VITAMIN-E:
VIsual Tracking And MappINg with Extremely Dense Feature Points

Masashi Yokozuka, Shuji Oishi, Thompson Simon, Atsuhiko Banno

Robot Innovation Research Center,
National Institute of Advanced Industrial Science and Technology (AIST)

CVPR2019

目次

1. 研究背景・目的
2. VITAMIN-E の発想に至るまで
3. 特徴点追跡
4. SLAM
5. 環境復元
6. 実験
7. まとめ・今後の課題
8. デモ

背景

- 現状の Visual SLAM の精度・ロバスト性は LiDAR SLAM の代わりにはなり得ない.
  - 移動体のナビゲーションには不十分な性能.
  - 精度・ロバスト性の追求が必要.

- 現状の Visual SLAM では LiDAR SLAM のように高密度・高精度な地図が作れない.
  - 移動体のナビゲーションに十分な環境復元ができない.
  - 現実の利用には実時間性が重要.

目的 1

Visual SLAM の 高精度化 及び 高ロバスト化

既存 SLAM ベンチマーク上で
State-of-the-art の SLAM 手法 を
精度・ロバスト性の両面で超える手法の開発

目的２

単眼カメラ による 実時間3次元環境復元

既存 SLAM ベンチマーク上で

State-of-the-art の SLAM 手法では未実現な

CPUのみによる実時間3次元環境復元の実現

Visual SLAM の State-of-the-art

- LSD-SLAM　(直接法，高密度)
  - J. Engel and D. Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *Proc. of European Conference on Computer Vision (ECCV)*, 2014.
- SVO　(直接法，低密度)
  - C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza. SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2):249–265, 2017.
- ORB-SLAM　(特徴点法，低密度)
  - R. Mur-Artal and J. D. Tardos. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.
- DSO　(直接法，中密度)
  - J. Engel, V. Koltun, and D. Cremers. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI)*, 2018.

State-of-the-art 分類

	直接法 (色の 誤差 最小化)	特徴点法 (射影 誤差 最小化)
Sparse (地図が粗)	**SVO** **DSO**	**ORB-SLAM**
Dense (地図が密)	**LSD-SLAM**	未提案

