

確率場と深層学習に関する第1回CRESTシンポジウム

シーンモデリングにおける 畳み込みネットの応用と課題

櫻田 健 (東京工業大学)

@sakuDken

2016/1/13 早稲田大学

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

略歴

東北大学

学部 (機械知能・航空工学科)

2005年4月



2009年3月

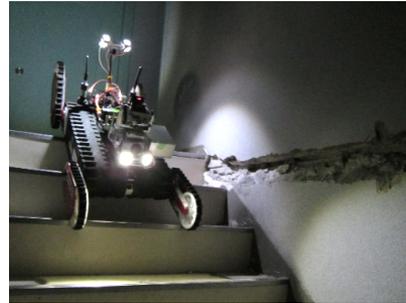
修士 (情報科学研究科)

2009年4月



2011年3月

(田所研)



ロボティクス
(レスキューロボット)



コンピュータービジョン

東日本大震災

博士 (情報科学研究科)

2011年4月



2015年3月

(岡谷研)

日本学術振興会
特別研究員DC2

2012年4月



2014年3月

客員研究員
(カーネギーメロン大学)

2013年4月



2014年3月

東京工業大学

博士研究員

2015年4月～

(奥富研)

CMUと日産の自動運転
プロジェクトに参加 (半年)

学部・修士の研究

レスキューロボット

- 研究：レーザーSLAM, 遠隔操縦支援 [IROS2010, SII2010]
- 開発：3次元レーザースキャナ, 操縦インターフェース

実績

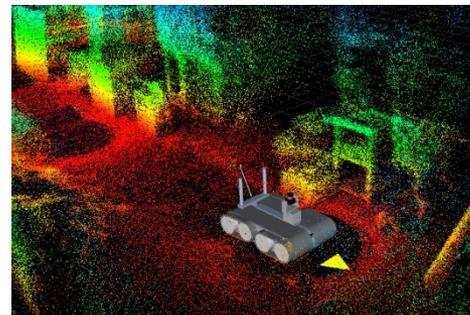
- NEDO 閉鎖空間内高速走行探査群
- ロボカップ2009世界大会 (チーム名: Pelican United)
 - レスキューロボットリーグ 世界優勝 2部門, 世界準優勝 2部門



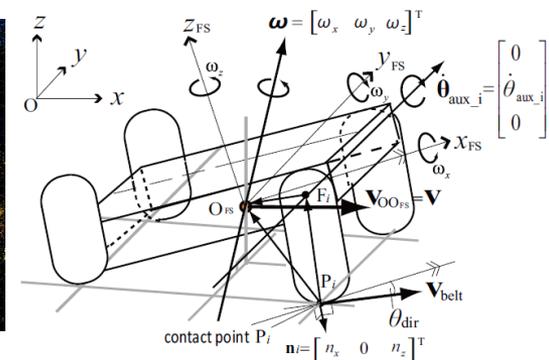
サブクローラを有する
クローラロボット



3次元レーザースキャナ



3次元地図と
操縦インターフェース

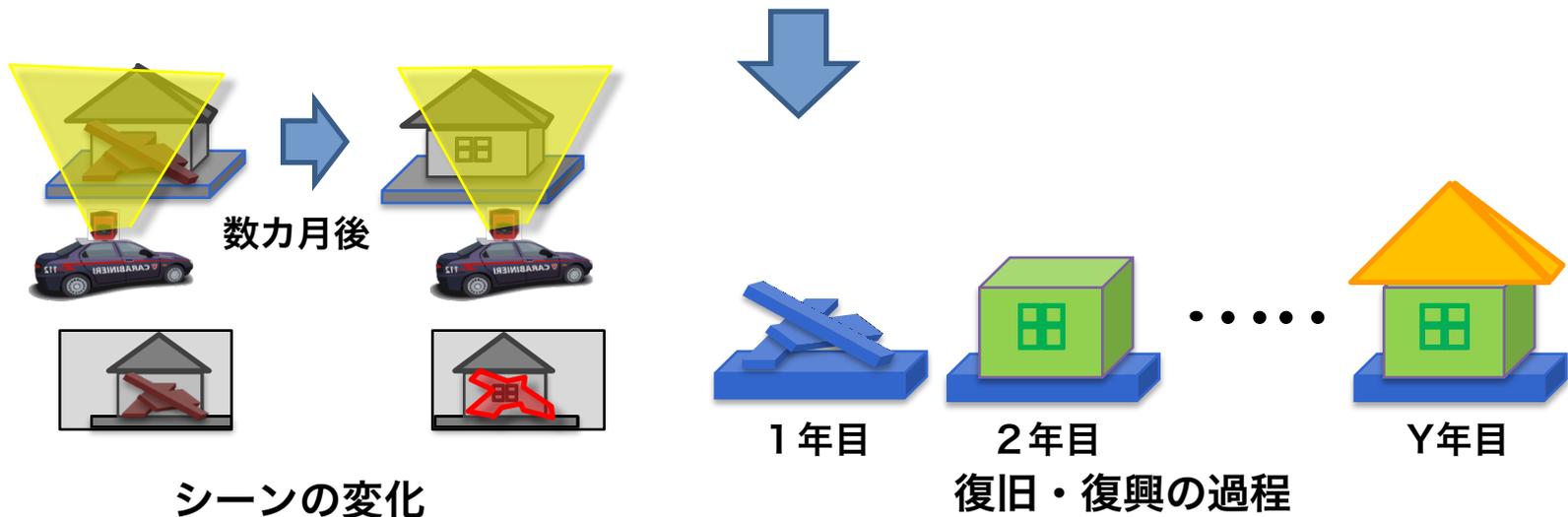


位置推定モデル

博士の研究

車載画像を用いた市街地の時空間モデリング

- 市街地がどう変わってゆくかをモデル化



目次

- 自己紹介
- **研究背景**
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

3.11被災地の映像記録 (1/2)

- システム

- 車の屋根に取り付けた全方位カメラ
 - Ladybug 3/5, Point Grey Research inc.
- GPSセンサーあり・デプスセンサーなし
- 路上約2m間隔で画像を撮影



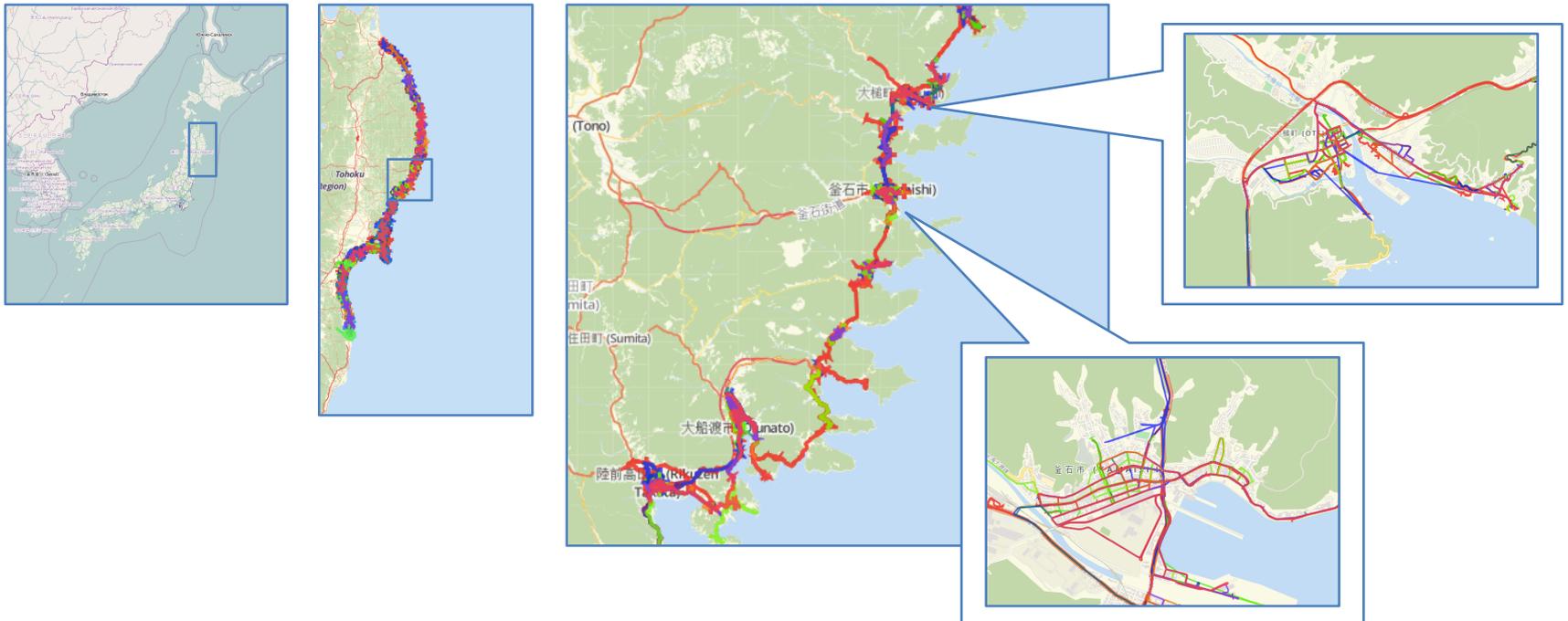
画像例

- 岩手県大槌町，地震後約1ヶ月



3.11被災地の映像記録 (2/2)

- プロジェクトの概要
 - 沿岸部被災地のほぼ全体 ~ 長さ500km
 - 地震発生後約1ヶ月から**定期的に撮影**; 初年度は3ヶ月に一度, 以降6ヶ月
 - 現在のデータ: **40テラバイト** = 数千万枚の画像



プロジェクトの目的

- 被害の記録
 - 教育, 土木・建築, 都市計画, 防災等
- 復興の記録
 - 3-6ヶ月ごとの定期的撮影

2011/4



2011/7



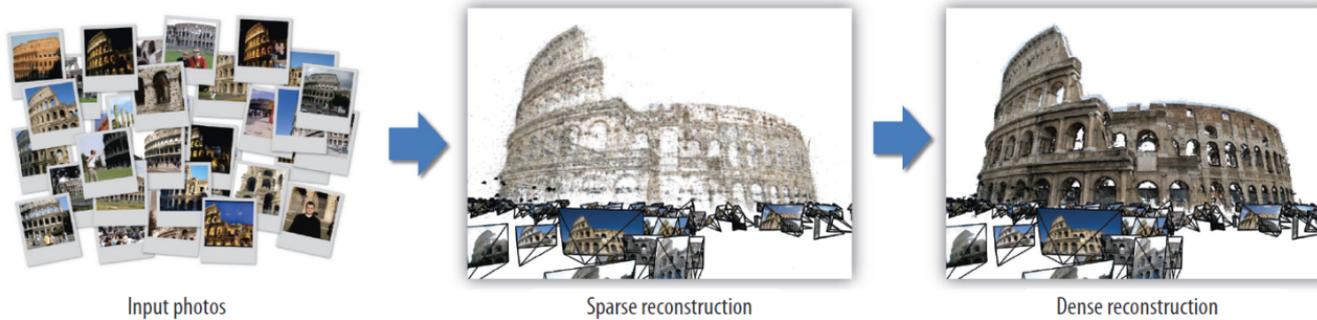
2012/1



3次元復元 (SfM+MVS) の研究状況

- 研究から応用の段階へ

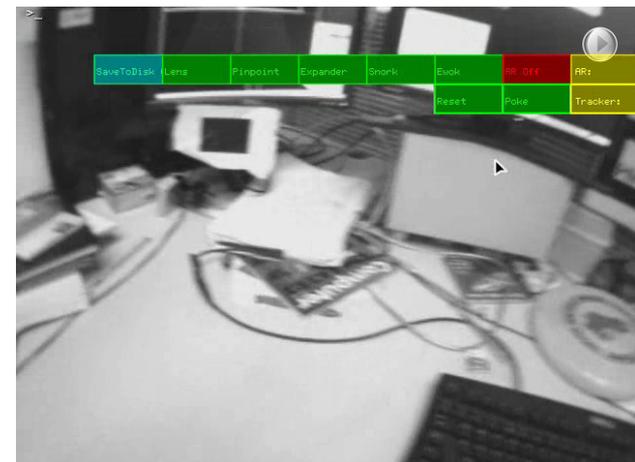
未整列画像からの大規模復元 [Snavely+04, Agarwal+10]



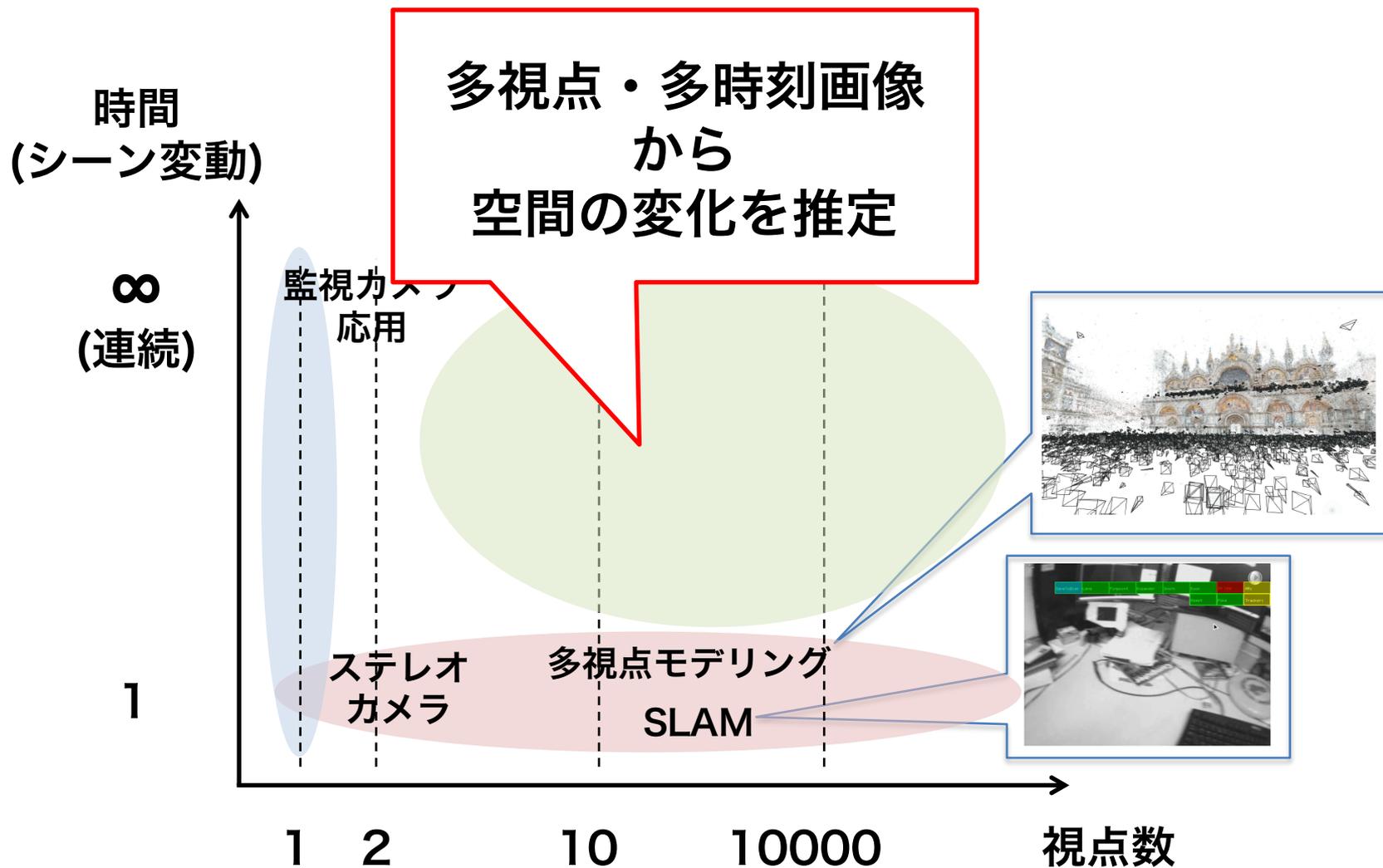
Autodesk 123D Catch



PTAM [Klein+07]



次の課題：3Dから4Dへ



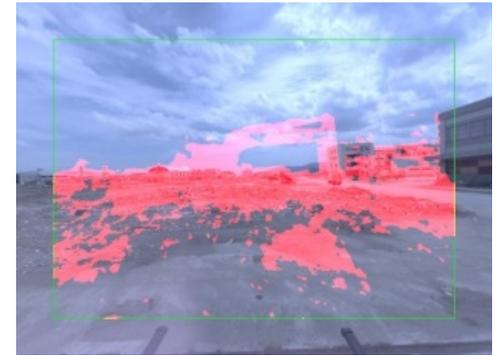
戦略

- 様々な技術を用いた変化検出手法を提案

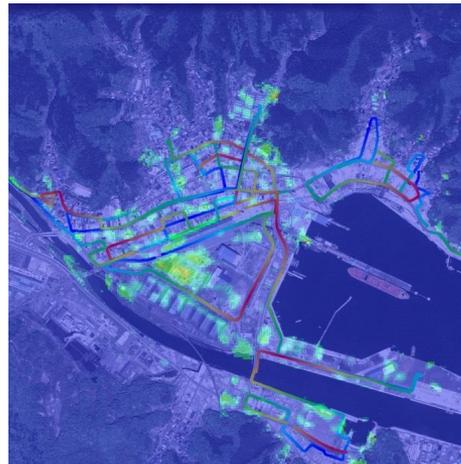
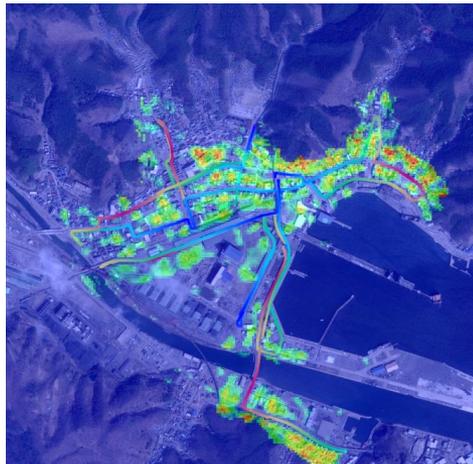
2次元ベース
高速なシーン変化検出



