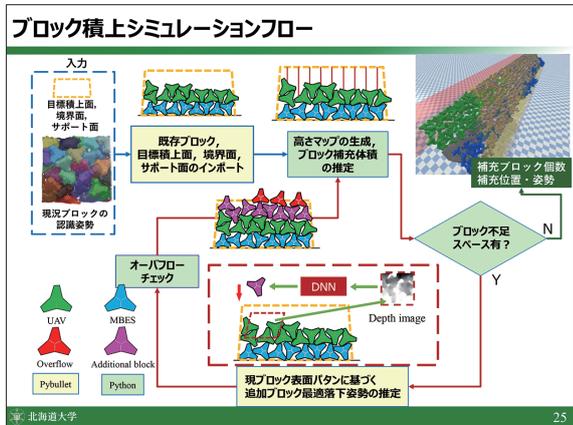
それからブロックのタイプの分類も、結構しっかりできていて、サイズの違いもちゃんと分類できているということが分かります。

計算処理時間

サイト	Sawara (砂原港)		Todohokke (税法華港)		Era (江良港)	
計測方法	UAV	MBES	UAV	MBES	UAV	MBES
テスト領域サイズ [m]	~10 × 10	~10 × 10	~15 × 15	~15 × 15	~25 × 25	~25 × 25
テスト領域数	14	15	20	10	6	7
1領域あたりブロック数	40~50	35~50	25~35	25~30	30~50	30~40
計測点密度 [点/m <sup>2</sup> ]	~600	~260	~370	~250	~580	~800
使用ブロックタイプ	A1	A1	B1	B1+B2	A2+A3	A2+A3
1領域あたり点数 (×10 <sup>3</sup> 点)	100~120	30~35	70~100	65~80	400~700	700~1,000
1領域あたりブロックインスタンスセグメンテーション時間 [s]	0.8	0.4	0.7	0.5	3.6	4.2
1領域あたりブロック位置・姿勢推定時間 [s]	127	72	130	85	175	162
1領域あたりブロックタイプ分類時間 [s]	32	24	35	26	160	158
1領域あたり総計算時間 [s]	160	96	166	112	339	324

北海道大学

計算の処理時間も非常に速くて、20m×20mぐらいの領域に分けて、ブロックの認識とCADモデルのフィッティングをおこなっているのですが、1領域あたり数分から5~6分ぐらいでできます。ですから1時間も待っていれば数百mぐらいのところは全部認識できるということになります。



こうして、現況がどういふブロックから作られているのかというのは分かるのですが、次はこの既存ブロックの上に目標形状になるまで、上からブロックのモデルをばらばら落としていくというシミュレーションを行います。この時、ブロックが積まれている表面にできるだけ密に新しいブロックが入るような姿勢の計算にも、実はディープラーニングを使っている、なるべくタイトにブロックが入るような姿勢を推定して補充個数を推定計算しています。

例えばこういう緑色の既存のブロックのシーン(スライド右上)があった時に、ディープラーニングを使わ

DNNによるブロック落下初期姿勢の推定結果

- ブロックタイプ: Clinger 6Ton
- 実装タイプ: Pytorch
- 学習セット: 姿勢付き深さ画像800枚 (512×512画素)
- テストセット: 深さ画像200枚 (512×512画素)
- Optimizer: ADAM

DNNで推定した落下初期姿勢推定誤差

使用したDNN	回転誤差 (°)	変位誤差 (m)
Transforme [r [Parmar, 2018]	42.45	0.42
ResNet16 [He, 2016]	36.2	0.36
ResNet134 [He, 2016]	28.48	0.37
ResNet150 [He, 2016]	25.15	0.31

北海道大学

ないで上からただランダムにブロックを入れていくと、このように、他のブロックに引っかかったような状態の姿勢になってしまいますけれども(スライド下段2つの図のうち左側)、ディープラーニングを使った方法でやると突起がちゃんと下に向けて、この空間になるべく密に入るように、姿勢を推定することができます(スライド下段2つの図のうち右側)。

要補充ブロック個数の比較結果 - DNNによる落下姿勢最適化の有無による違い -

- いずれの方法でも、実工事の追加ブロック数(864個)と同程度の追加数を推定可能。
- 同一目標空間に対し、DNNによる落下姿勢最適化を用いて推定された補充ブロック個数は、ランダムボース落下で推定したブロック個数に対し、100個程度多いブロックを挿入可能
- ➡ DNNによる落下姿勢最適化が、より強固な消波ブロック構造の実現に寄与する可能性