高密度な特徴点 ＝ **高精度＆ロバスト**

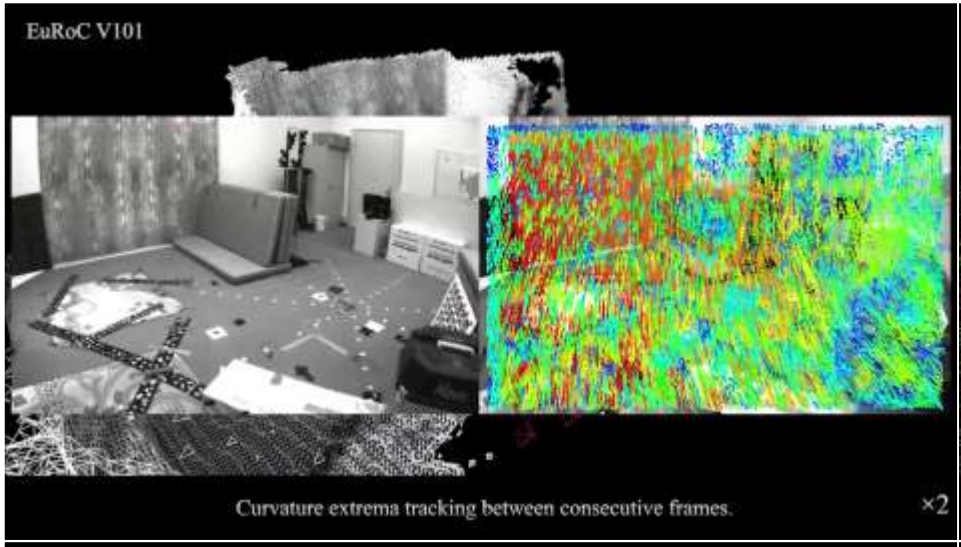
従来の特徴点法の問題点

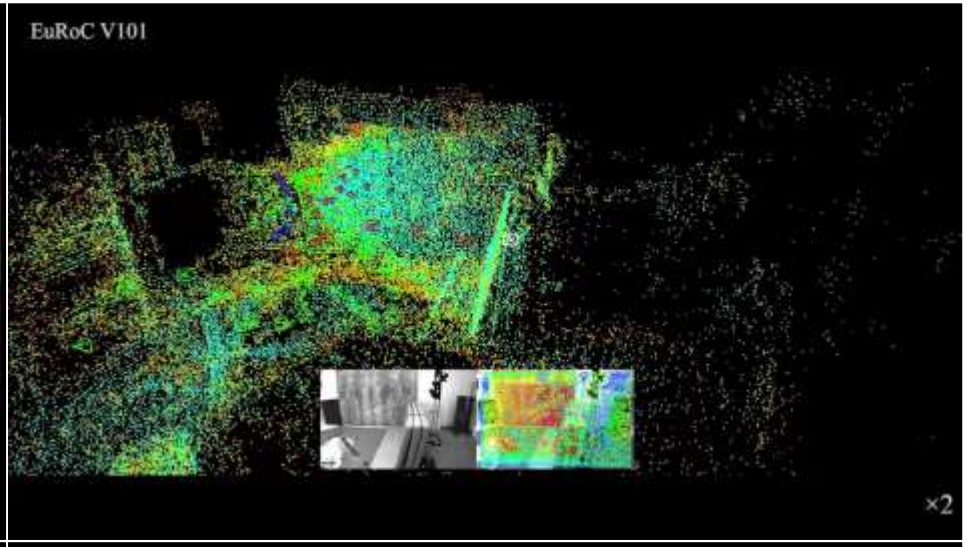
- 特徴点を密にマッチングできない.
  - 【計算コスト問題】特徴点抽出・特徴量記述が必要.
    - 密に点検出 → 大量の記述 → 計算コスト大.
  - 【ロバスト性問題】マッチングが安定しない.
    - 低テクスチャ → 特徴量記述が不安定 → 誤マッチング .

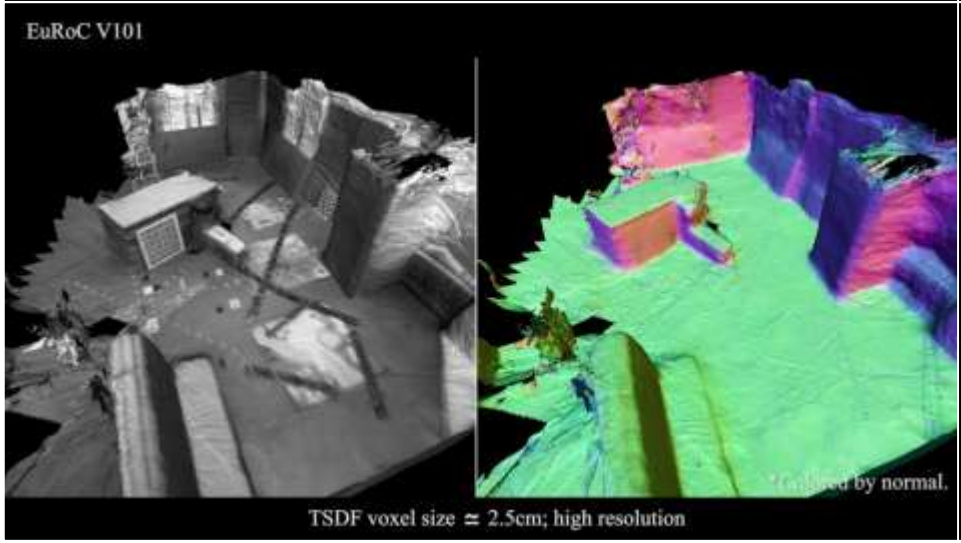
- 特徴量記述が根本原因.
  - 特徴量を利用しない → 計算コスト削減・不安定性を回避.

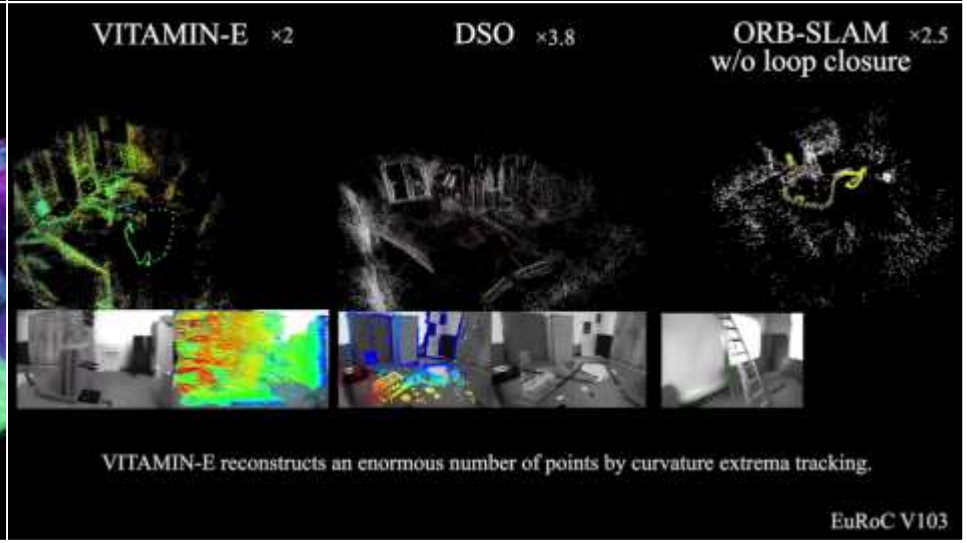
- 本研究では特徴量記述なしでトラッキングを行う.
  - 特徴点群の位置関係(全体的な形状)を重視.