3次元ベース
正確な3次元構造の変化検出

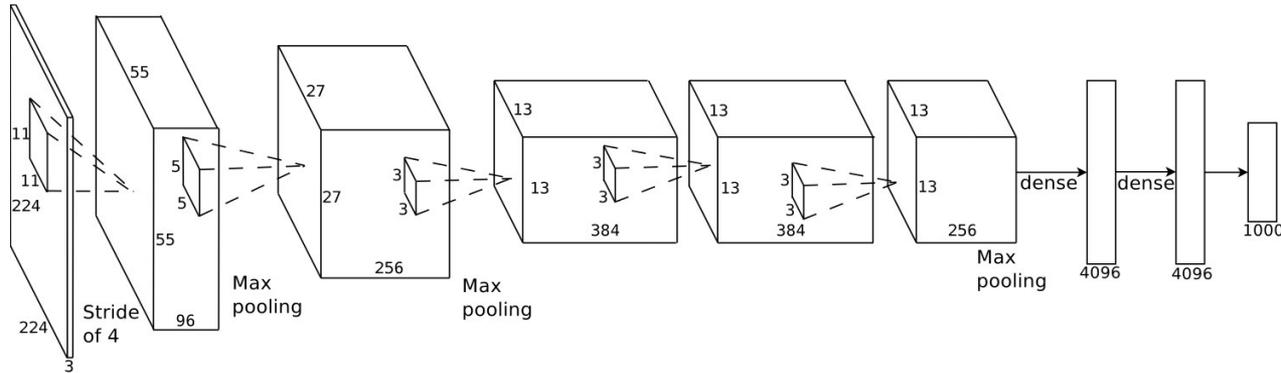


物体認識ベース
市街地全体の瓦礫や植物の分布推定

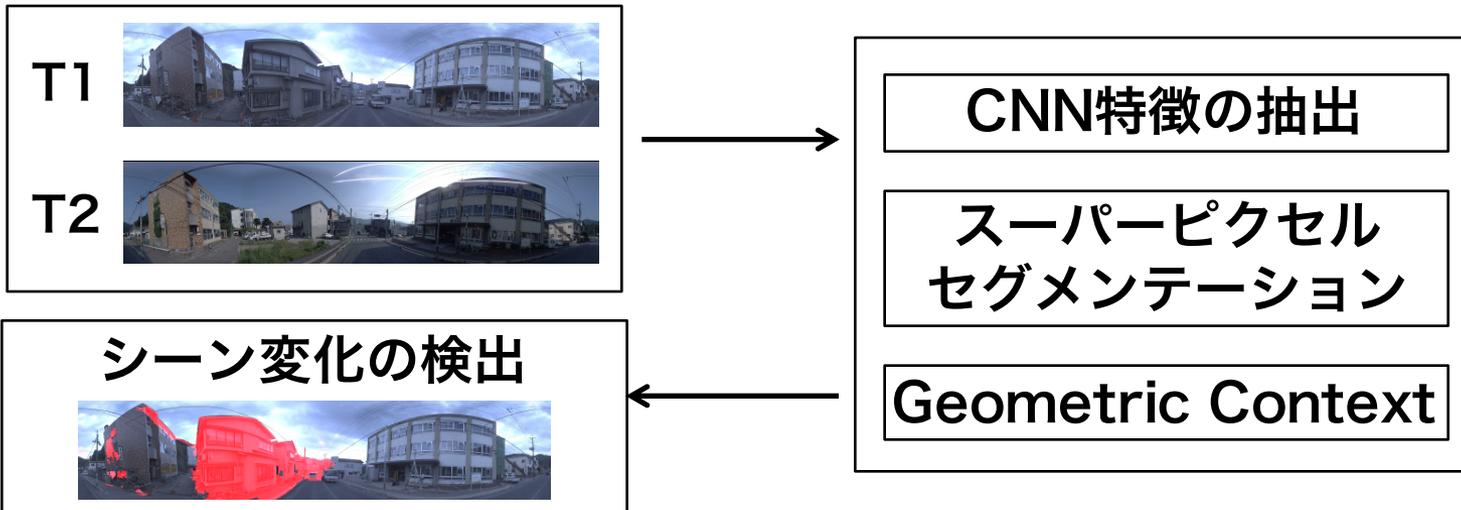


CNN特徴を用いた高速な変化検出

[BMVC2015]



“ImageNet Classification with Deep Convolutional Neural Networks”, Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton, NIPS 2012



CNN特徴を用いた高速な変化検出

[BMVC2015]

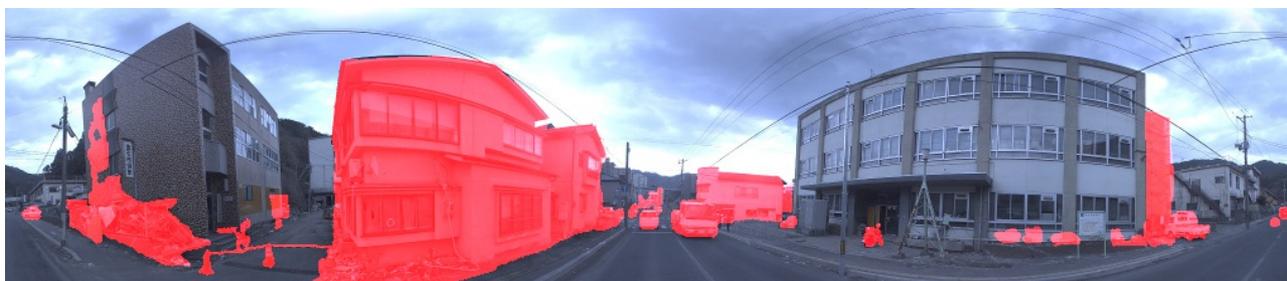
時刻 1



時刻 2
検出結果



人による
判定結果

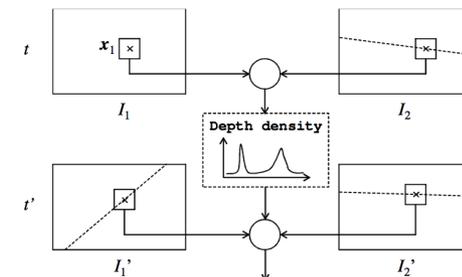
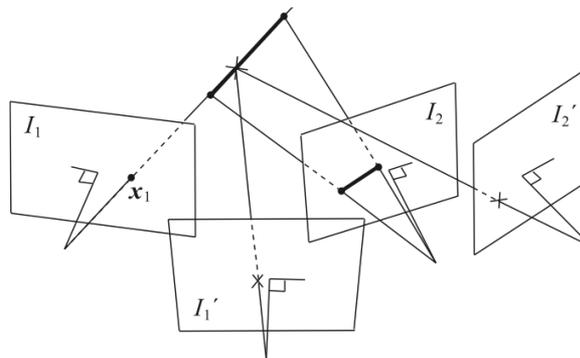
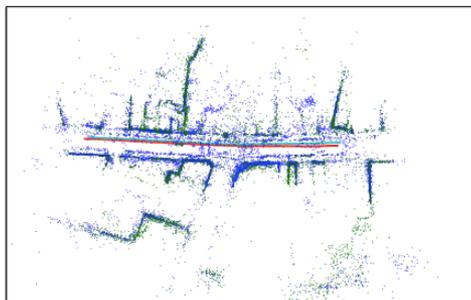
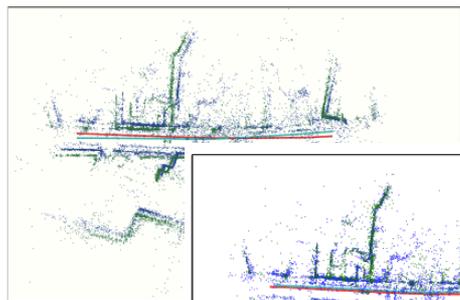
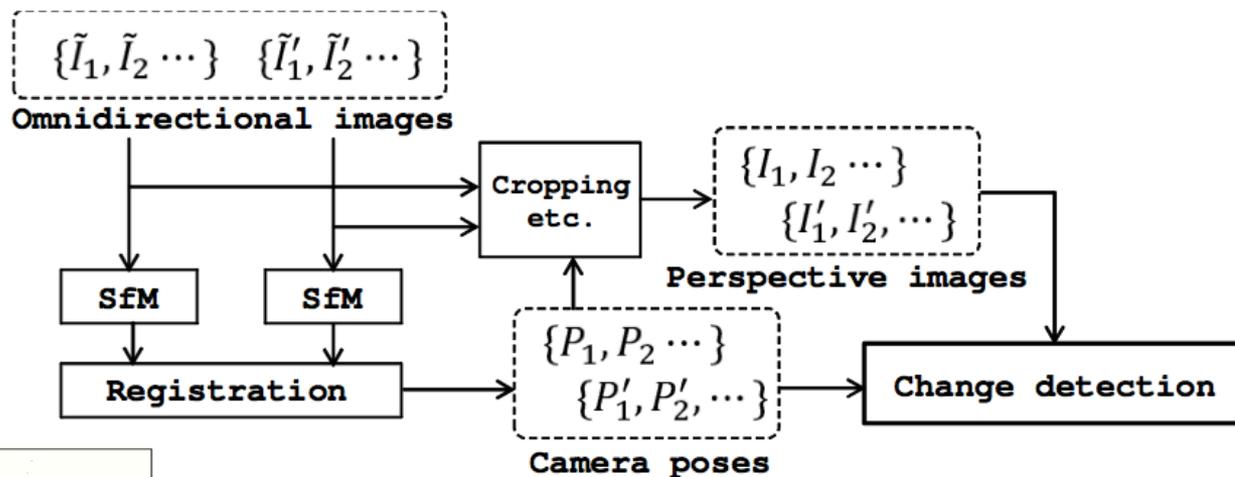


画像の
直接比較



確率的デプス推定に基づくシーン変化検出

[CVPR2013]



確率的デプス推定に基づくシーン変化検出

[CVPR2013]

2011/4



2011/7



<https://www.youtube.com/watch?v=u3kXJQuy7lo>

市街地の状態推定

[ACCV2014]

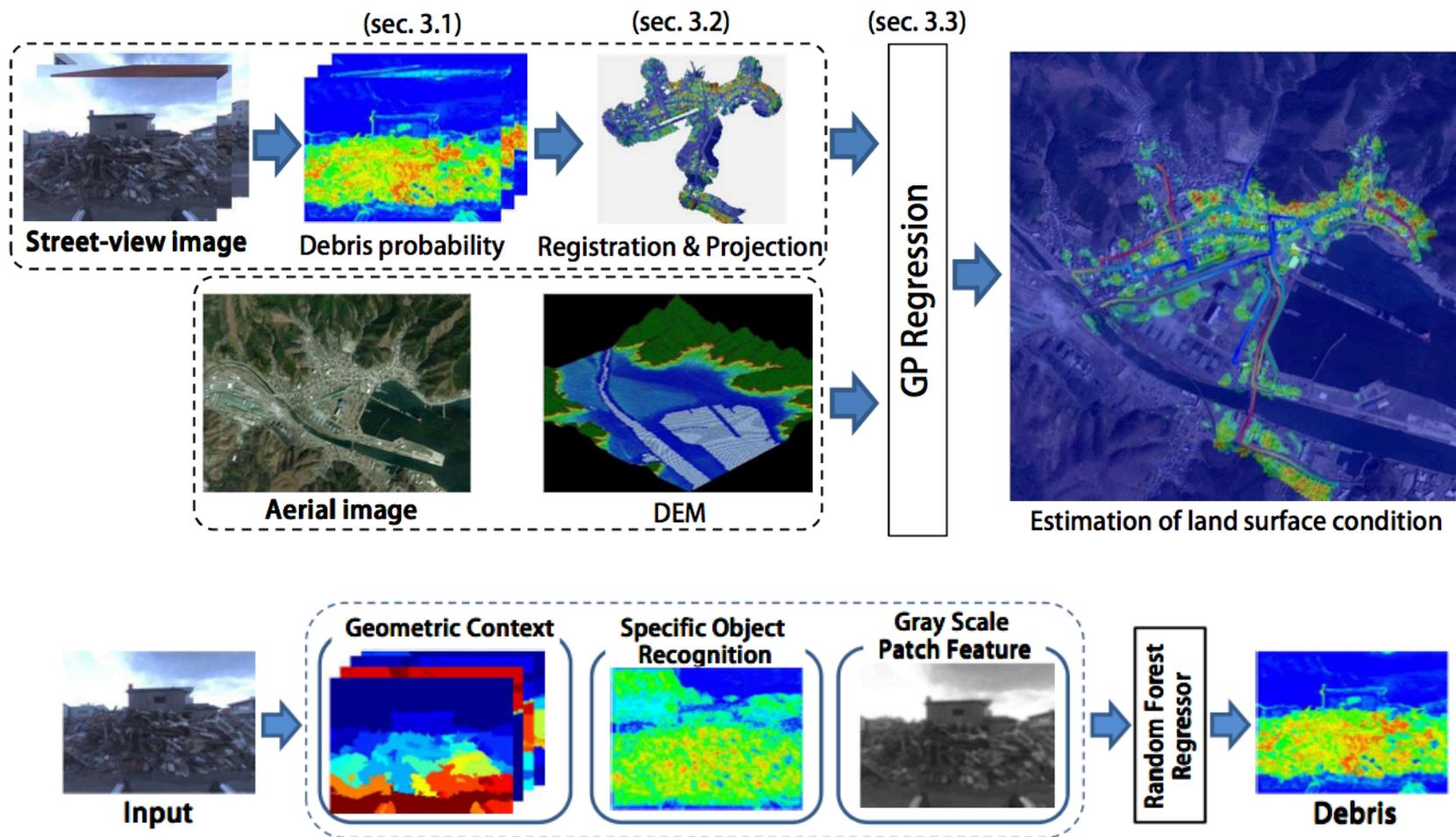
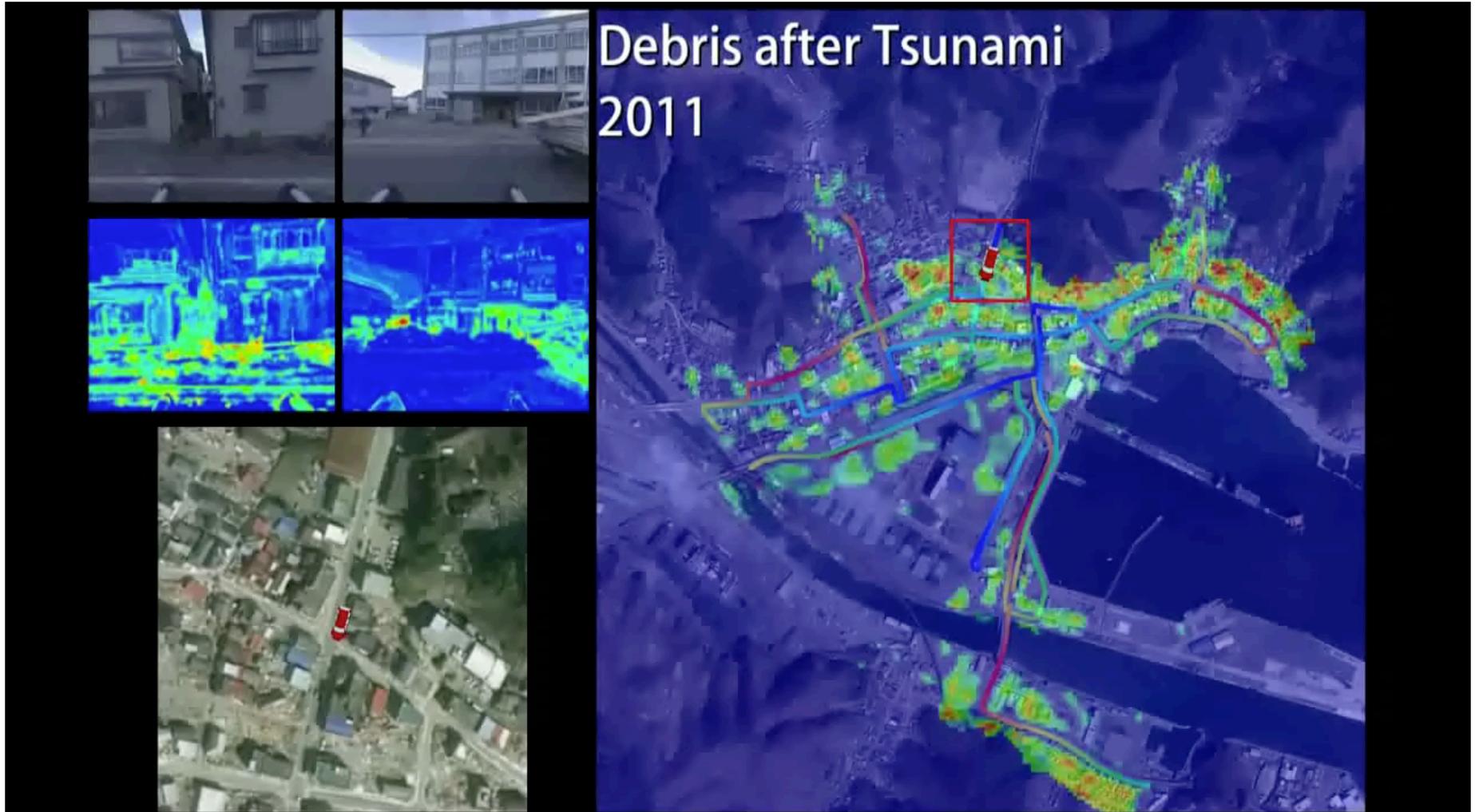


Fig. 4. Data flow diagram of debris detection. As features of debris, the probabilities of geometric context, specific object recognition and patch features are employed.