要補充ブロック個数推定値 (DNN未使用)				要補充ブロック個数推定値 (DNN使用)					
# of blocks	f = 0m	f = 0.3m	f = 0.35m	f = 0.5m	# of blocks	f = 0m	f = 0.3m	f = 0.35m	f = 0.5m
v = 0.4m <sup>3</sup>	645	788	852	903	v = 0.4m <sup>3</sup>	733	889	955	1016
v = 1.0m <sup>3</sup>	645	792	844	900	v = 1.0m <sup>3</sup>	698	891	925	1017
v = 2.0m <sup>3</sup>	645	786	823	892	v = 2.0m <sup>3</sup>	676	894	906	1012
v = 4.0m <sup>3</sup>	645	790	821	884	v = 4.0m <sup>3</sup>	671	842	868	898
v = 6.0m <sup>3</sup>	645	756	795	836	v = 6.0m <sup>3</sup>	668	761	768	780

目標面: v = 0.4m<sup>3</sup>, f = 0.35m, Number = 852  
 目標面: v = 0.4m<sup>3</sup>, f = 0.35m, Number = 955

北海道大学

ディープラーニングを使っても使わなくても、シミュレーションにより実際の工事で追加したブロックの個数と大体同等の追加数が推定できるということが分かっています。ただ、ディープラーニング使ったほうが、場合によっては外側の面が同じでも100個ぐらい多くブロックが入られるような姿勢が求まるということなので、消波ブロックをできるだけ密に積むのがいいのかどうかは分かりませんが、少なくともこうやって最適化をしてあげると、より強固な消波ブロックの構造の実現には寄与できると考えています。

また将来、これと例えばクレーンのオペレーションの自動化とを結び付けられれば、施工自動化などにも寄与できるのではないかと考えています。現在、実補充

28

### まとめ・今後の課題

**■ まとめ**

1. 消波ブロックの補充作業計画のため、ディープラーニングと物理エンジンを組み合わせ、積上げ時の重心点高さが最小で、より密にかつ物理的に安定となる最適な追加ブロックの挿入位置・姿勢を推定する手法を提案
2. 北海道内の港湾を対象としたケーススタディにより、提案手法によるブロック追加の計画は、同一目標補充空間に対し、より多くのブロックを密に挿入可能で、より強固な消波ブロック構造の実現に寄与できる可能性があること、その推定個数は実作業で追加されたブロック個数と同程度であることを確認

**■ 今後の課題**  
実補充作業計画への適用と検証



北海道大学

作業計画への適用ということで、別のサイトでの検証を行っている最中です。

② 動画からの路肩堆雪部の体積推定

### 背景

**■ 北海道地区の除排雪作業の現状**

- 作業員の高齢化、人員不足
- 深夜労働や残業の多さ
- 自治体の除雪予算削減

→ 作業の合理化・効率化  
・排雪対象の堆雪部体積の高精度見積りが  
運搬トラック台数の適正発注に必要

**■ 現状の堆雪部体積測定作業とその課題**

- 一定断面上での代表点測量 + 平均断面法による体積計算
- 非効率/危険な作業
- 低い体積見積精度 (過剰なダンプ発注台数)
- 雪捨て場残量の過少評価

→ 排雪作業前に、路肩の堆雪部体積を一定精度で効率良く、かつ低コストで測定可能な技術が必要



排雪作業の例



路肩堆雪部代表点測量作業の例

29

さて、時間がなくなってきたので、この路肩堆雪部の体積推定については少し飛ばしながら説明します。これは路肩の雪山。北海道の方だとお分かりだと思いますけれども、これをロータリー除雪車で排雪をするという作業が冬には必要になります。この排雪作業に何台のダンプを発注したらいいのかを検討するときに、予め現場で測量して雪山の体積を求めるということをやっているのですが、見ていただくと分かる通り、非常に危険な作業で効率も悪く、しかもある断面しか測れないので見積もりの精度も悪いです。従って、排雪作業前に路肩の堆雪部の体積をそれなりの精度で効率よく、かつ安全に低コストで測定できるような技術が求められています。

そこで路肩堆雪部の体積測定に求められる要件を、道内の建設会社で実際に除雪の作業をされている方へのヒアリング結果に基づき1.適度な推定処理速度、2.低コストな機器・ソフト、3.短時間の計測・体積推定、