提案手法概要



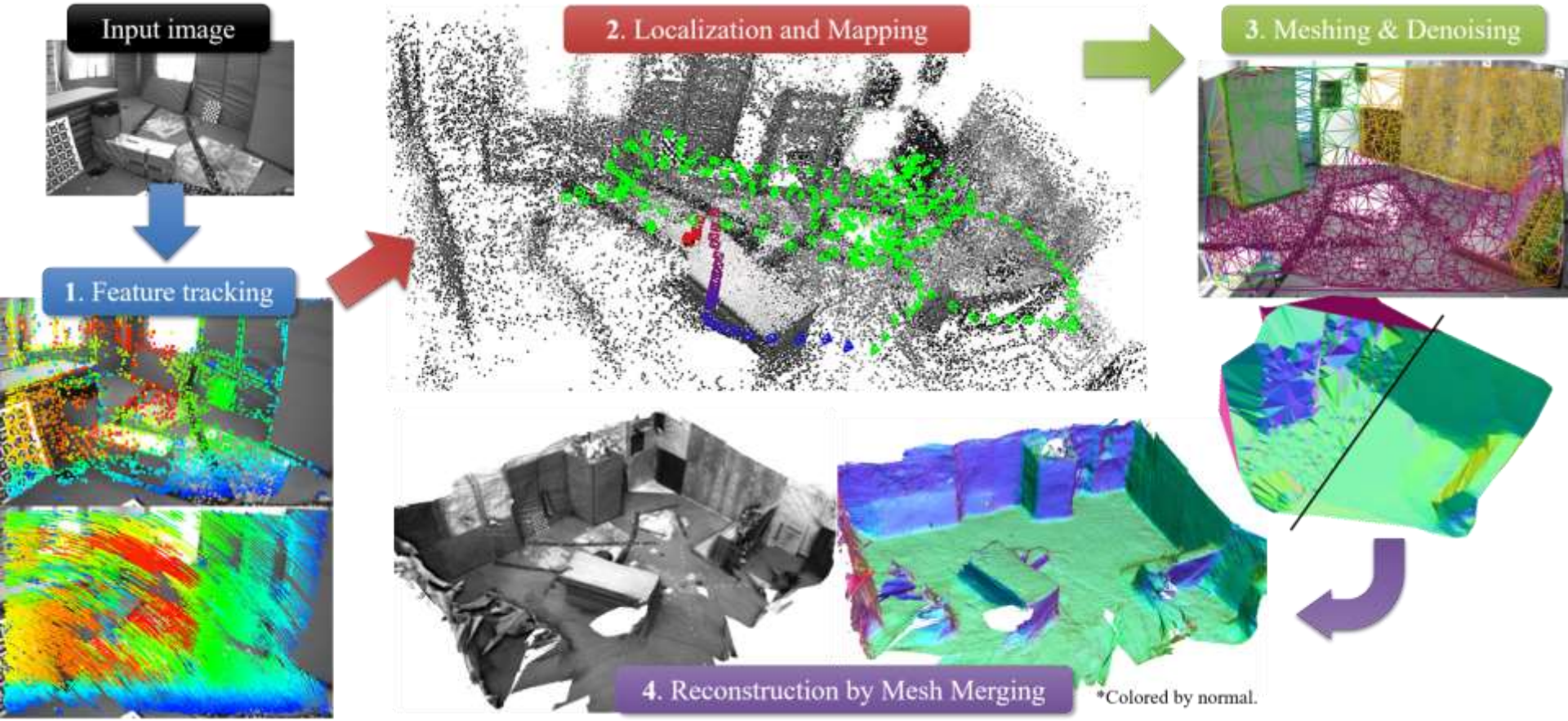






VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points
Masashi Yokozuka, Shuji Oishi, Thompson Simon, Atsuhiko Banno, **CVPR 2019**.