市街地の状態推定

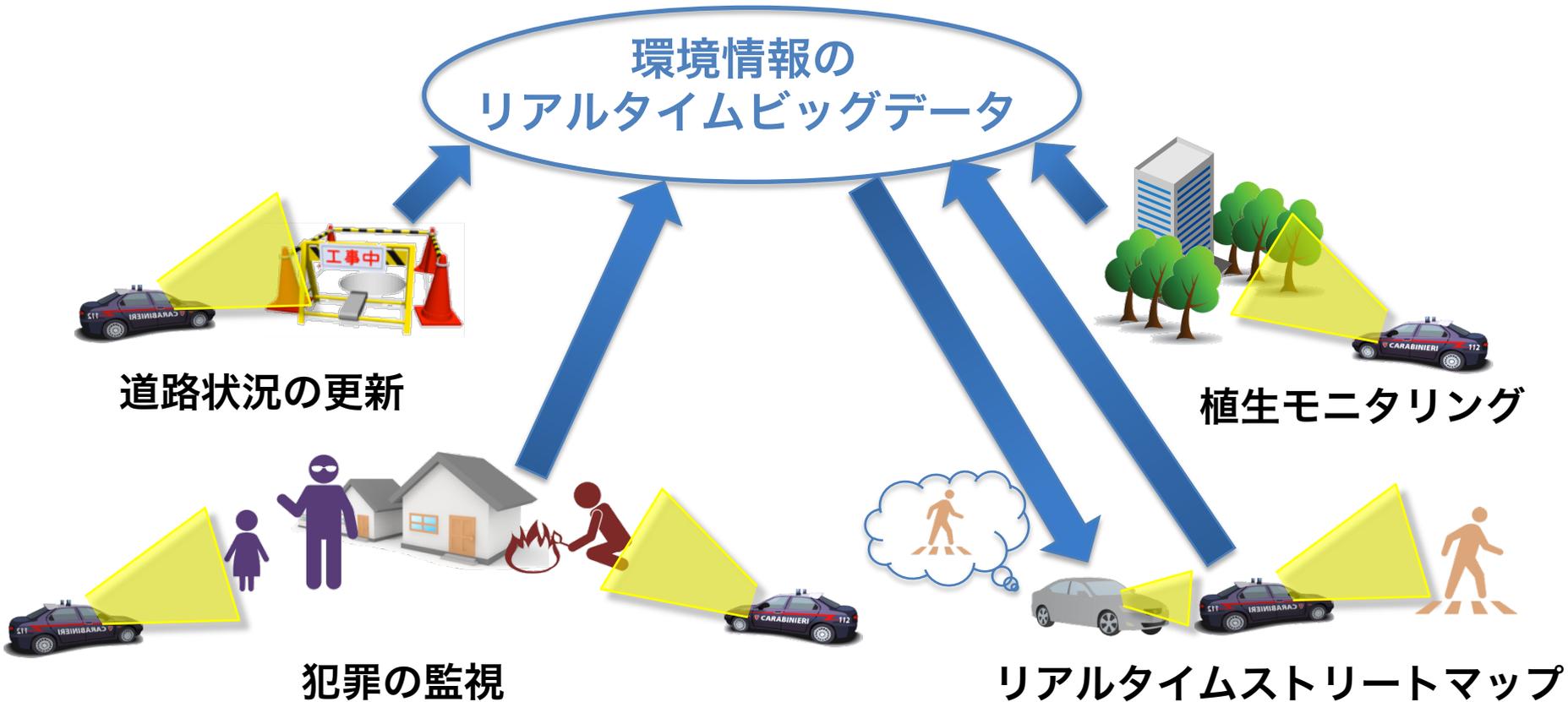
[ACCV2014]

- 地上＋上空視点画像の融合



将来像

- 移動体センサーネットワークを利用したダイナミックな環境のモデリング
 - 地上での出来事をリアルタイムで学習



目次

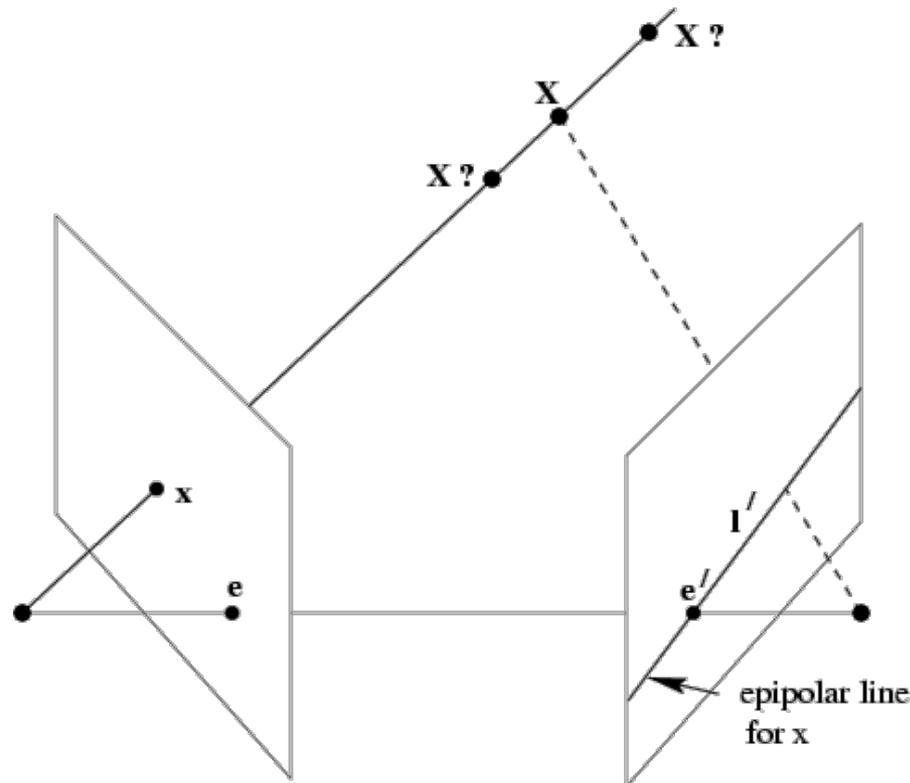
- 自己紹介
- **研究背景**
 - **画像を用いた市街地の時空間モデリング**
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- **解析的手法の現状と課題**
- **CNNのシーンモデリングへの応用**
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

3次元モデリングの基礎

- 推定した対応点から三角測量の原理により3次元点（デプス）を推定

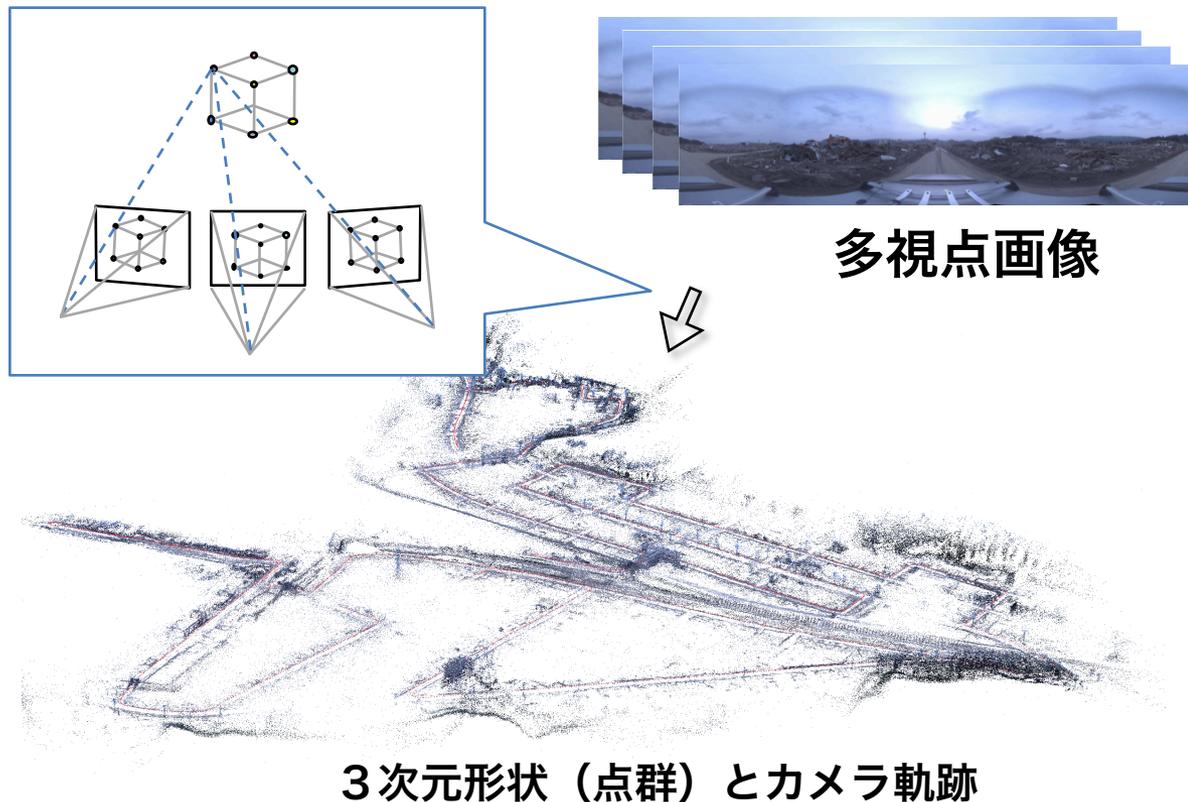


エピポーラ幾何

(参照 : Multiple View Geometry)

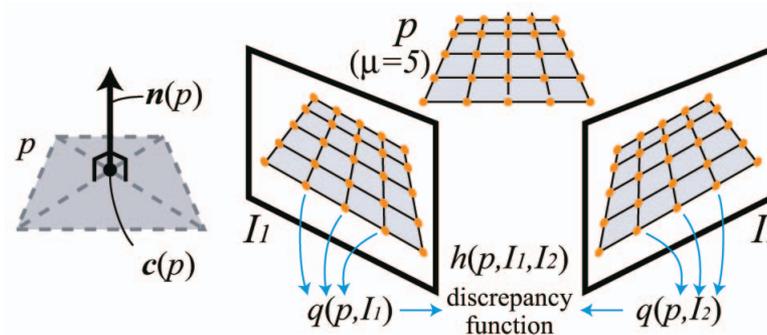
Structure-from-Motion

- 多視点画像からカメラ姿勢と3次元形状を同時に推定
 - 特徴点の対応付け
 - 対応点から3次元点を復元
 - 再投影誤差の最小化



多視点ステレオ (MVS)

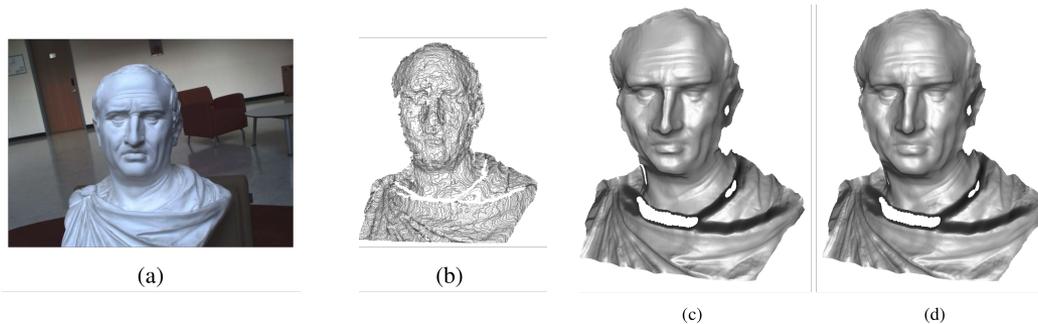
- 姿勢が既知の多視点画像からシーンを密に復元
 - “Accurate, Dense, and Robust Multi-View Stereopsis”, Yasutaka Furukawa and Jean Ponce, TPAMI, 2010



3次元形状の高精度化

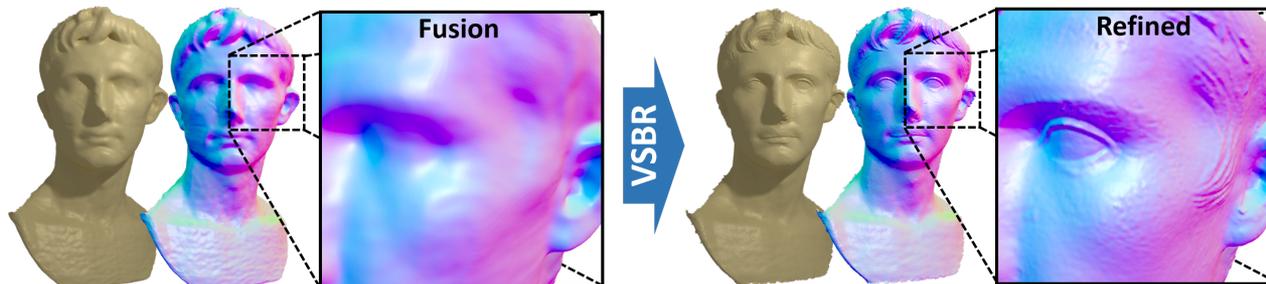
- 単眼RGB-D画像 + フォトメトリ

- “Intrinsic scene properties from a single rgb-d image”, J. T. Barron and J. Malik, CVPR, 2013
- “High quality shape from a single rgb-d image under uncalibrated natural illumination”, Y. Han, J.-Y. Lee, and I. S. Kweon, ICCV, 2013



- 多視点RGB-D画像 + フォトメトリ

- “Shading-based refinement on volumetric signed distance functions”, M. Zollhofer, A. Dai, M. Innman, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, ACM Trans. Graphics, 2015

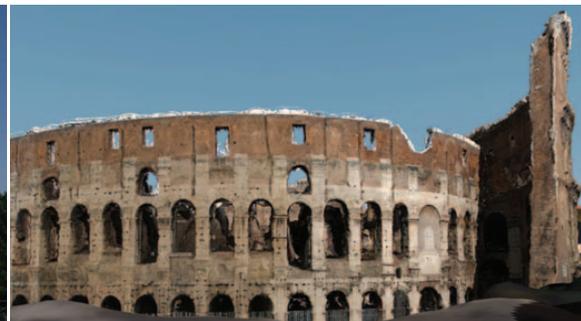


3次元形状の高精度化

- 多視点ステレオ (MVS) + フォトメトリ
– 法線ベクトルをリファイン



Reference photo



Rendered view

“The visual turing test for scene reconstruction”, Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. M. Seitz, 3DV, 2013

3次元再構成の例



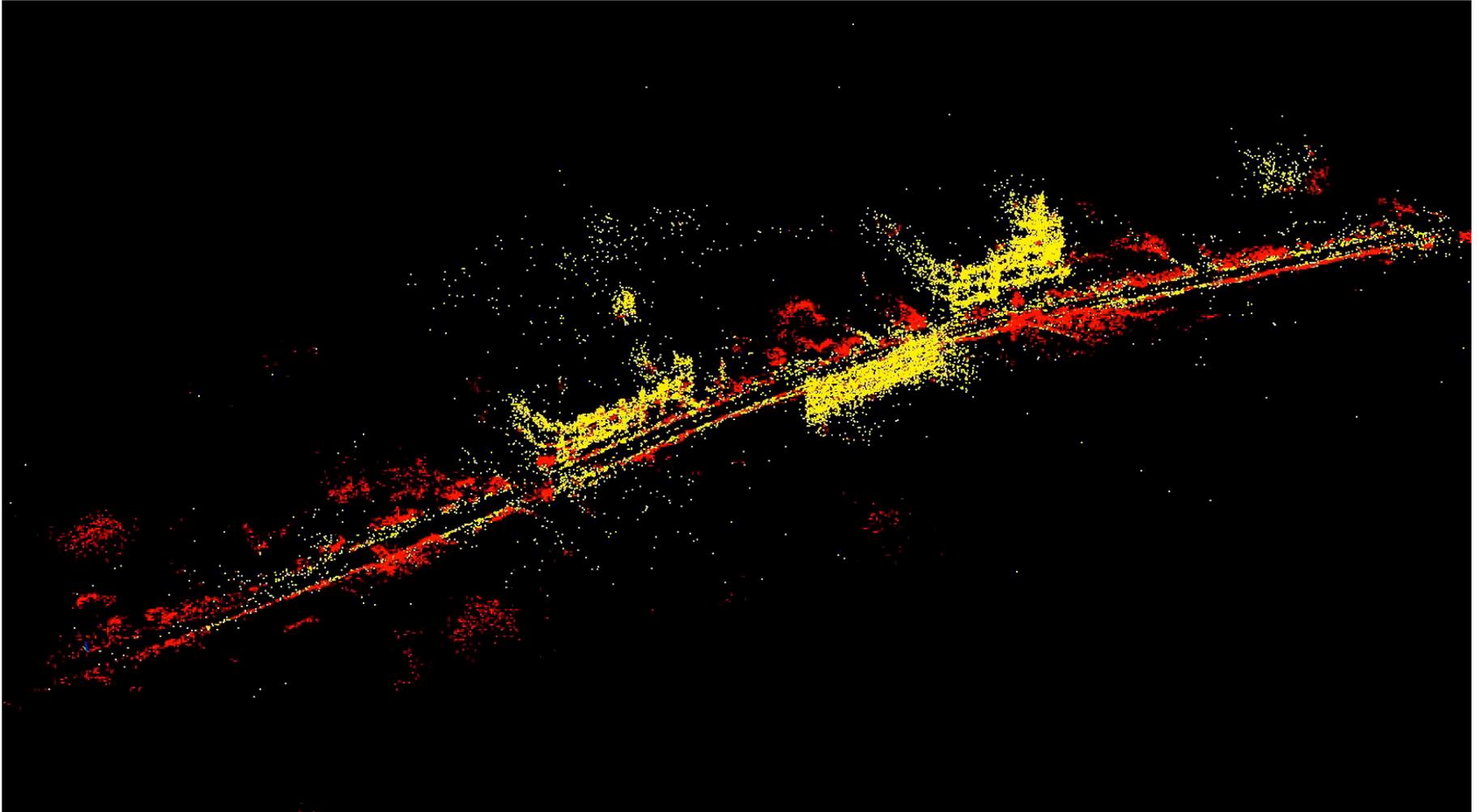
密な 3次元復元の例

- 岩手県釜石市, 地震後約1ヶ月



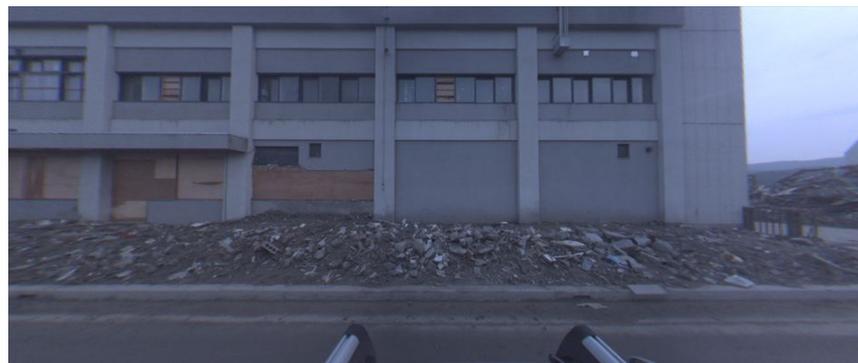
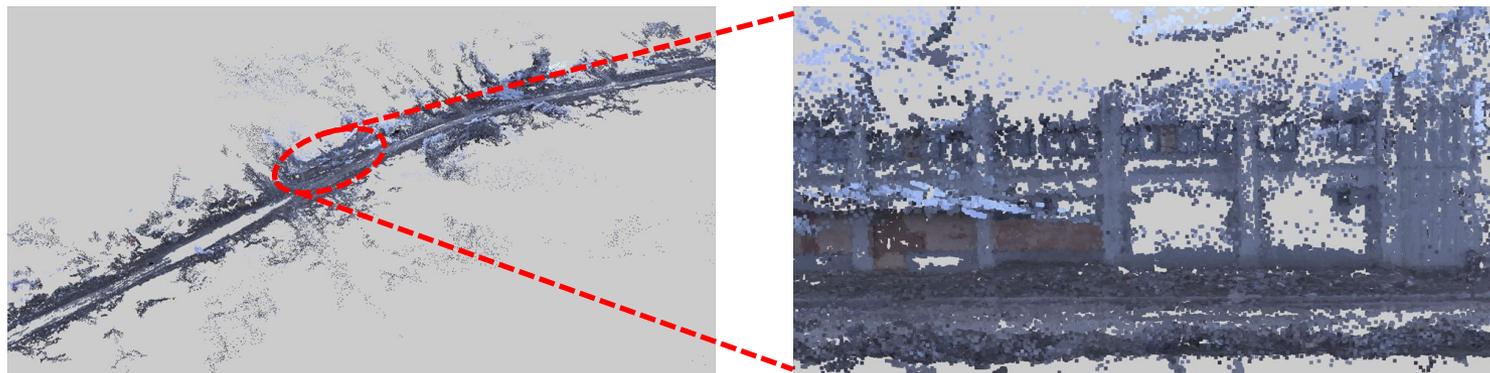
3次元形状の差分による変化検出