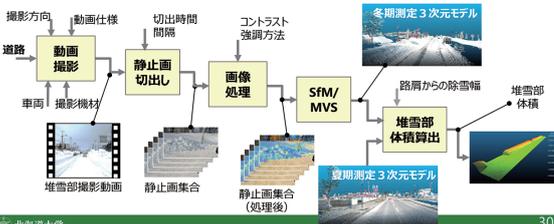
### 本試行の目的・手法

**■ 目的**

- 車両から撮影した動画から、SfM/MVS\*を活用し路肩堆雪部の体積を一定精度(ダンプ1台分以内程度の誤差範囲)で測定可能な手法の開発
- 堆雪部3次元モデルを安定生成可能な条件の実験的検討
  - 撮影機材/方向、偏光フィルタの有無、画像処理の必要性、静止面切出し時間間隔
- 夏期/冬期測定モデルの併用による高精度な堆雪部体積推定自動化手法の検討

\*Structure-from-Motion / Multi-View Stereo

**■ 手法**



30

4.作業員の少なさ、5.バックエンド作業の少なさといった要件を満たす写真計測を応用した路肩堆雪部の体積計測と推定方法を開発しました。車の上にGoProというアクションカメラを2台付けて、夏と冬の2回、道路を走ります。40キロぐらいで走って撮ってきた動画から先ほどご紹介したSfM/MVSを使って雪の表面と、夏の雪が降っていない表面を撮って3次元点群化し、その差分を取って堆雪部の体積を自動計算しています。

### 実験3-画像コントラスト強調の効果-

**■ コントラスト強調の必要性**

- 雪面から抽出可能な特徴点を増加させ、SfM処理の成功確率の向上が期待
- 晴天下の新雪部の一部がMVSで3次元モデル化できない現象改善が期待

**■ 使用コントラスト強調方法**

- ヒストグラム均一化(HE)
- コントラスト制限適応型ヒストグラム均一化(CLAHE)



コントラスト強調によるSURF特徴点抽出の改善効果

コントラスト強調による鳥面撮影画像の違い

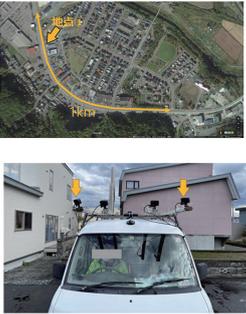
31

ご存じだと思いますが、雪の表面が完全に真っ白だと写真計測できません。ですがある画像処理を施すと、計測できないと思っていた雪の表面が3次元化できるようになるという場合もあるということが分かってきました。

留萌市内の1kmぐらいの区間、軽車両の上にGoProを2台付けて走って、GPSの受信機がこの中へ入っています。撮影した動画から静止面に切り出して、その静止画一枚一枚に、RTKではありませんがGoProの普通のGPSの座標を付けてSfM/MVSの処理をしました。画像は大体2,000枚ぐらい。それでもモデルの生

### 実験概要 -計測地点・計測方法-

- 計測区間**
  - 留萌市南町3丁目～湖静3丁目付近
  - 国道233号線 深川方向車線側路肩部
  - 区間長：約1km
- 計測方法**
  - カメラ：GoPro-11、2台(左右)、CPLフィルタ付
  - 動画撮影条件：4K、120FPS、約2min間
  - 軽乗用車両輪フリップアップ上に固定
  - 走行速度：40km/h、信号停止あり
  - モデルスケール計測
    - カメラ内蔵GPSを用いて動画各フレームへGPS位置付加
- 計測日時**
  - 夏期：2022年11月
  - 冬期：2023年1月
- 3次元モデル作成方法**
  - 静止画抽出し時間間隔：0.3s
  - 総静止画枚数：約2000枚
  - SfM/MVS処理：Metashape Pro, v1.8
  - モデル生成総時間：約1h強 (PCスペックに依存)



32 北海道大学

成時間は意外と速くて、1時間強でできます。40km/hぐらいで走りながら動画を撮って、そこから3次元化しています。

### 夏期計測点群 (地点①拡大)



計測点数：8,078K

33 北海道大学

これは夏の間の点群のある部分の拡大図ですが、路肩部分は結構ちゃんと撮れて3次元化できています。ただ、MMSによるレーザ計測ではないので、上のほうを見ていただくと、看板などあまりうまく点が取れていないのが分かりますが、少なくとも問題としている雪山計測に必要な路肩の部分はしっかりと取れています。