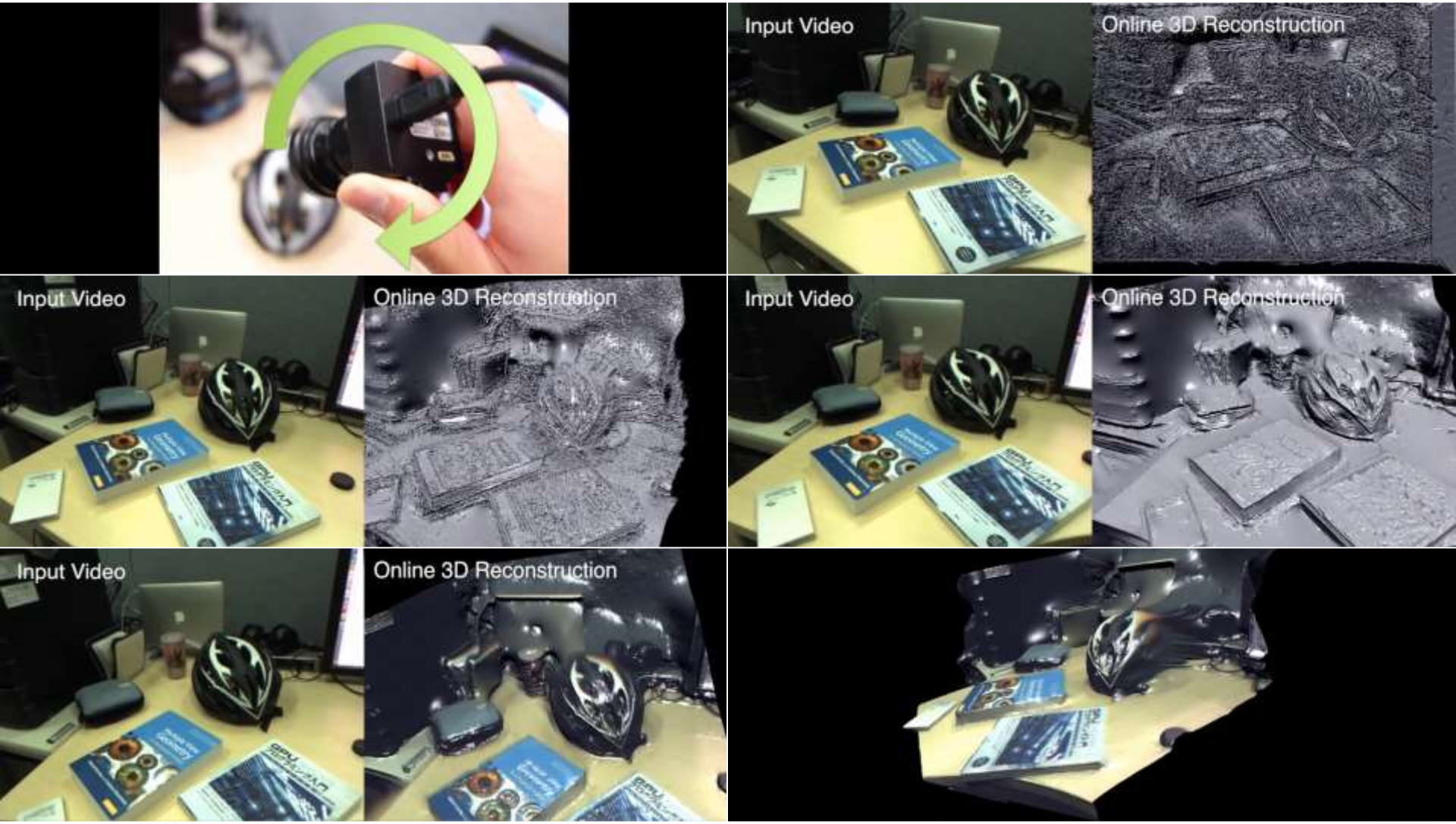
提案手法概要


Input image
1. Feature tracking
2. Localization and Mapping
3. Meshing & Denoising
4. Reconstruction by Mesh Merging
*Colored by normal.