- 岩手県陸前高田市, 3ヶ月後の変化 (黄: 変化なし 赤: 消失 青: 出現)



既存手法の限界

- 復元できない部分が多い存在



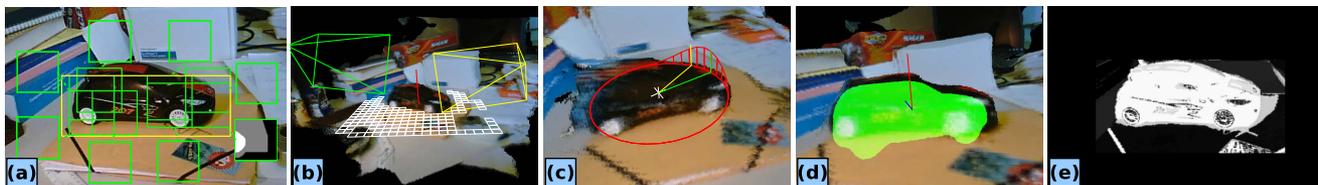
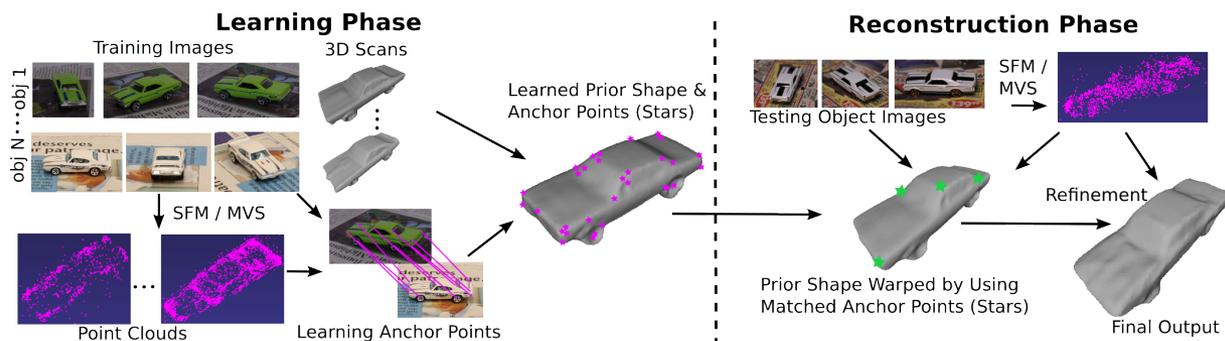
問題点

- 情報の不足

- 画像のテクスチャや視点数が不十分な場合シーンを密に復元できない

- 既存の機械学習手法(カテゴリーベース)

- “Dense Object Reconstruction with Semantic Priors”, Sid Yingze Bao, Manmohan Chandraker, Yuanqing Lin, Silvio Savarese, CVPR, 2013
- “Dense Reconstruction Using 3D Object Shape Priors”, Amaury Dame, Victor A. Prisacariu, Carl Y. Ren, Ian Reid, CVPR, 2013



ディープネットワークへの期待

- 幾何学的に十分な情報が得られない場合にも学習データから推定
 - 幾何学的な情報をどのようにネットワークの入力・出力として設定するか
 - 従来の解析的手法と学習手法の組み合わせ方法
 - データセットの作成
- 学習結果を解析的手法にフィードバック
 - 特徴量の抽出
 - e.g. SIFT, SURF
 - パッチの類似度計算
 - e.g. SSD, NCC

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- **解析的手法の現状と課題**
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

目次

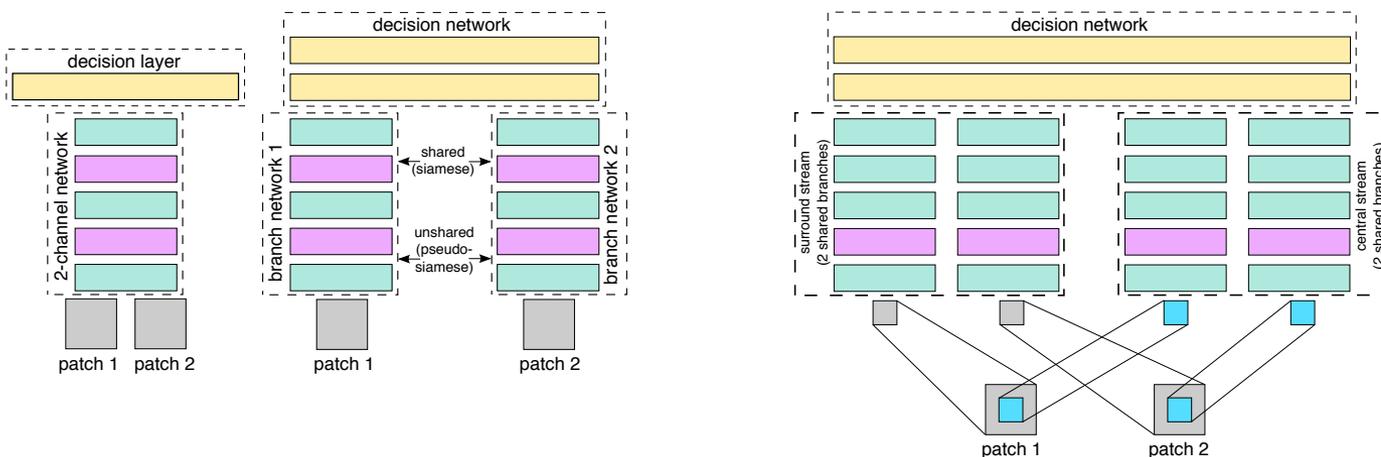
- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- **CNNのシーンモデリングへの応用**
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - **パッチの類似度計算**
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

CNNを用いたパッチの類似度推定

- パッチの類似度推定を様々なネットワーク構造で比較
 - 入力：2ch > 1ch (siamese)
 - スケール：マルチ > シングル



Train	Test	2ch-2stream	2ch-deep	2ch	siam	siam- l_2	pseudo-siam	pseudo-siam- l_2	siam-2stream	siam-2stream- l_2	[19]
Yos	ND	2.11	2.52	3.05	5.75	8.38	5.44	8.95	5.29	5.58	6.82
Yos	Lib	7.2	7.4	8.59	13.48	17.25	10.35	18.37	11.51	12.84	14.58
ND	Yos	4.1	4.38	6.04	13.23	15.89	12.64	15.62	10.44	13.02	10.08
ND	Lib	4.85	4.55	6.05	8.77	13.24	12.87	16.58	6.45	8.79	12.42
Lib	Yos	5	4.75	7	14.89	19.91	12.5	17.83	9.02	13.24	11.18
Lib	ND	1.9	2.01	3.03	4.33	6.01	3.93	6.58	3.05	4.54	7.22
mean		4.19	4.27	5.63	10.07	13.45	9.62	13.99	7.63	9.67	10.38
mean(1,4)		4.56	4.71	5.93	10.31	13.69	10.33	14.88	8.42	10.06	10.98

“Learning to Compare Image Patches via Convolutional Neural Networks”, Sergey Zagoruyko, Nikos Komodakis, CVPR, 2015

目次

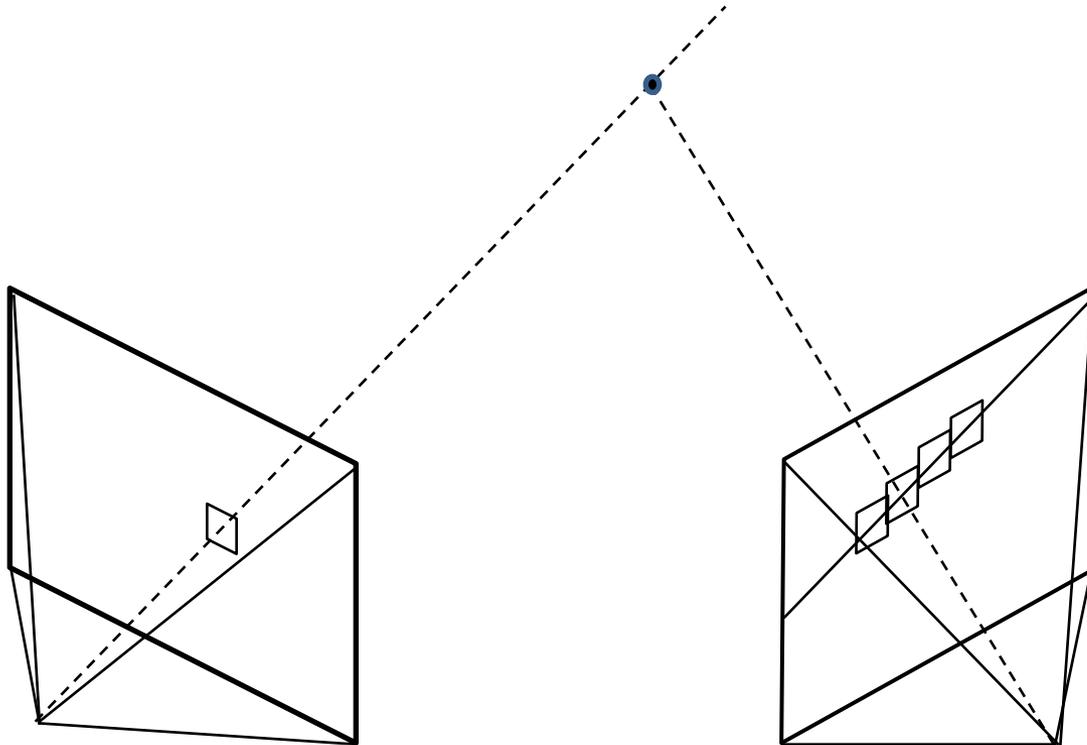
- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - **パッチの類似度計算**
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

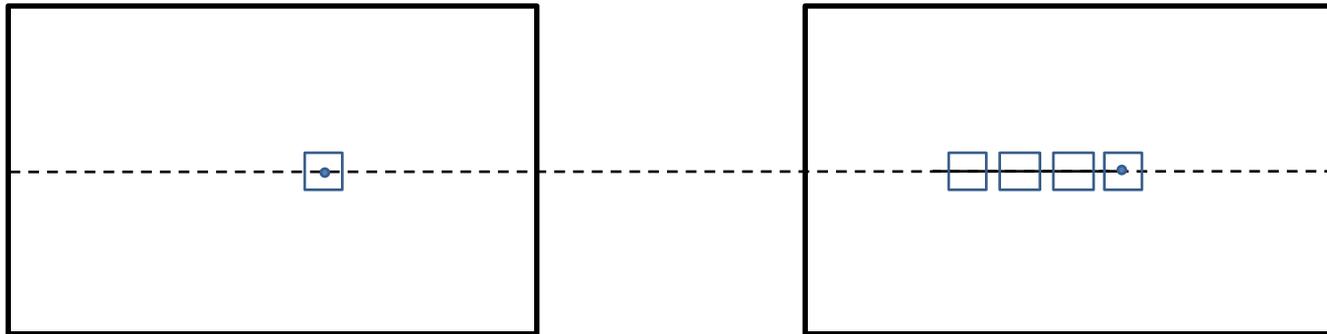
ステレオマッチング

- 多視点画像から各ピクセルのデプスを推定
 - エピポーラ線上のパッチの類似度を計算



平行ステレオ

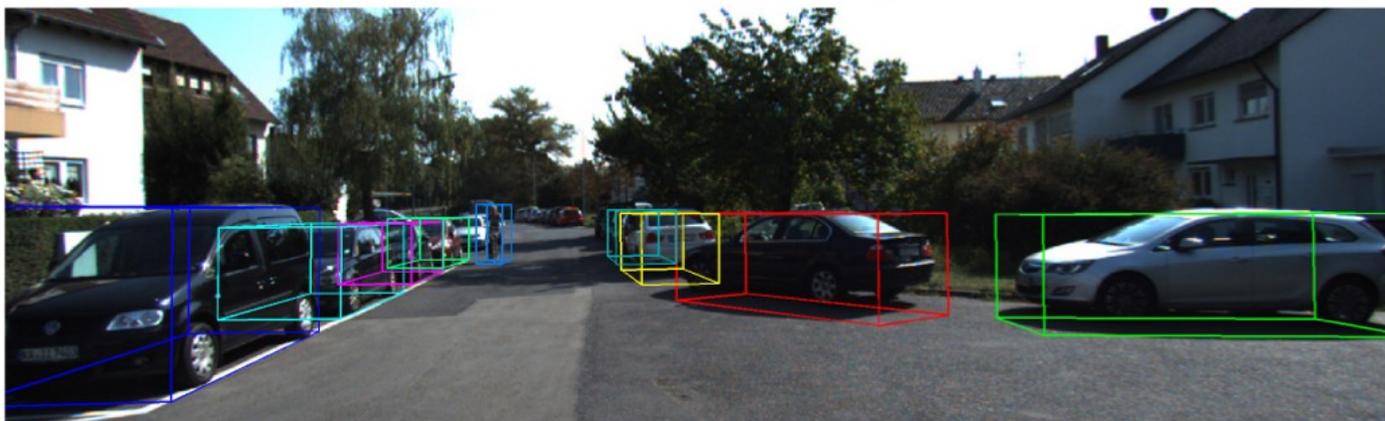
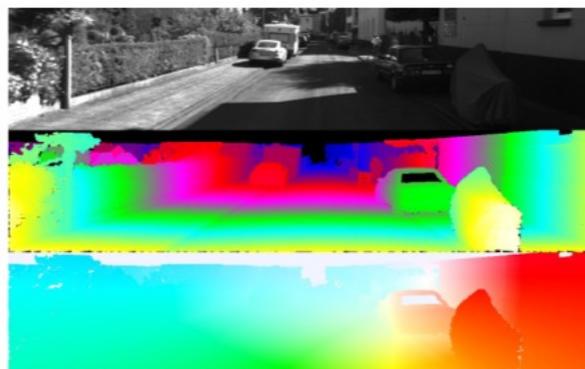
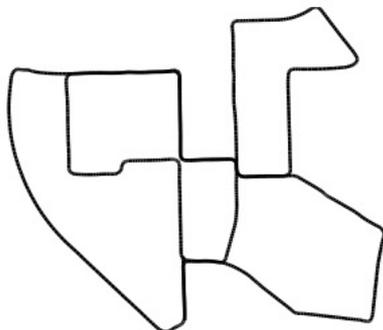
- 同一水平線上を探索してデプスを推定
 - 連続的にメモリへアクセスするため計算が高速



KITTIのベンチマークデータセット

- カメラ, レーザー, GPS, IMUの搭載車両で一般道を走行
 - ステレオ, オプティカルフロー, ビジュアルオドメトリ, 3次元物体検出・追跡

360° Velodyne Laserscanner
Stereo Camera Rig



“Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite”, Andreas Geiger, Philip Lenz, Raquel Urtasun, CVPR, 2012