### 冬期計測点群 (地点①拡大)



34 北海道大学

これは雪が積もった後の計測結果です。雪山の表面の点群がそれなりにしっかりと出ていますね。

### 夏期+冬期計測点群位置合わせ結果 (地点①)

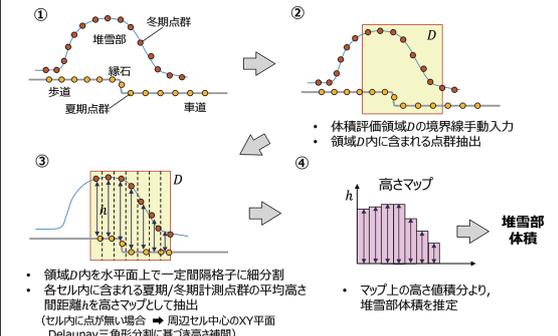


- 点群位置合わせ：道路標識、バス停、看板上の対応点6点を手動入力し、自動位置合わせ
- 目視範囲での位置ずれは見られず、両点群の高さ比較により堆雪部体積の推定は可能

35 北海道大学

この2時期に計測した点群に共通して含まれる道路標識、バス停、看板上の対応点を数点手動入力して自動位置合わせをするとこんな形になります。結構ちゃんと合います。

### 冬期/夏期点群からの堆雪体積の推定処理フロー

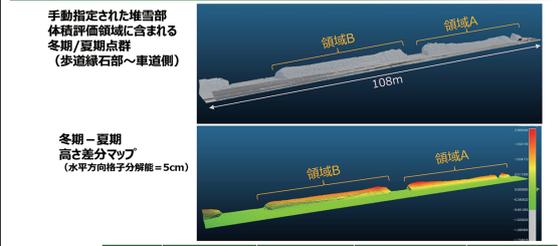


- 冬期点群、夏期点群、歩道、車道、緑石、堆雪部
- 領域Dの境界線を手動入力し、領域D内に含まれる点群を抽出
- 領域D内を水平面上で一定間隔格子に細分割し、各セル内に含まれる夏期/冬期計測点群の平均高さ間距離hを高さマップとして抽出 (セル内に点がない場合 → 周辺セル中心のXY平面 Delaunay三角形分割に基づき高さ補間)
- 高さマップを作成し、堆雪部体積を推定

36 北海道大学

そこから、排雪の対象となる領域 (スライド右上の矩形) を、手で境界線を入力して、このように格子に切って、高さマップを作り、冬の点群の高さから夏

### 冬期/夏期点群高さ差分と堆雪体積の推定 (地点①)



各領域の堆雪体積推定値	点群より平均断面法*で求めた推定体積 $V_1$ [m <sup>3</sup> ] (従来手法)	高さ差分マップから求めた推定体積 $V_2$ [m <sup>3</sup> ] (提案手法)	堆雪部体積推定誤差 ( $V_1 - V_2$ ) [m <sup>3</sup> ] (提案手法)	グッドトラック台数に換算した推定誤差 [台]
領域A	124.67	107.11	+17.56 (16%)	+1.25
領域B	68.3	63.13	+5.17 (8.2%)	+0.37

提案手法による体積推定は、従来手法の過剰体積見込みを防止可能

\*平均断面法：体積と各領域の左右代表断面の平均断面積×長さ方向長さ

37 北海道大学

の点群の高さを引いて、堆雪部の体積を推定しています。

このように撮ってきて推定した雪山の体積は、従来の平均断面法に比べると大体10%~15%ぐらいの違いがあるということが分かりました。仔細な検証は行っていませんが、おそらく提案手法のほうが正確だと思います。平均断面法のほうがたぶん誤差が大きいということになります。これをダンプの台数に換算すると、従来方法だと100mで大体0.3~1.2台ぐらいダンプが余計にいるというような推定になっています。他の区間だともっと誤差が大きいところもありました。

**まとめ**

- 車両撮影した動画から、写真計測により路肩堆雪部を安定に3次元モデル化可能な条件を実験的に調査
  - 撮影機材/方法: アクションカメラ、動画、CPLフィルター付き撮影
  - 事前画像処理: コントラスト強調が雪面再構成領域の拡大に有効
  - 静止画抽出間隔: 再構成領域を最大化する最適切り出し時間間隔が存在
- 夏期/冬期測定モデル (3次元点群) の比較による堆雪部体積推定は技術的に実現可能
  - 車両搭載のアクションカメラ2台による動画撮影
  - 計測終了後、堆雪部3次元点群は2時間程度で推定可能
  - 但し、体積推定に必要な点群位置合わせ、体積評価領域指定の自動化が今後必要
- 作業人員、測定・計算時間 → 従来手法の約1/5に削減可能