2.VITAMIN-Eの発想に至るまで

VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points

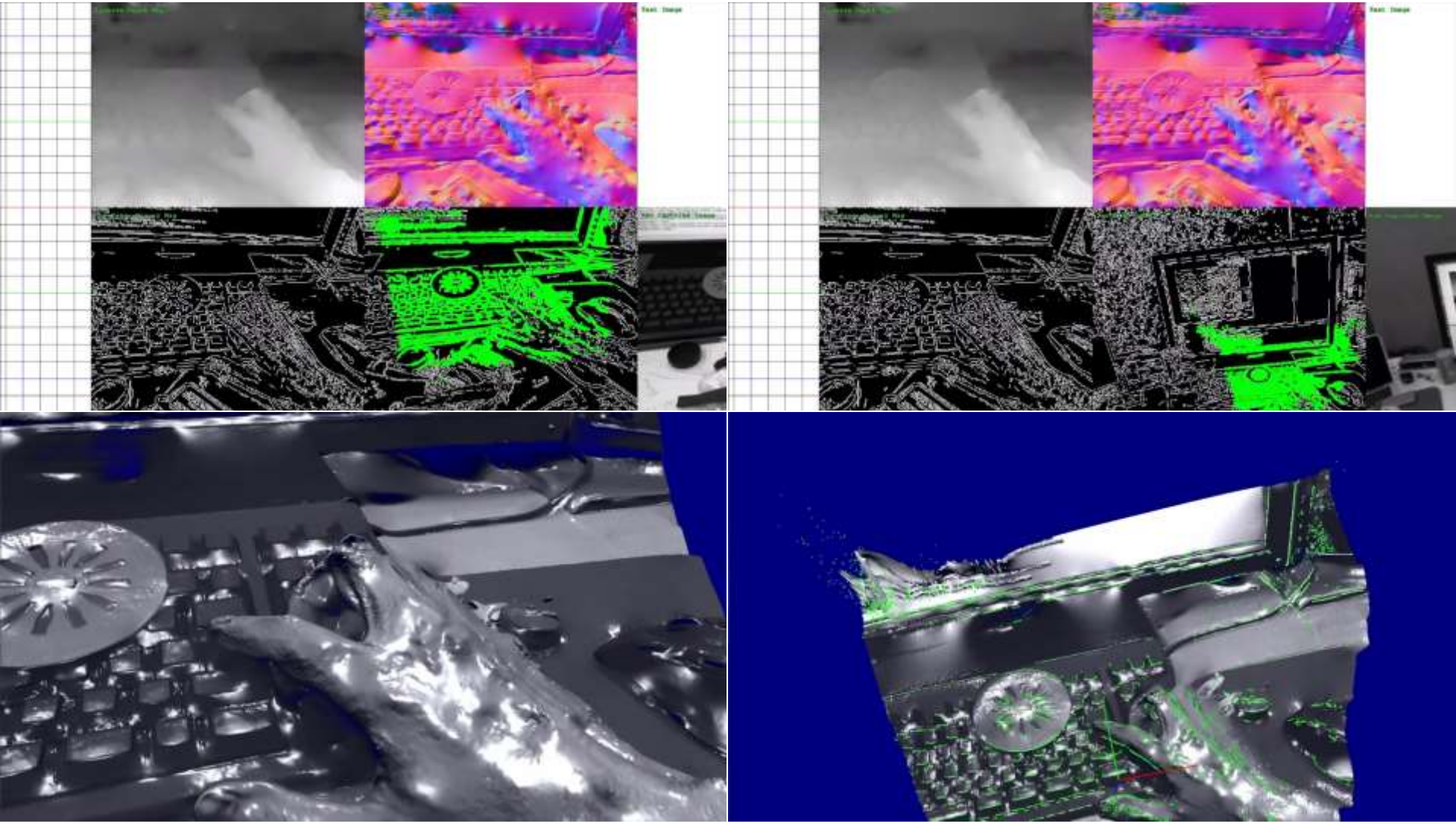
VITAMIN-E 以前の研究1:
DTAM追実装



DTAM原著：R.A.Newcombe, S.J.Lovegrove, A.J..Davison, "DTAM: Dense tracking and mapping in real-time", IEEE International Conference on Computer Vision (ICCV), 2011

VITAMIN-E 以前の研究2:

Visual SLAM from Binary Images



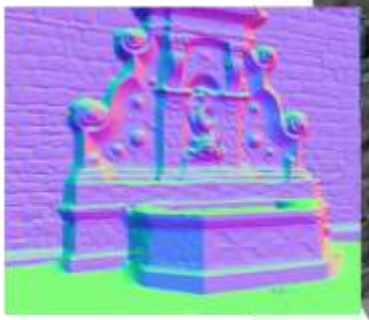
VITAMIN-E 以前の研究3:

Dense Reconstruction from Sparse Points

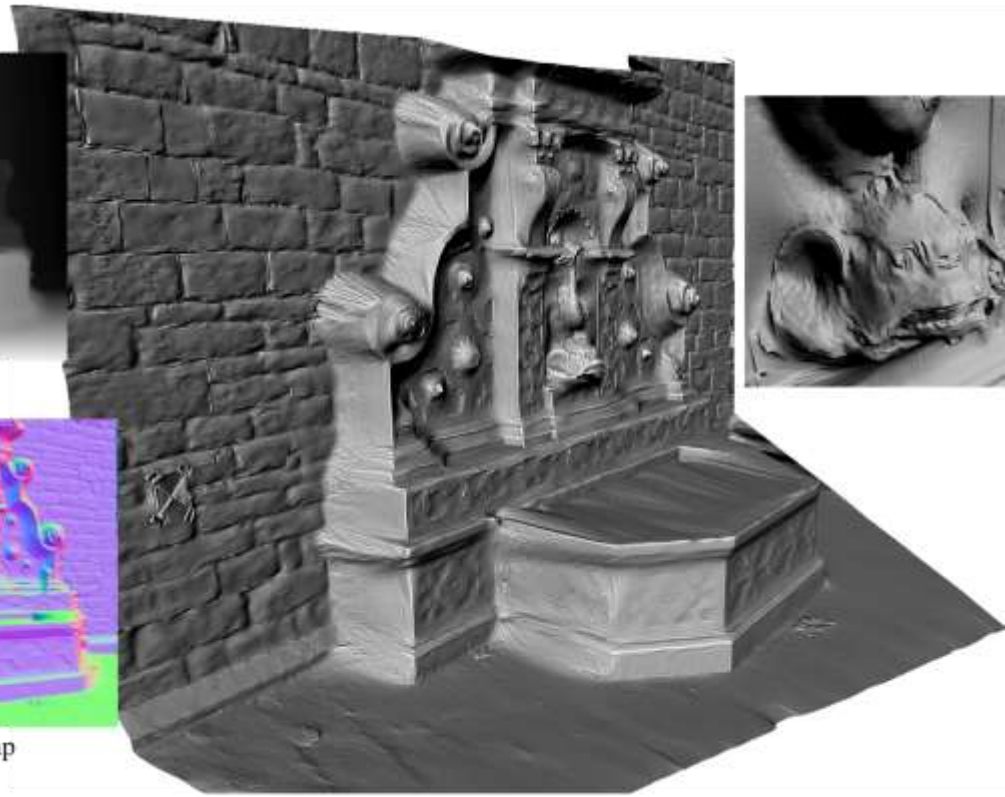
Stereo Measurement (Block Matching)