KITTIのベンチマークデータセット

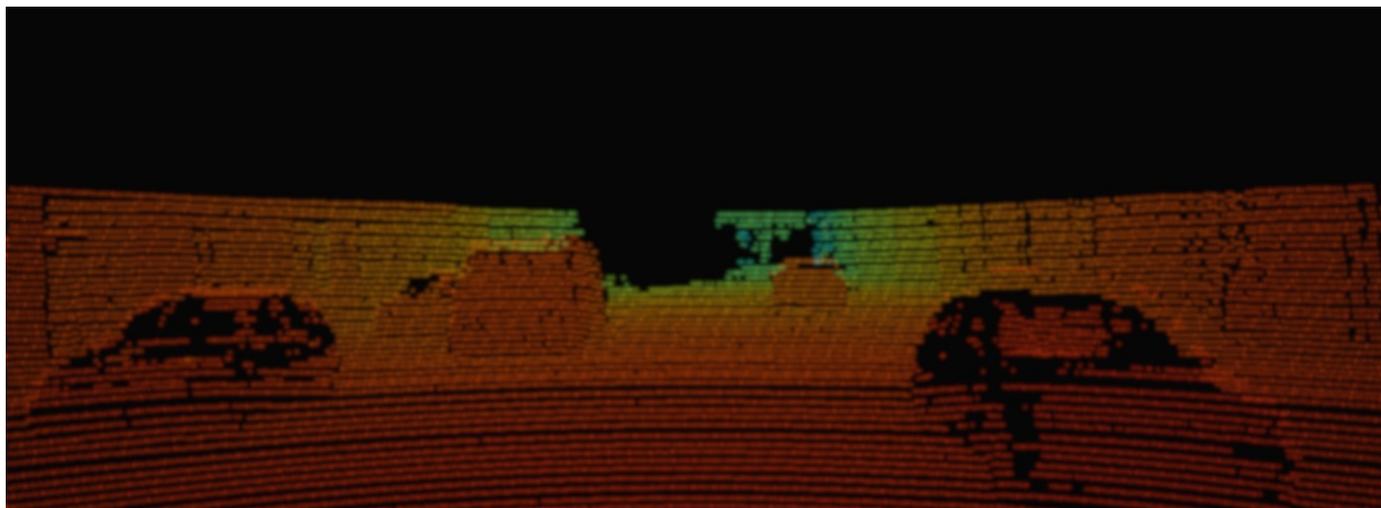
- ステレオマッチングの場合
 - 3次元点群データを視差マップに変換



左画像



右画像



レーザーセンサの3次元点群を左画像に投影

KITTIのステレオベンチマーク

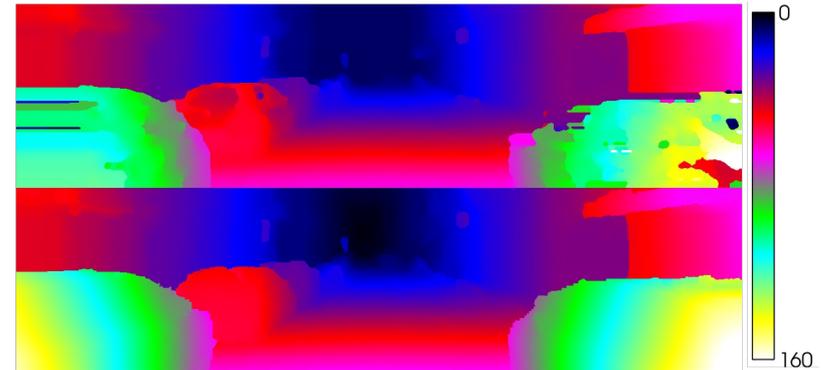


	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment
1	Displets v2		code	2.37 %	3.09 %	0.7 px	0.8 px	100.00 %	265 s	>8 cores @ 3.0 Ghz (Matlab + C/C++)
F. Guey and A. Geiger: Displets: Resolving Stereo Ambiguities using Object Knowledge . Conference on Computer Vision and Pattern Recognition (CVPR) 2015.										
2	MC-CNN-acrt		code	2.43 %	3.63 %	0.7 px	0.9 px	100.00 %	67 s	Nvidia GTX Titan X (CUDA, Lua/Torch7)
J. Zbontar and Y. LeCun: Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches . Submitted to JMLR.										
3	Displets		code	2.47 %	3.27 %	0.7 px	0.9 px	100.00 %	265 s	>8 cores @ 3.0 Ghz (Matlab + C/C++)
F. Guey and A. Geiger: Displets: Resolving Stereo Ambiguities using Object Knowledge . Conference on Computer Vision and Pattern Recognition (CVPR) 2015.										
4	MC-CNN			2.61 %	3.84 %	0.8 px	1.0 px	100.00 %	100 s	Nvidia GTX Titan (CUDA, Lua/Torch7)
J. Zbontar and Y. LeCun: Computing the Stereo Matching Cost with a Convolutional Neural Network . Conference on Computer Vision and Pattern Recognition (CVPR) 2015.										
5	PRSM		code	2.78 %	3.00 %	0.7 px	0.7 px	100.00 %	300 s	1 core @ 2.5 Ghz (C/C++)
C. Vogel, K. Schindler and S. Roth: 3D Scene Flow Estimation with a Piecewise Rigid Scene Model . ijcv 2015.										
6	SPS-StFl			2.83 %	3.64 %	0.8 px	0.9 px	100.00 %	35 s	1 core @ 3.5 Ghz (C/C++)
K. Yamaguchi, D. McAllester and R. Urtasun: Efficient Joint Segmentation, Occlusion Labeling, Stereo and Flow Estimation . ECCV 2014.										
7	VC-SF			3.05 %	3.31 %	0.8 px	0.8 px	100.00 %	300 s	1 core @ 2.5 Ghz (C/C++)
C. Vogel, S. Roth and K. Schindler: View-Consistent 3D Scene Flow Estimation over Multiple Frames . Proceedings of European Conference on Computer Vision. Lecture Notes in, Cor										
8	Deep Embed			3.10 %	4.24 %	0.9 px	1.1 px	100.00 %	3 s	1 core @ 2.5 Ghz (C/C++)
Z. Chen, X. Sun, Y. Yu, L. Wang and C. Huang: A Deep Visual Correspondence Embedding Model for Stereo Matching Costs . ICCV 2015.										
9	JSOSM			3.15 %	3.94 %	0.8 px	0.9 px	100.00 %	105 s	8 cores @ 2.5 Ghz (C/C++)
Anonymous submission										
10	OSF		code	3.28 %	4.07 %	0.8 px	0.9 px	99.98 %	50 min	1 core @ 3.0 Ghz (Matlab + C/C++)
M. Menze and A. Geiger: Object Scene Flow for Autonomous Vehicles . Conference on Computer Vision and Pattern Recognition (CVPR) 2015.										
11	CoR		code	3.30 %	4.10 %	0.8 px	0.9 px	100.00 %	6 s	6 cores @ 3.3 Ghz (Matlab + C/C++)
A. Chakrabarti, Y. Xiong, S. Gortler and T. Zickler: Low-level Vision by Consensus in a Spatial Hierarchy of Regions . CVPR 2015.										
12	TCD-CRF			3.32 %	5.24 %	0.9 px	1.9 px	100.00 %	60 s	4 cores @ 3.5 Ghz (C/C++)
Anonymous submission										
13	SPS-St		code	3.39 %	4.41 %	0.9 px	1.0 px	100.00 %	2 s	1 core @ 3.5 Ghz (C/C++)
K. Yamaguchi, D. McAllester and R. Urtasun: Efficient Joint Segmentation, Occlusion Labeling, Stereo and Flow Estimation . ECCV 2014.										

関連研究

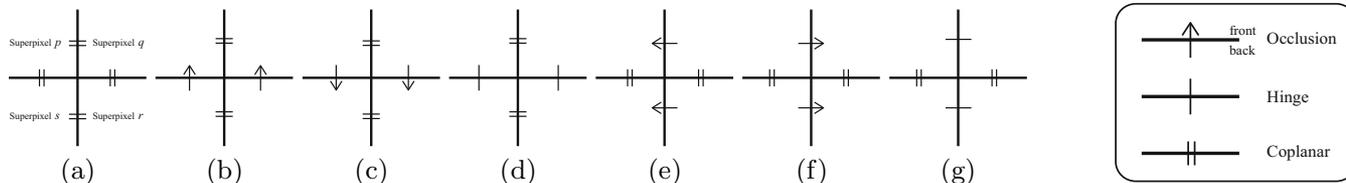
- テクスチャレスや透明な領域を物体形状の事前知識から推定

- “Displets: Resolving Stereo Ambiguities using Object Knowledge”, Fatma Guney, Andreas Geiger, CVPR, 2015



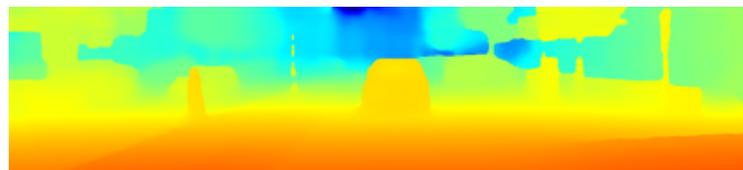
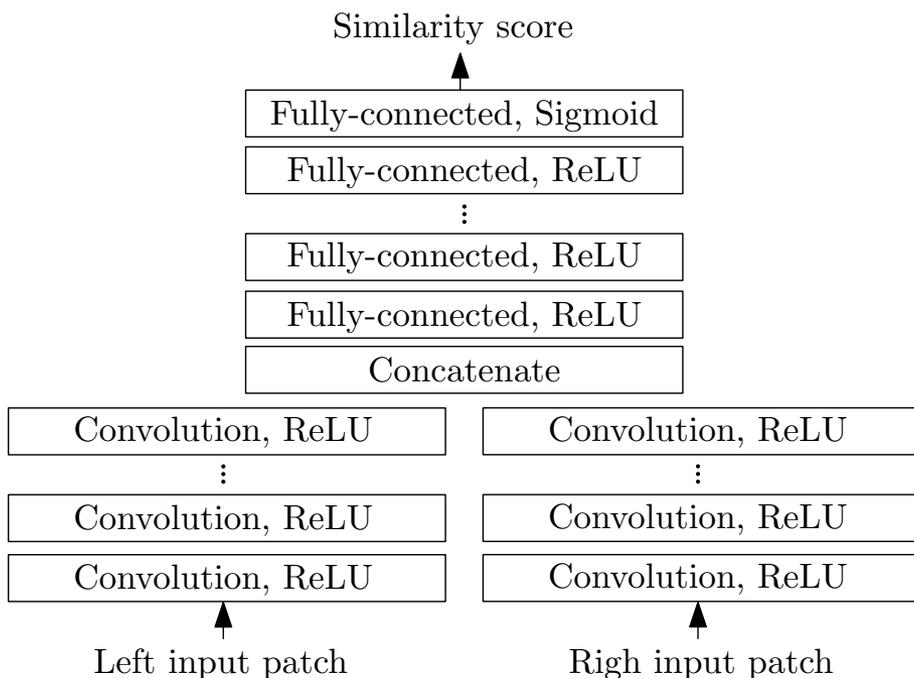
- 局所平面の遮蔽関係とデプスをMRFで同時推定

- “Continuous markov random fields for robust stereo estimation”, K. Yamaguchi, T. Hazan, D. A. McAllester, and R. Urtasun. ECCV2012



CNNを用いたステレオマッチング（例1）

- 基本は従来の枠組みと同じ
 - （例）Semi-Global Matching, Cost Aggregation
- パッチの類似度をCNNで推定



デプスの推定結果

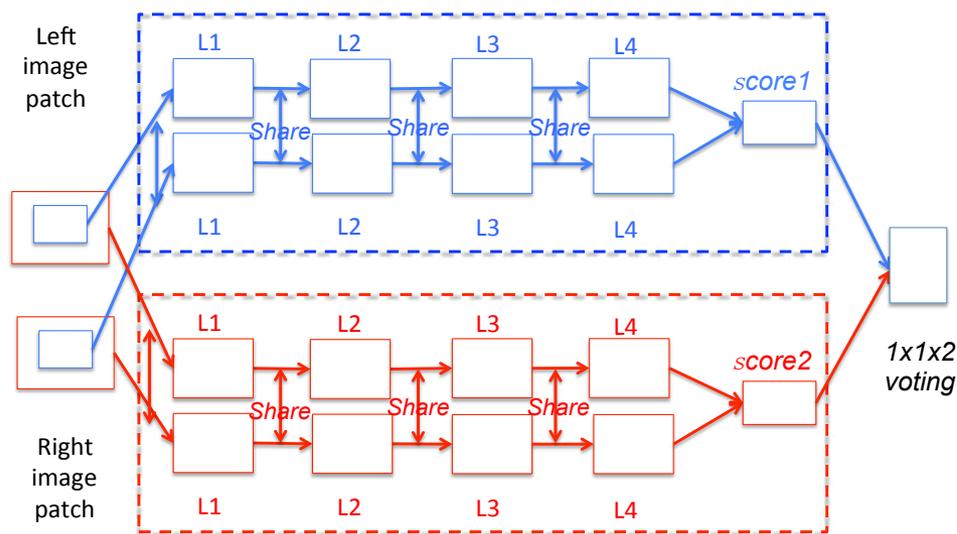


推定誤差（赤：誤差大）

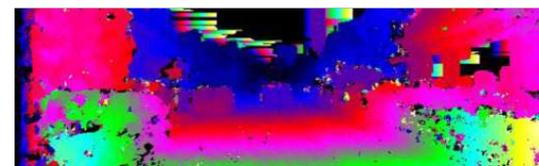
“Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches”,
Zbontar, Jure, and Yann LeCun, arXiv preprint arXiv:1510.05970, 2015

CNNを用いたステレオマッチング（例2）

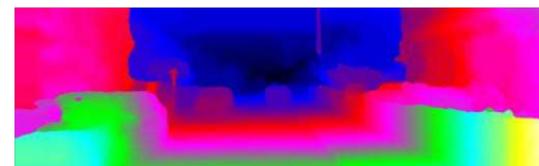
- 基本は従来の枠組みと同じ
- 異なるスケールでパッチの類似度を学習
 - ダウンサンプルしたパッチで広域の情報を考慮



(a) Reference image of a stereo pair



(b) Winner-takes-all result



(c) Result from an MRF-based stereo framework



(d) color-coded depth

“A Deep Visual Correspondence Embedding Model for Stereo Matching Costs”, Zhuoyuan Chen, Xun Sun, Liang Wang, Yinan Yu, Chang Huang, ICCV, 2015

目次

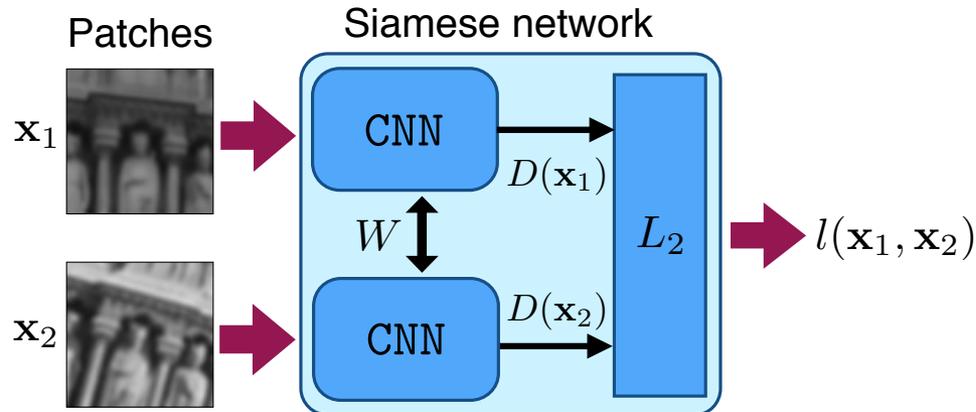
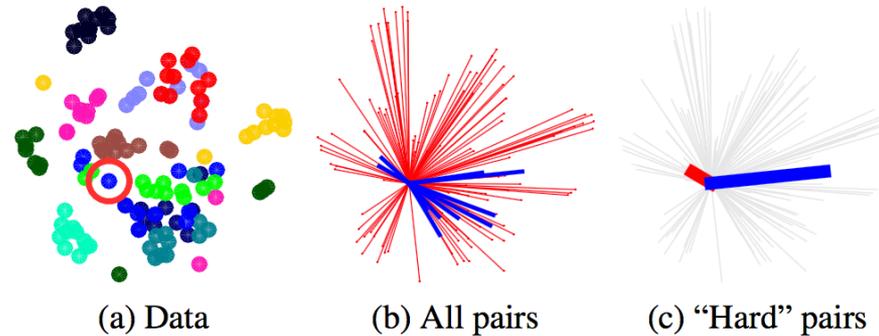
- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - **特徴点マッチング**
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

CNNを用いた特徴点マッチング

- 識別が難しい対応点を用いてパッチの類似度を学習
 - 距離が最も遠いポジティブサンプルと近いネガティブサンプルを利用



目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - **特徴点マッチング**
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

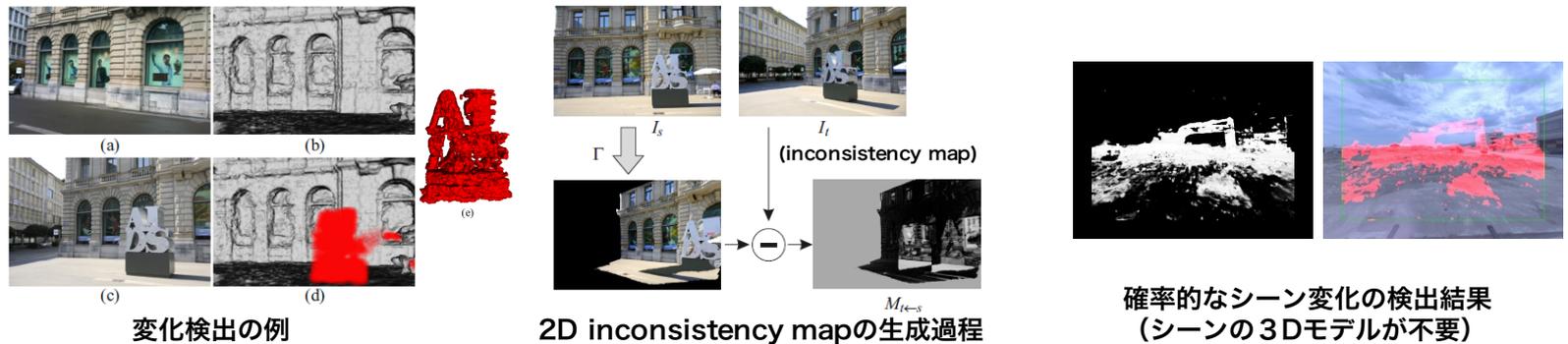
目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - **変化検出**
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

シーン変化の検出

• 3次元構造の変化

- “Image Based Detection of Geometric Changes in Urban Environments”, Aparna Taneja, Luca Ballan, Marc Pollefeys, ICCV, 2011
- “City-Scale Change Detection in Cadastral 3D Models using Images”, Aparna Taneja, Luca Ballan, Marc Pollefeys, CVPR, 2013
- “Detecting Changes in 3D Structure of a Scene from Multi-view Images Captured by a Vehicle-mounted Camera”, Ken Sakurada, Takayuki Okatani, Koichiro Deguchi., CVPR, 2013



• テクスチャの変化

- “Scene Chronology”, Kevin Matzen and Noah Snavely, ECCV, 2014



CNNを用いたシーン変化の検出

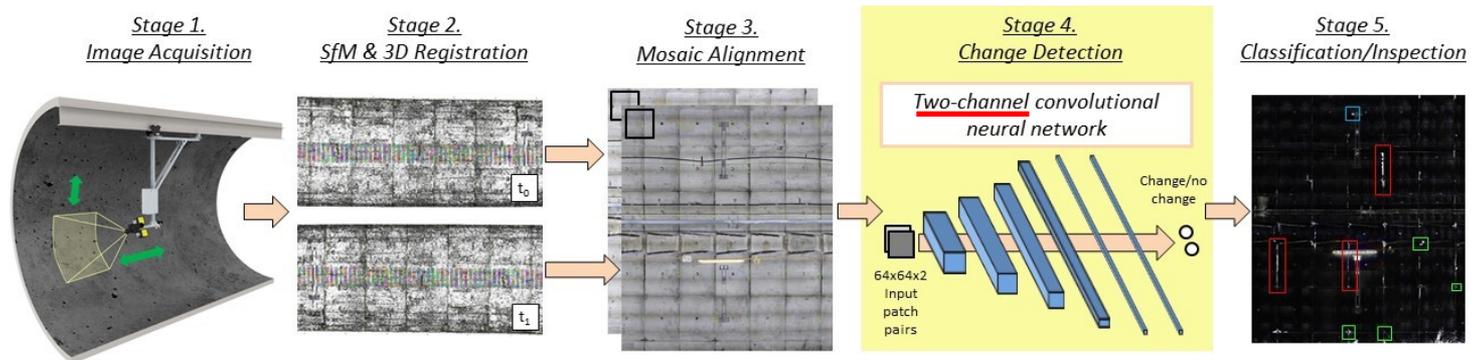
• 物体レベルの変化

- “Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation”, Ken Sakurada, Takayuki Okatani, BMVC, 2015