北海道大学 38

これを工数の削減効果で見ると、従来手法で測量し手で書類を書くような方法に比べて大体5分の1ぐらいの工数で済むということも分かっています。これについても処理の自動化を今年度のプロジェクトでやる予定になっています。

4. 関連技術動向

**最新研究: NeRF: Neural Radiance Fields**

- 多数の既存撮影画像から、シーンの3次元形状 (実際にはRadiance Field) を復元し、新たな視点からの画像を高速・高品質に生成可能な深層学習手法[Mildenhall 20]
- ただし、学習に多数のカメラポーズ付きの画像が必要 (通常は写真計測 (SfM) による推定)
- UAV撮影画像からの3Dモデル生成向けNuRF (Mega-NeRF) も近年提案[Turki 22]

北海道大学 39

さて最後に、写真計測技術の周辺の新しい技術についてお話させてください。SfM-MVSとちょっと違う

技術としてNeRF (ナーフ) というものが最近出てきています。これは、まず対象を多方向から撮影してSfMだけを行い、対象物に対する撮影画像の位置姿勢を推定します。その後MVSをやらないで、いわゆる機械学習を行うと、画像を撮影した方向のボリュームレンダリング結果がほぼ撮影画像に近いような対象物回りの3次元的な色分布と体積透過率分布の3次元ボリュームが出てきます。そしてこの3次元ボリュームを使うと、撮影されていない方向からの画像が、ほぼリアルタイムに合成できます。機械学習も非常に高速で1分ぐらいでできます。

**NeRFによる3Dボリュームモデルの推定例 (橋脚画像110枚)**

橋脚周辺全110枚の画像を入力して生成したNeRFボリューム

橋脚の右方向の撮影画像が11枚程度不足した状態で生成したNeRFボリューム

ポーズ付き画像→NeRFボリューム生成時間: 約1min

北海道大学 40

例えば、橋脚だとこのようなものが出てきます。これが1分ぐらいでできます。ただし、先ほど説明したようにNeRFの場合も、その入力となる撮影画像の位置姿勢を推定するためにSfMだけは行わなければなりません。さて面白いことにNeRFでの再構成では、撮影が非常に不足している空間に (スライド下の図) 雲みたいなぼやっとした「フロータ」と呼ばれるアーチファクト (画像のノイズ) をもつボリュームが出てきます。撮影が十分であればスライド上のようにフロータは現れません。従ってNeRFで作られた、こういうフロータのようなアーチファクトの多寡をうまく評価してあげると、本講演の最初に少し触れた計算機のガイドに従った自動撮影計画の従来手法に比べて、もう少し高精度かつ高速に、対象物の3次元化の際に撮影が不足している位置や方向の推定ができるのではないかと考えています。コンピュータービジョンの世界では、このNeRFの研究が今非常に盛んに行われていて、盛り上がっている状況です。

今日は写真計測の事例をご紹介いたしました。それからわれわれのところではやっている例をご紹介しまし

## まとめ・今後の展開

## ■ まとめ

- 写真計測 (SfM/MVS) : BIM/CIM, i-Construction導入のための低コストで簡便な現況の3次元計測手法として有効
- 小型カメラ+動画撮影+SfM/MVSの組合せによる低コストな3次元計測・3次元再構築の可能性
- SfM/MVSの計測原理と技術課題の理解
- 3次元計測データ+CADデータや、異時期計測データ組合せによる高度な3次元解析

## ■ 今後の展開

- (SLAM, NeRF)+SfM/MVSの統合による、高速かつ高品質な現況モデル化技術の実現
- モデル化対象の要求精度/利用目的に応じた、適切な3次元計測手法、計測プラットフォームを選択できる技術者の育成

©北海道大学

41

たが、やはりデータを取るだけではなくて、それとCADデータを合体させてあげるとか、2番目にご紹介したように、夏と冬のデータを組み合わせて差分を取って体積を計算するとか、そうするとそれなりの精度で写真計測ができ、特に動画撮影というのは結構いい

のではないかとということが分かってまいりました。今後の展開として、NeRFのような技術と組み合わせることで、さらにもう少し高品質な写真計測が可能になるのではないかと考えています。

以上で発表を終わらせていただきます。どうもありがとうございました。(拍手)



## ■ 講演者

金井 理 (かない さとし)

北海道大学

大学院情報科学研究所 特任教授

本稿は2023年9月22日に開催された、当協会主催「社会・技術動向講演会 2023 in 札幌」における、金井 理 氏の特別講演の内容をまとめたものです。