Initial Geometry

Depth Map



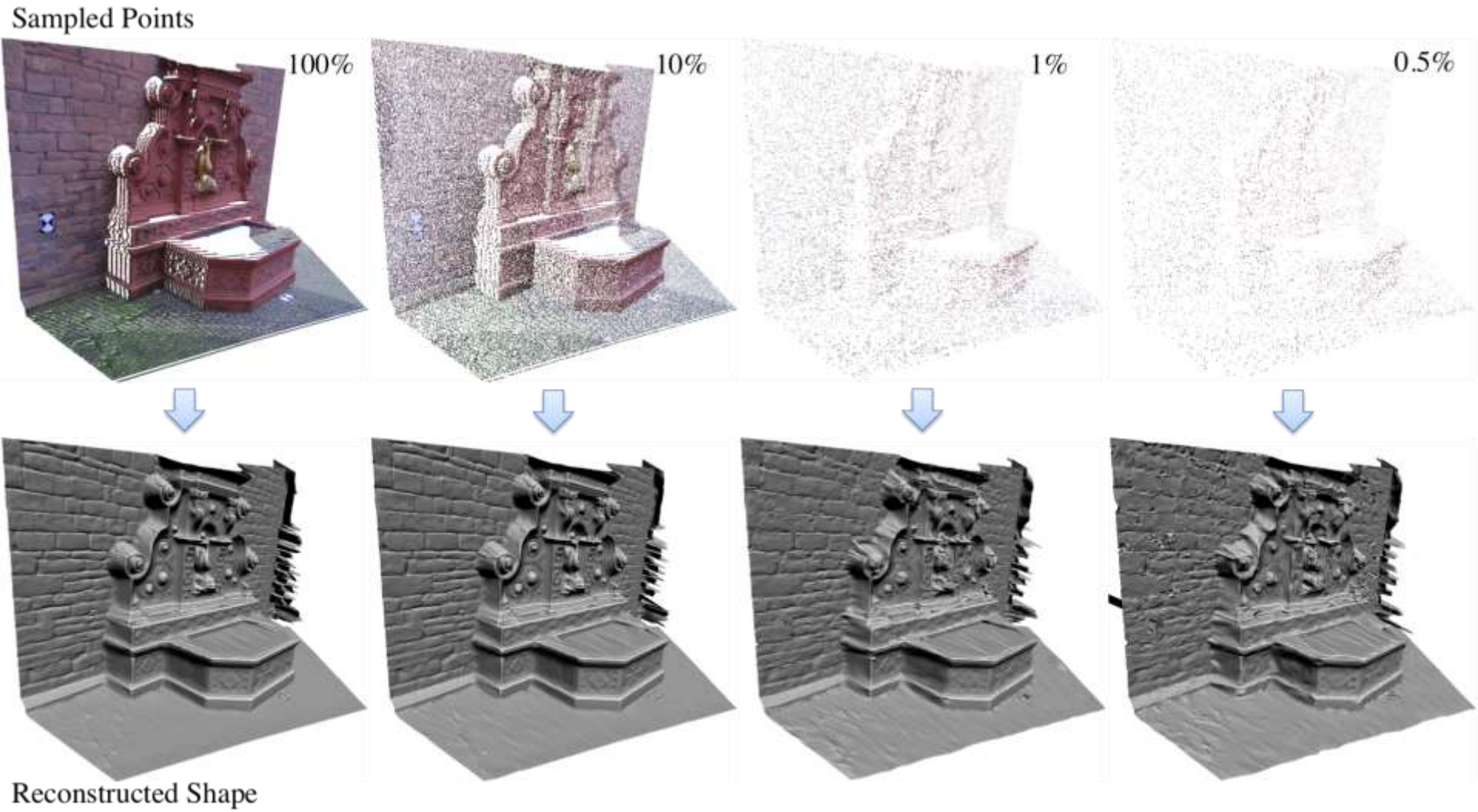
Normal Map



Result of applying our method (without texture, ***only lighting***)

VITAMIN-E 以前の研究3:

Dense Reconstruction from Sparse Points



VITAMIN-Eの着想

- DTAM再実装
  - Direct法は明度変化に弱い.
  - 特徴点法は比較的明度変化に強い → 密にできないか？
- Visual SLAM from Binary Images
  - 連続的に処理する場合 → 2値程度の情報量で充分.
    - ただし，Direct法でやるとトラッキングが外れやすい.
- Dense Reconstruction from Sparse Points
  - 密な形状復元は，疎な点の集合からできる.
  - 特徴点をある程度，密にとれば可能なはず.