• クラックなどの微細な変化

- “Detecting Change for Multi-View, Long-Term Surface Inspection”, Simon Stent, Riccardo Gherardi, Bjorn Stenger and Roberto Cipolla, BMCV, 2015



CNNを用いたシーン変化の検出

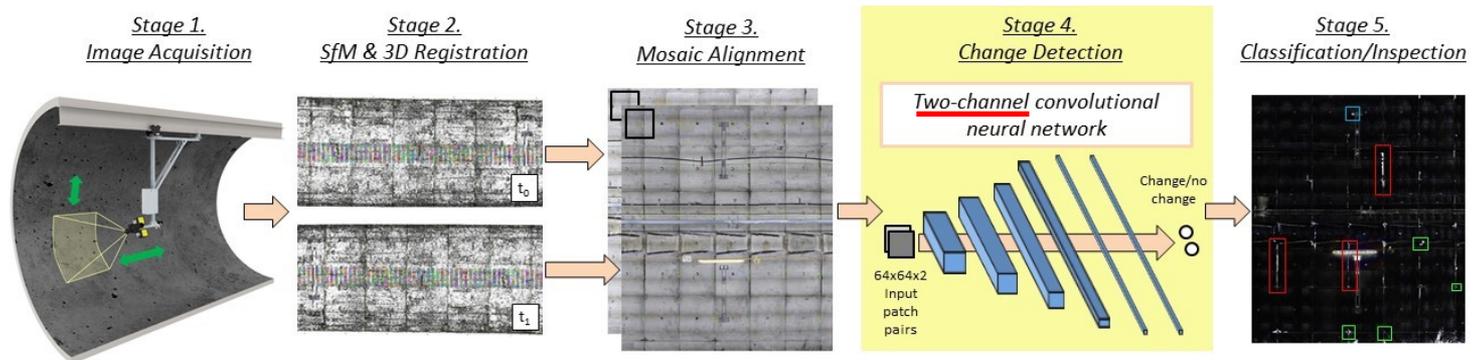
• 物体レベルの変化

- “Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation”, Ken Sakurada, Takayuki Okatani, BMVC, 2015



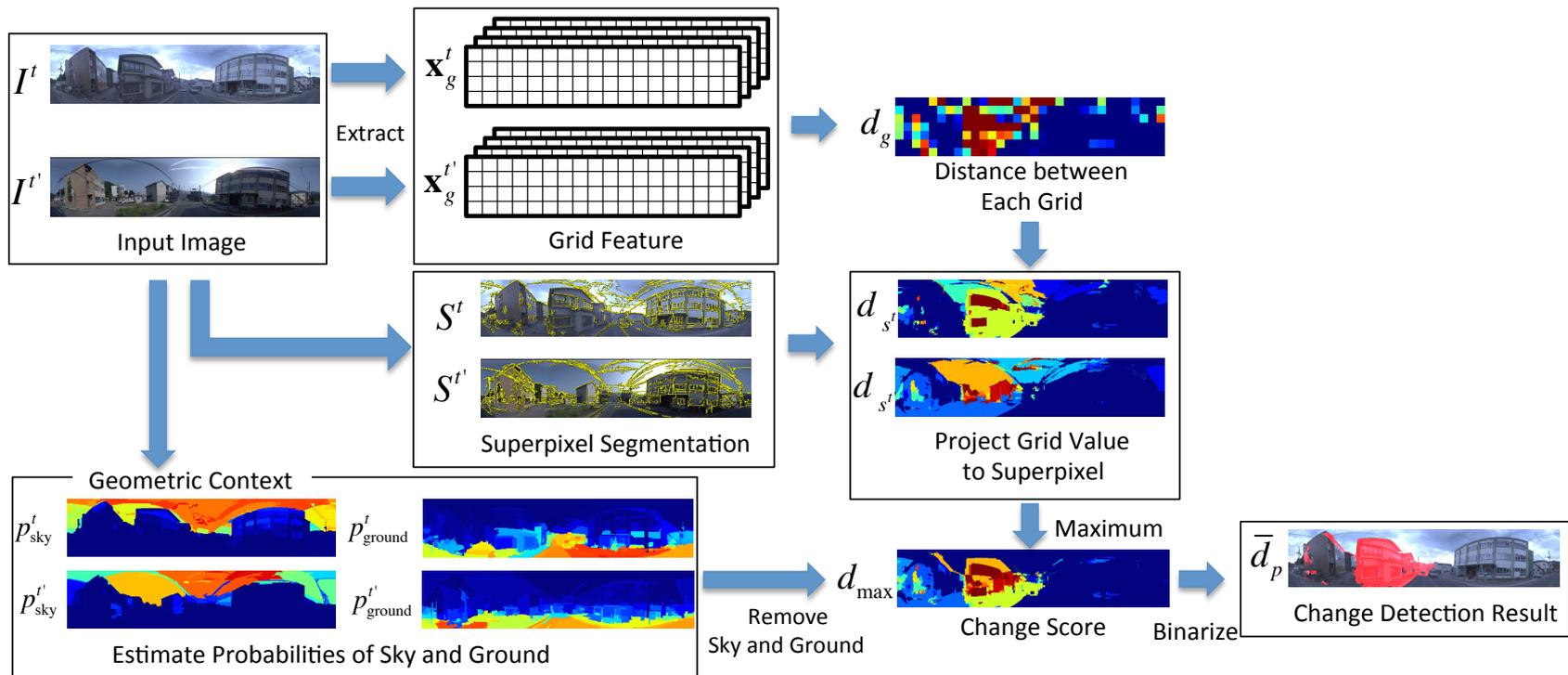
• クラックなどの微細な変化

- “Detecting Change for Multi-View, Long-Term Surface Inspection”, Simon Stent, Riccardo Gherardi, Bjorn Stenger and Roberto Cipolla, BMCV, 2015



CNNを用いたシーン変化の検出

- 一般物体認識用にトレーニングしたネットワークを転用
 - 上位レイヤーの特徴マップで大まかに変化を検出
 - スーパーピクセルセグメンテーションで正確な境界を推定



“Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation”, Ken Sakurada, Takayuki Okatani, BMVC, 2015

CNNを用いたシーン変化の検出

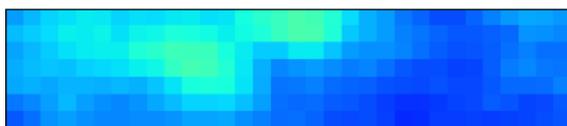


Input (database)

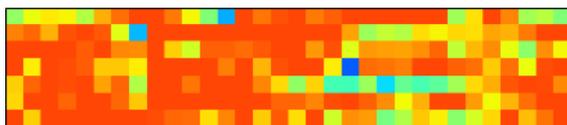


Hand-labeled ground-truth

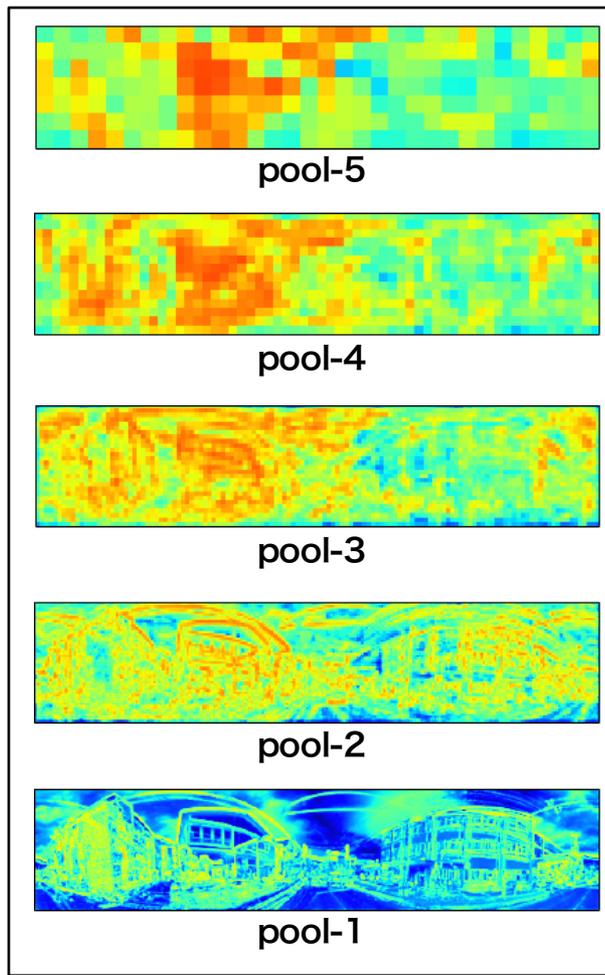
Baseline



Dense-SIFT



Local-patch



Change

Not Change

“Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation”,
Ken Sakurada, Takayuki Okatani, BMVC, 2015

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - **変化検出**
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

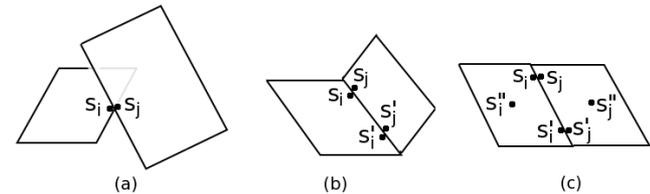
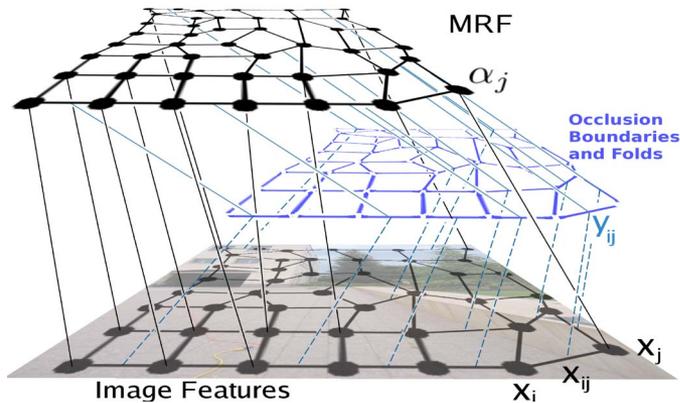
目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - **単一視点画像のデプス推定**
 - 視点補間
- まとめ

単一視点画像のデプス推定

- MRFで各パッチの平面パラメーターを推定

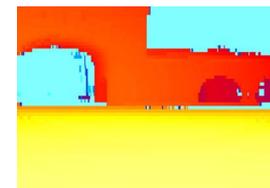
- “Learning 3-D Scene Structure from a Single Still Image”, Ashutosh Saxena, Min Sun and Andrew Y. Ng, TPAMI2009



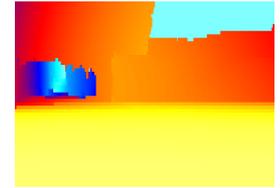
平面の隣接関係



入力



グランドトゥールース



提案手法

- 大まかなカテゴリーに分類 (地面, 空, 垂直平面)

- “Automatic photo, pop-up”, D. Hoiem, A. A. Efros, and M. Hebert, SIGGRAPH, 2005



(a) input image



(b) superpixels



(c) constellations



(d) labeling

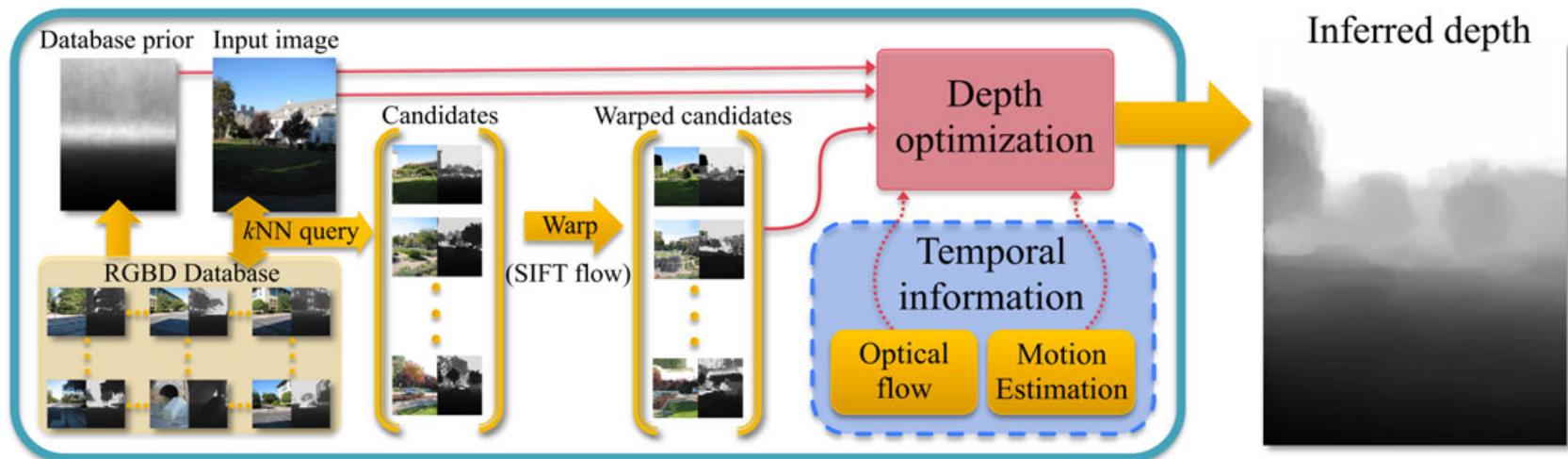


(e) novel view

単一視点画像のデプス推定

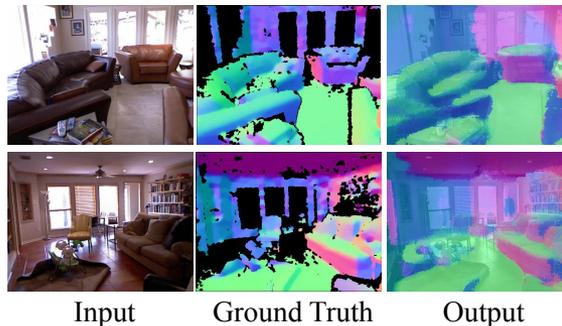
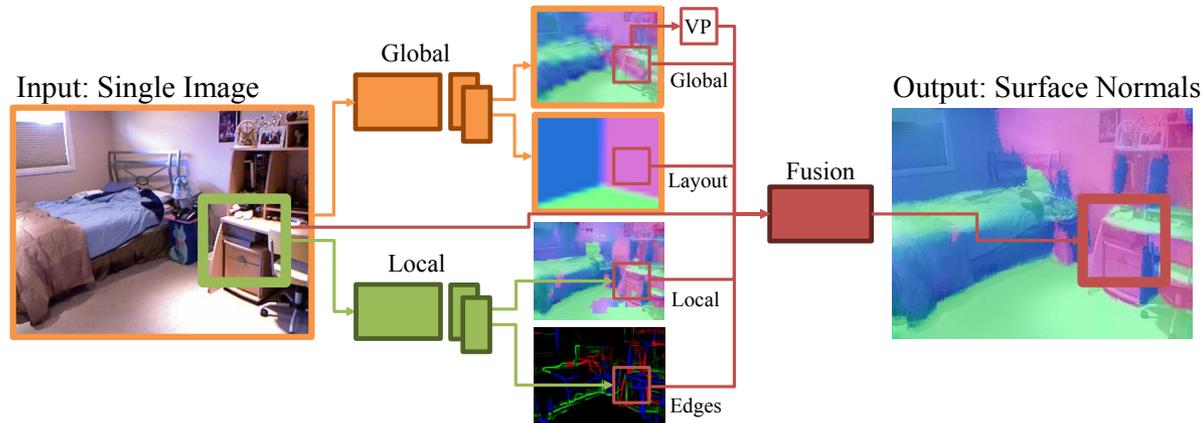
- データベースから類似シーンを検索

- “DepthTransfer: Depth extraction from video using non-parametric sampling”, K. Karsch, C. Liu, and S. B. Kang, TPAMI2014
- “2D-to-3D Image Conversion by Learning Depth from Examples“, J. Konrad, M. Wang, and P. Ishwar. CVPR Workshops 2012