3.特徴点追跡

VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points

特徴点追跡：概要

1. Coarse Matching

- 低解像度：画像全体を支配するフローを推定.
  - 特徴点マッチングからアフィン変換を推定.

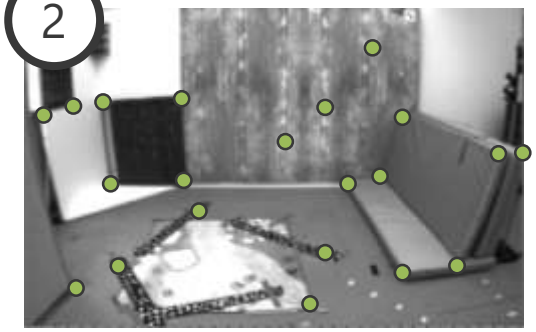
2. Fine Tracking

- 高解像度：密に特徴点トラッキング.
  - 特徴量を利用せずに特徴点の対応付け.

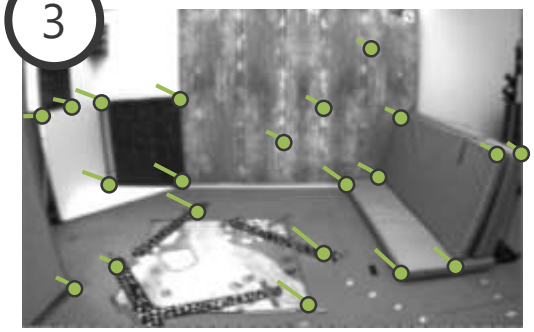
特徴点追跡：1. Coarse Matching

① 

低解像度化
(1/4～1/6程度)

② 

特徴点抽出&特徴量記述
(BRIEF を 32bit で利用)

③ 

特徴点対応付け
(BRIEFによるマッチング)

④ **支配的フロー推定**

$$y_i = Ax_i + b$$

y_i 現フレームの特徴点位置 (2次元ベクトル)
x_i 前フレームの特徴点位置 (2次元ベクトル)
A, b **支配的フロー** (2x2行列, 2次元ベクトル)

$$A, b = \underset{A,b}{\operatorname{argmin}} \rho(\|y_i - (Ax_i + b)\|_2)$$

ガウス・ニュートン法で最適化し支配的フローを求める

$$\rho(x) = \frac{x^2}{x^2 + \sigma^2}$$

M推定カーネル

特徴点追跡：2. Fine Tracking

① 曲率画像 κ の生成 (高解像度画像上で)

$$\kappa = f_y^2 f_{xx} - 2 f_x f_y f_{xy} + f_x^2 f_{yy}$$

f_x : X方向に 1 回 Sobel Filter を適用

f_{xy} : X方向に適用後，y方向に Sobel Filter を適用

② 支配的フローによる予測

$$\bar{x}_{t_1} = A x_{t_0} + b$$

$\bar{x}_{t_1}$ 現フレームの予測位置 (2次元ベクトル)

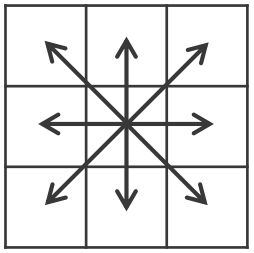
x_{t_0} 前フレームの特徴点位置 (2次元ベクトル)

A, b 支配的フロー (2x2行列, 2次元ベクトル)

③ 山登り法による曲率極大探索

$$x_{t_1} = \underset{x_{t_1}}{\operatorname{argmax}} \underbrace{\kappa\left(\boldsymbol{x}_{t_1}, t_1\right)}_{\text{曲率画像}} + \lambda \underbrace{w\left(\left\|\boldsymbol{x}_{t_1} - \bar{\boldsymbol{x}}_{t_1}\right\|_2\right)}_{\text{正則化項}}$$

初期値を予測位置 $\bar{x}_{t_1}$ から探索開始.
曲率画像上で上式が極大になるように，
山登り法(8方向)で探索.