CNNを用いた単一視点画像のデプス推定（例1）

- マルチスケールで法線ベクトルを推定
 - グローバルパッチ + レイアウト
 - ローカルパッチ + エッジ

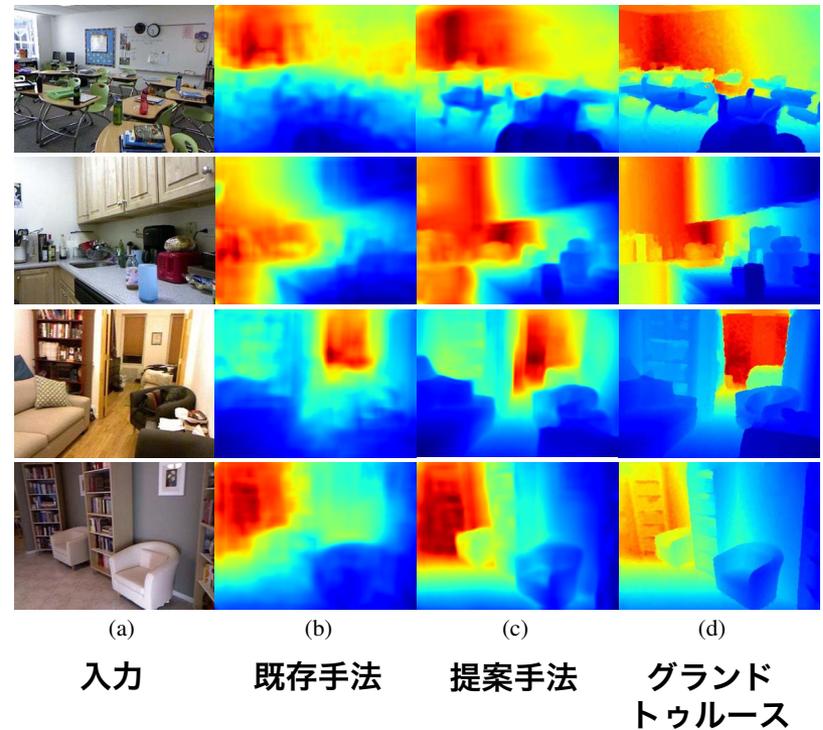
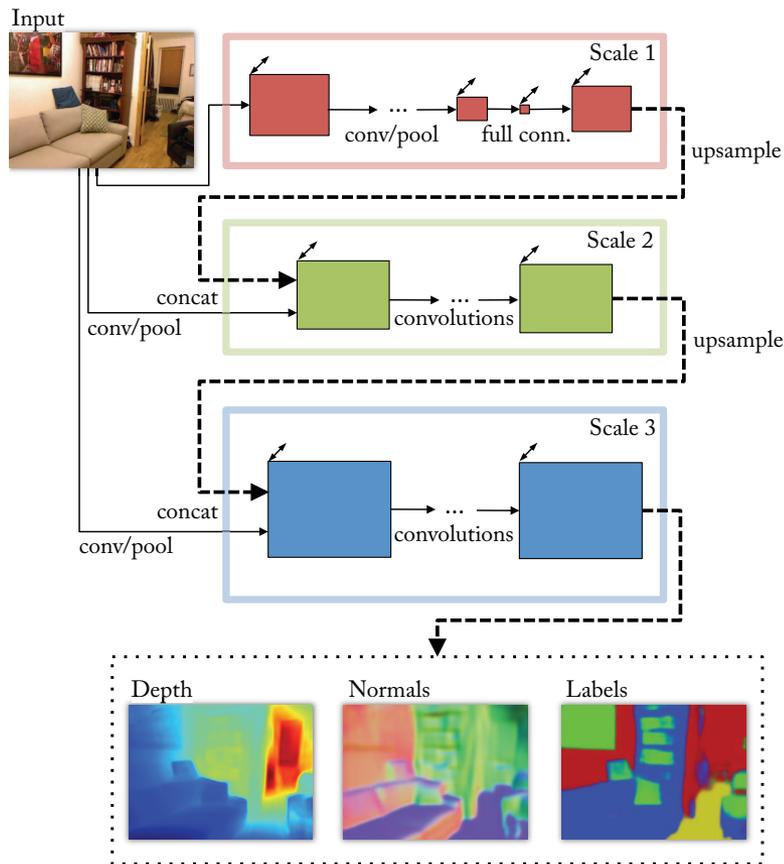


	Mean	Median	11.25°	22.5°	30°
Full	26.9	14.8	42.0	61.2	68.2
Full w/o Global	28.8	17.7	34.6	57.8	66.0
Fusion (+VP)	27.3	15.6	40.2	60.1	67.5
Fusion (+Edge)	27.8	16.4	37.5	59.4	67.4
Fusion (+Layout)	27.7	16.0	38.8	59.9	67.4
Fusion	27.9	16.6	37.4	59.2	67.1
Local	34.0	25.1	25.6	46.4	56.2
Global	30.9	20.8	31.4	52.3	60.5
Coarse CNN [8]	30.1	24.7	24.1	46.4	57.9

“Designing Deep Networks for Surface Normal Estimation”, X. Wang, D. F. Fouhey, and A. Gupta, CVPR2015

CNNを用いた単一視点画像のデプス推定（例2）

- デプス，法線ベクトル，カテゴリーラベルを同時推定



“Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture”, David Eigen, Rob Fergus, ICCV2015

目次

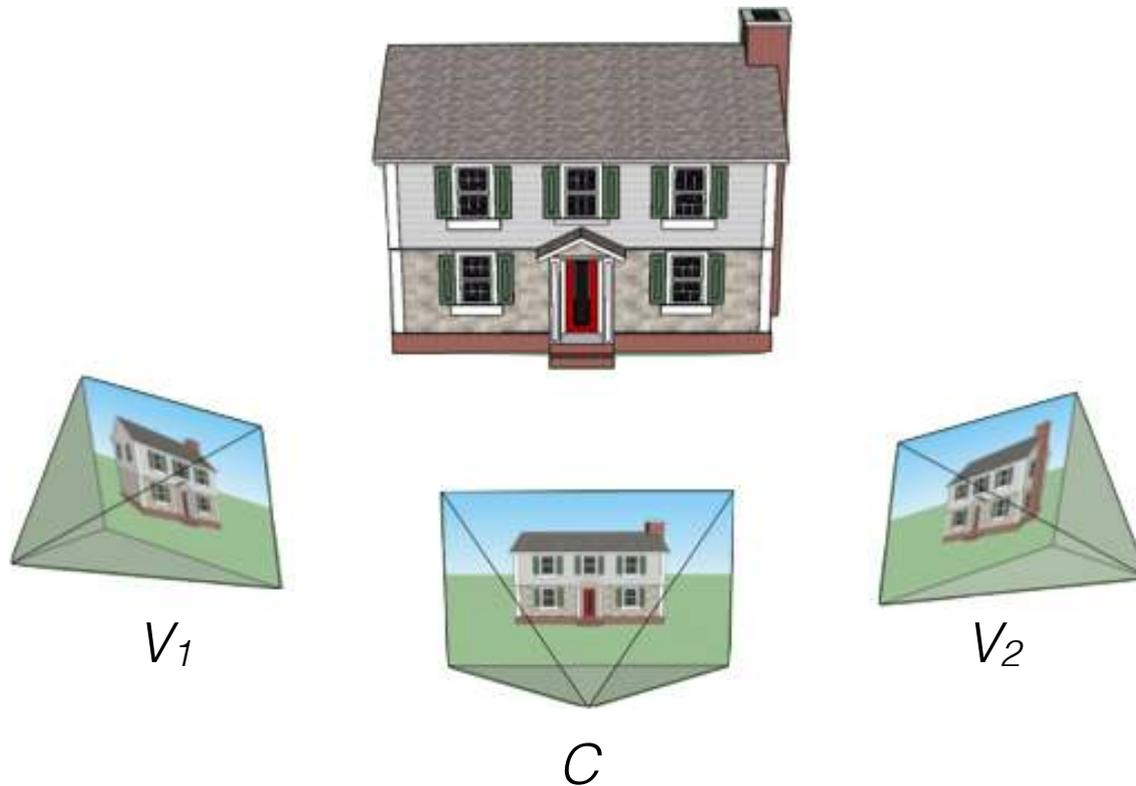
- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - **単一視点画像のデプス推定**
 - 視点補間
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

視点補間

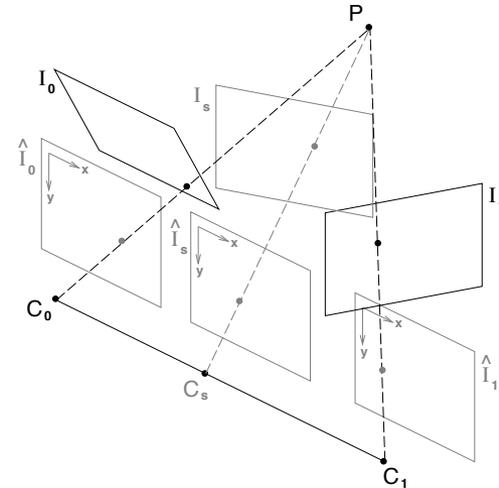
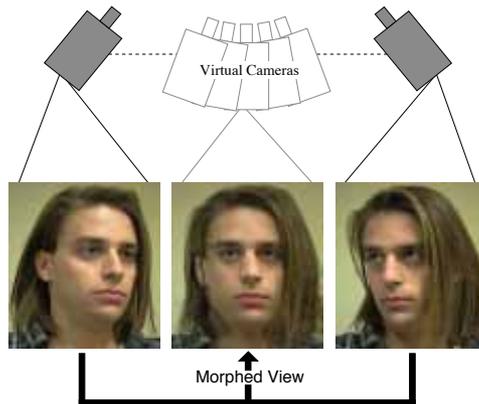
- 既存の視点 V_1 , V_2 の画像から新しい視点 C の画像を生成



関連研究

- モーフィング

- “View morphing”, S. M. Seitz and C. R. Dyer, SIGGRAPH1996



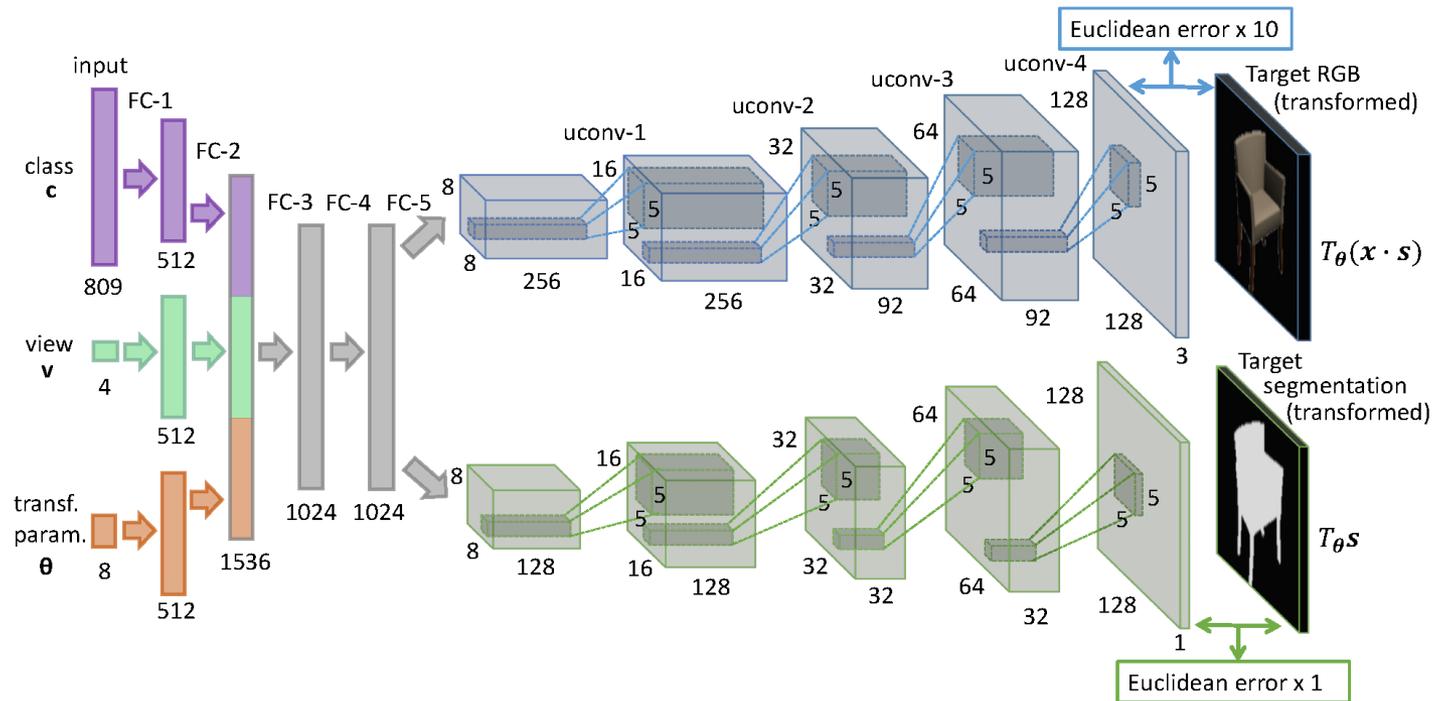
- 陽な形状復元とアピアランスの推定

- “The visual turing test for scene reconstruction”, Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. M. Seitz. 3DV2013
- “Three-dimensional scene flow”, S. Vedula, P. Rander, R. Collins, and T. Kanade, TPAMI2005



CNNを用いた視点補間（例1）

- 一般物体認識のネットワークを逆さにして利用
 - クラス, 視点, 画像変形のパラメータを入力しその画像を生成
 - セグメンテーションマスクのコスト関数を導入し精度向上
 - unpooling+ convでアップサンプリング

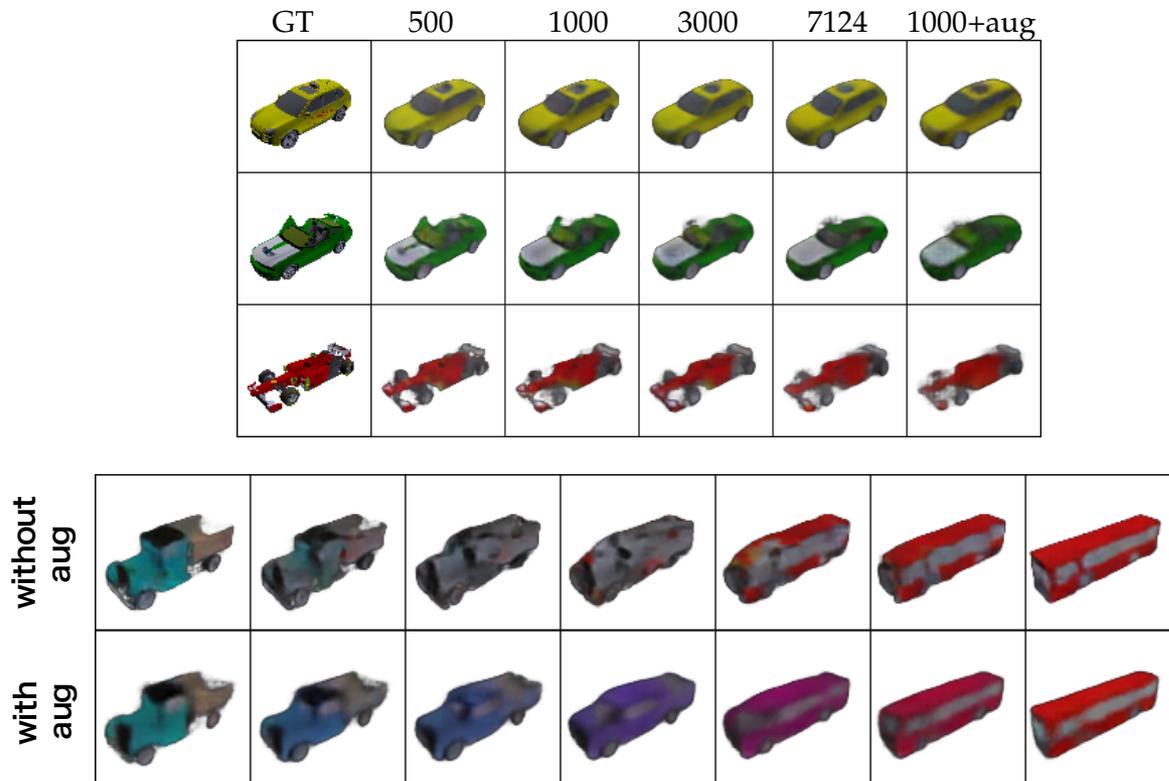


“Learning to Generate Chairs With Convolutional Neural Networks”, Alexey Dosovitskiy, Jost Tobias Springenberg, Thomas Brox, CVPR, 2015

CNNを用いた視点補間（例1）

- 学習時のモデル数

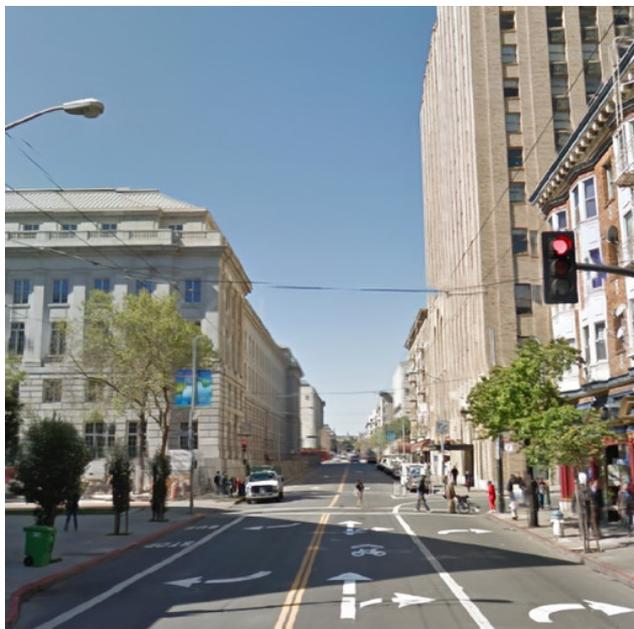
- 多いと生成された形状は不鮮明
- データ拡張するとモデル間のモーフィングがスムーズ（一般化）



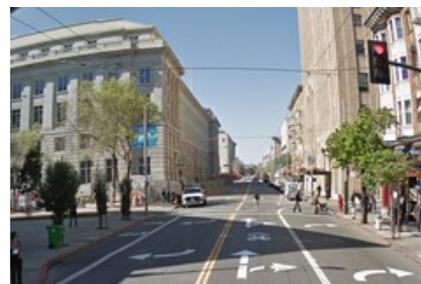
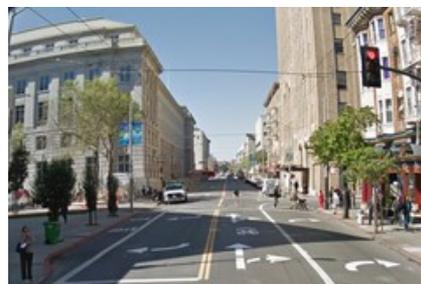
“Learning to Generate Chairs With Convolutional Neural Networks”, Alexey Dosovitskiy, Jost Tobias Springenberg, Thomas Brox, CVPR, 2015

CNNを用いた視点補間（例2）

- 姿勢が既知の大量の画像からシステム全体を学習
 - カメラの幾何学的関係を陰に含むようにプレーンスweepを利用



生成された新しい視点の画像



入力画像の一部

“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

DeepStereo

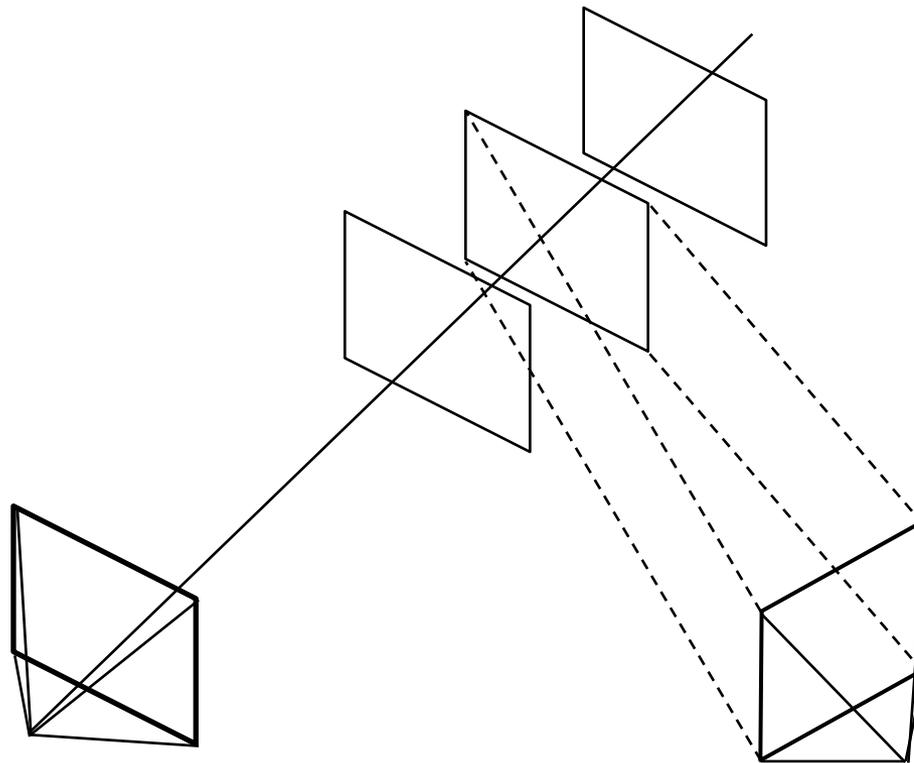
DeepStereo:

Learning to Predict New Views From the World's Imagery

“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

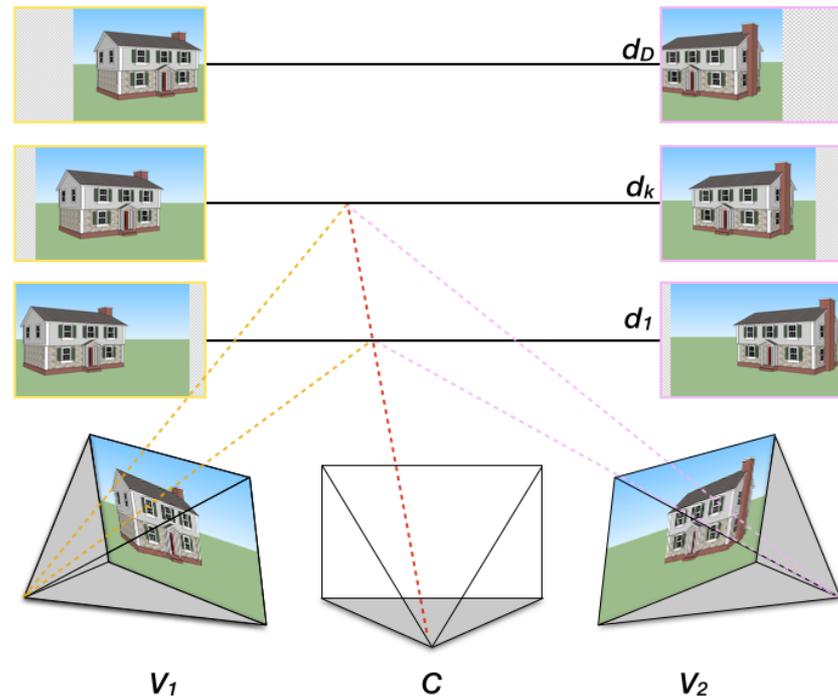
プレーンスweepステレオ

- 各奥行きに対してキーフレームに平行な面を仮定してペア画像を投影
 - エピポーラ線上の探索 → 画像ボリュームの奥行き方向の探索



プレーンスイープステレオ

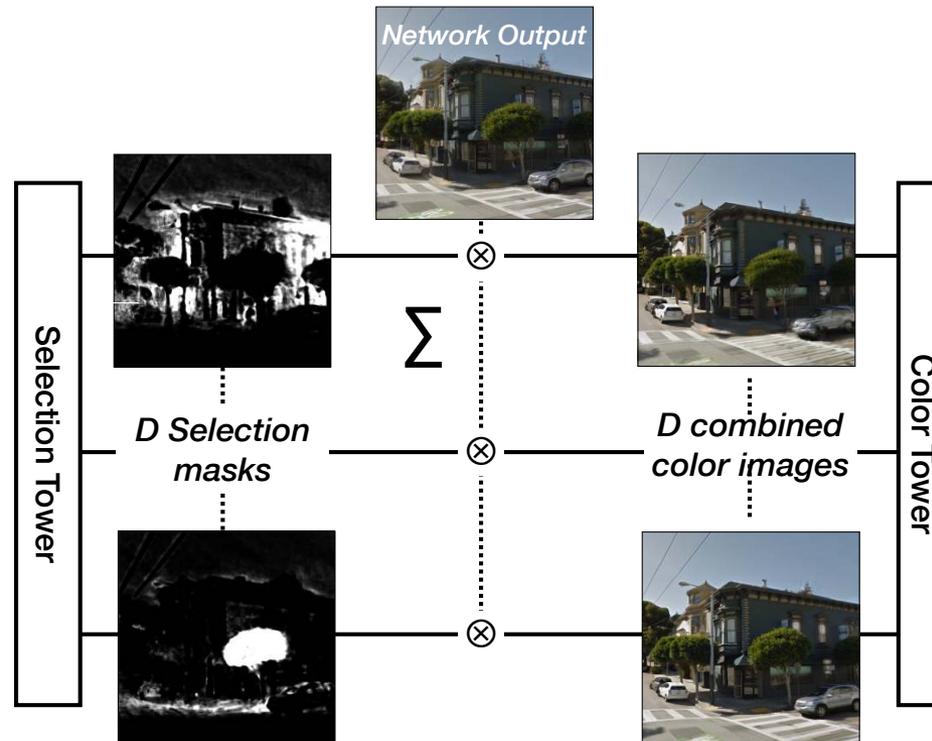
- 既存の視点 V_1 , V_2 の画像 I_1 , I_2 を新しい視点 C の画像に投影
 - 各デプス $d \in d_1 \dots d_D$ ごとに投影
 - 各視点の画像ボリュームをデプス方向に連結して入力



“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

基本的なネットワーク構造

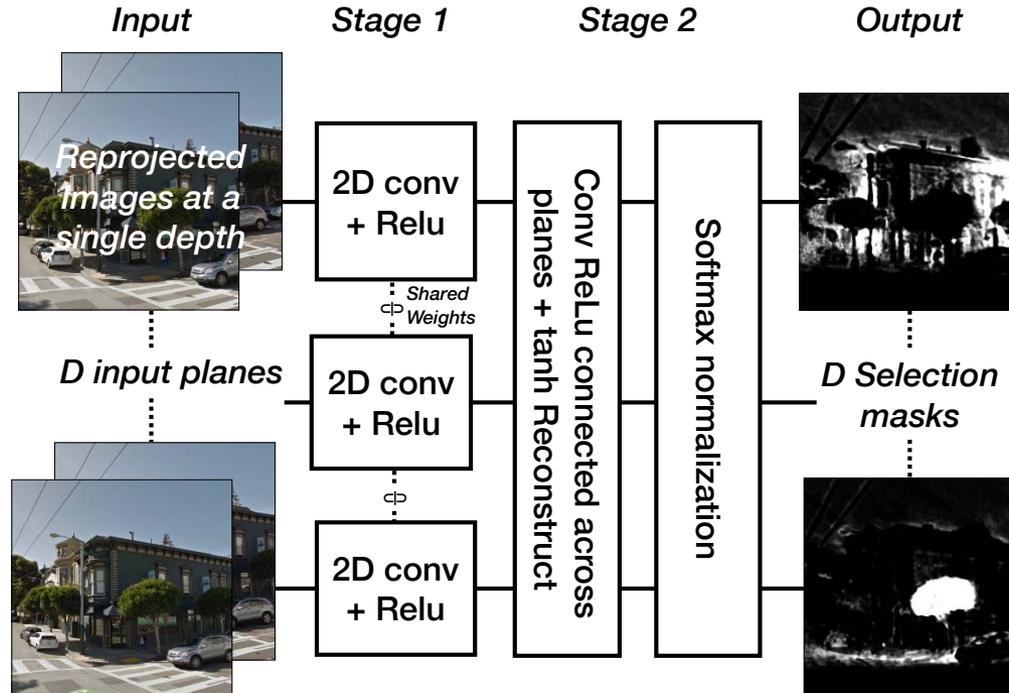
- セレクションタワーとカラータワー
 - 各タワーの出力の要素積
 - デプス方向に和を計算



“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

セレクションタワー

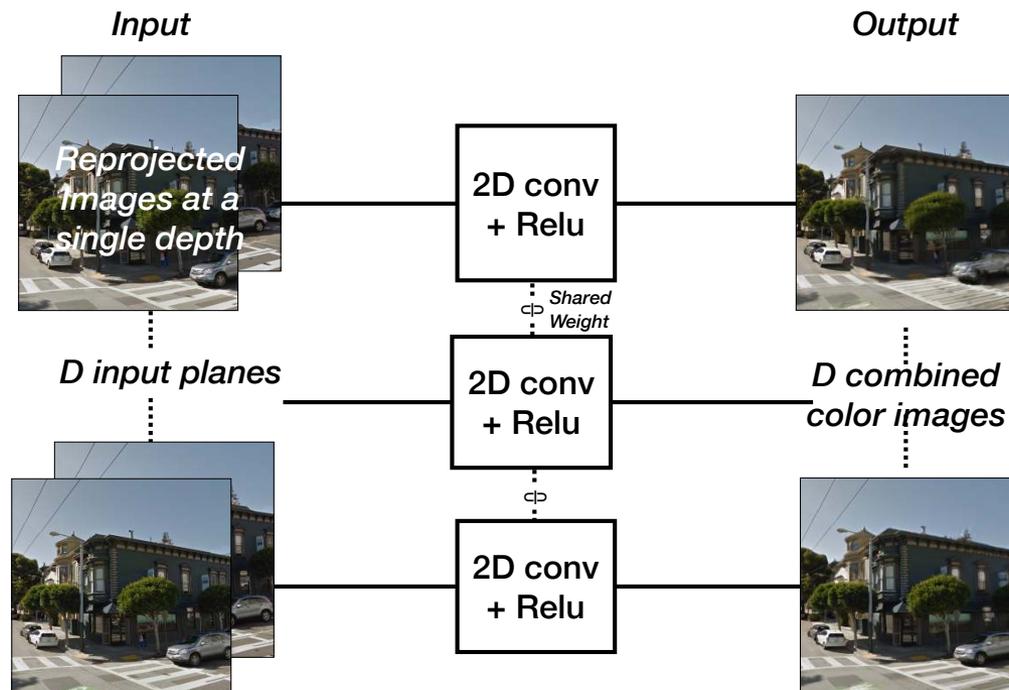
- 各デプス平面 P_z 上の各ピクセル $p_{i,j}$ を選択する確率 $S_{i,j,z}$ を生成
 - 陰にデプスを推定



“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

カラータワー

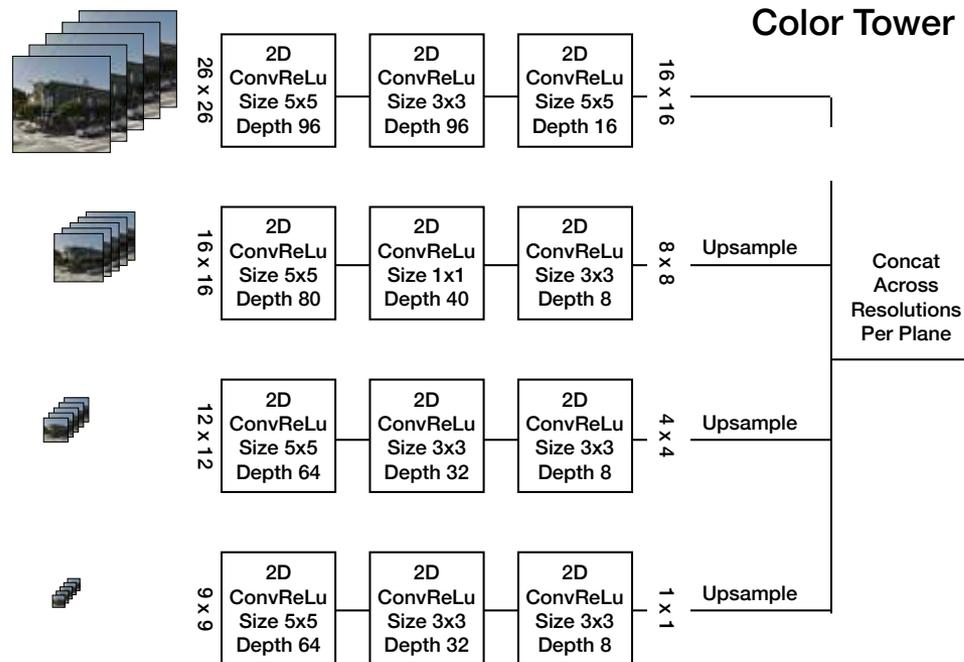
- 各デプス平面 P_z の各ピクセル $p_{i,j}$ の色 $c_{i,j,z}$ を生成



“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

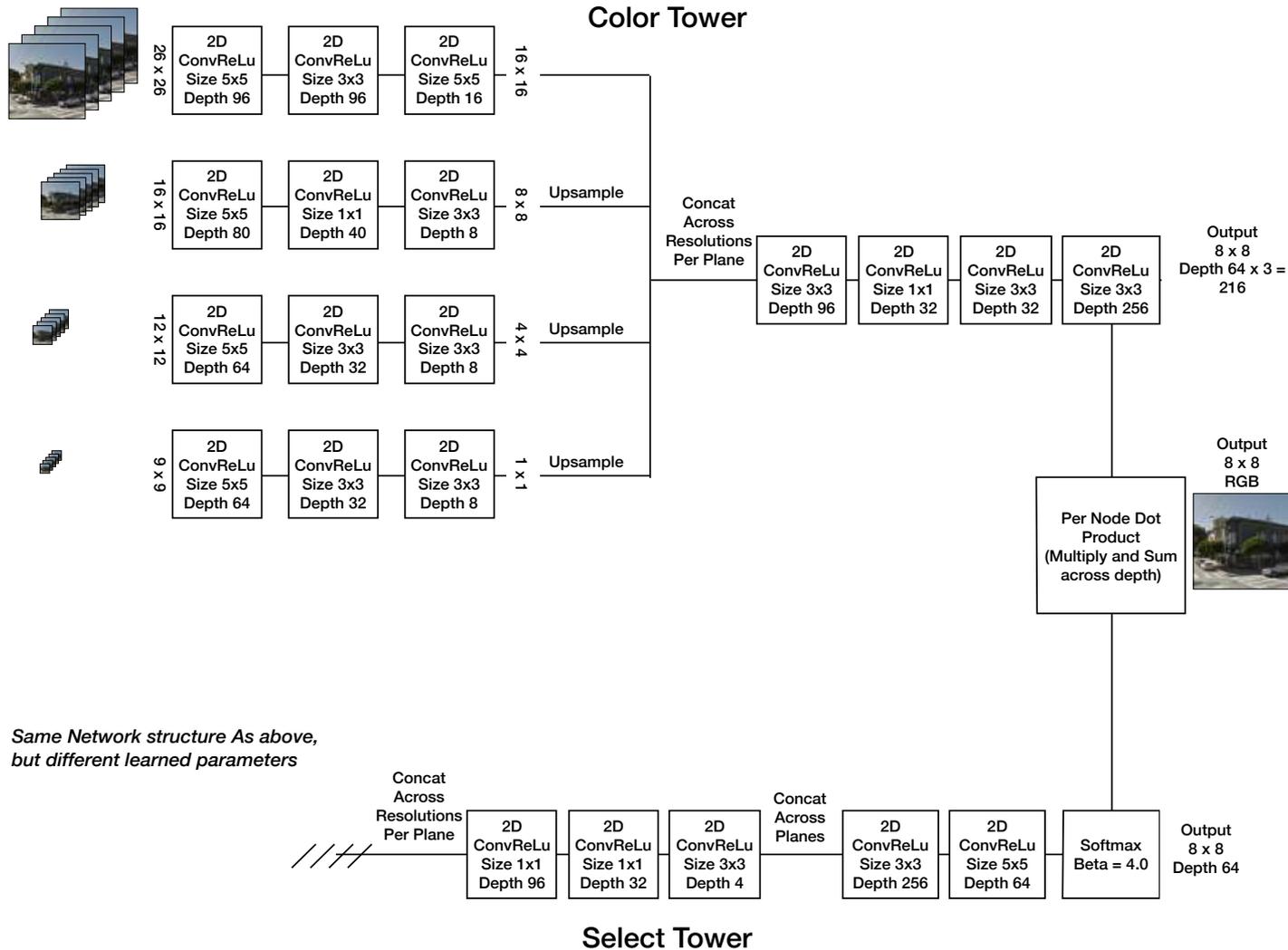
複数解像度のパッチ

- 4つの異なる解像度を入力
 - 各解像度ごとに独立に計算
 - 各解像度の畳み込み結果をアップサンプリングして統合し最終段に入力



“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

ネットワークの全体構造



“DeepStereo: Learning to Predict New Views from the World’s Imagery”, John Flynn, Ivan Neulander, James Philbin, Noah Snavely, arXiv:1506.06825, 2015

推定結果の一例



(a) Our result.



(b) Reference image.



(c) Crops of the five input panoramas.

課題

- 処理時間が長い
 - 12分 (512×512, マルチコアワークステーション)
- 入力画像の枚数とデプス解像度が固定
 - デプス解像度を96に設定
 - デプスの推定精度を制限
 - デプス解像度を上げるためにはより多くのデプスプレーンが必要
 - 予測と学習に要する時間が肥大化
 - 大きなRAM容量が必要

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - **視点補間**
- まとめ

目次

- 自己紹介
- 研究背景
 - 画像を用いた市街地の時空間モデリング
- 解析的手法の現状と課題
- CNNのシーンモデリングへの応用
 - パッチの類似度計算
 - ステレマッチング
 - 特徴点マッチング
 - 変化検出
 - 単一視点画像のデプス推定
 - 視点補間
- まとめ

まとめ

- CNNをシーンモデリングへ適用し大幅な精度向上
 - マルチスケールの画像を入力
 - 解析的なモデルの一部にCNNを利用
 - (例) ステレオマッチング
 - 解析的なモデルでネットワークに拘束条件を付与
 - (例) 視点補間
 - 関連のある全プロパティを入力
 - (例) 単一視点のデプス推定
 - デプス, 法線ベクトル, 物体カテゴリー
- 今後の課題
 - 学習結果を解析的手法にフィードバック
 - 3Dモデリング用大規模データセットの構築