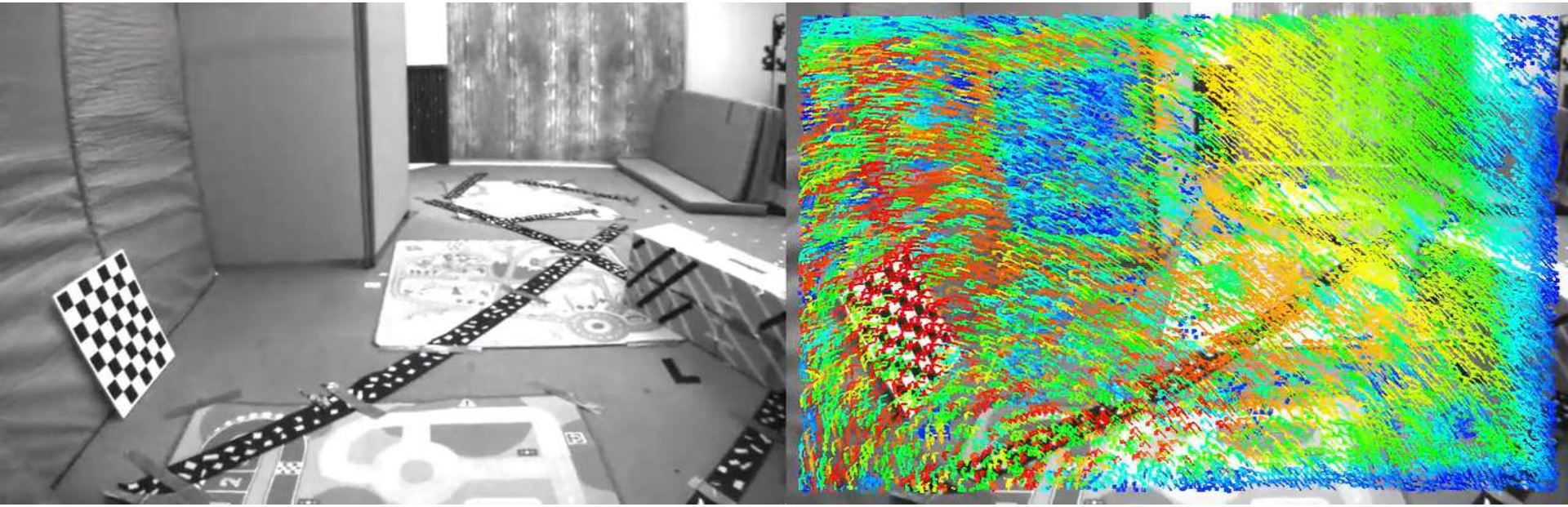
$$w(x) = 1 - \rho(x)$$

$$\rho(x) = \frac{x^2}{x^2 + \sigma^2}$$

特徴点追跡

1. Coarse Matching：フローの要約を求める.

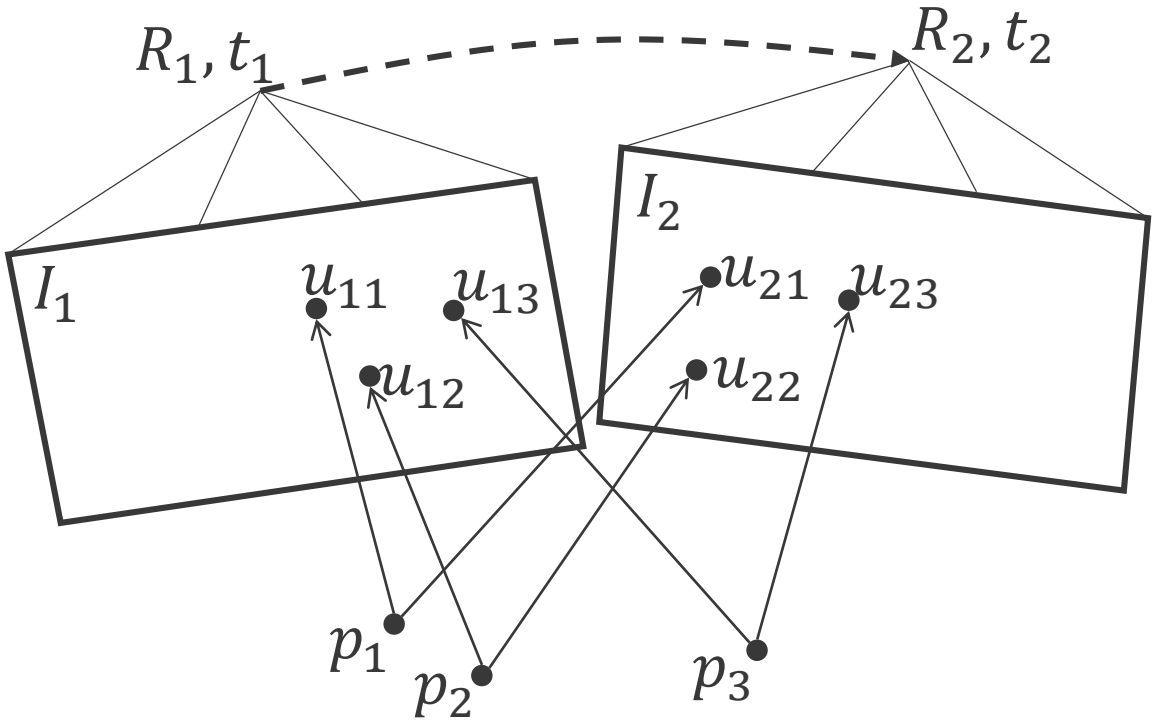
2. Fine Tracking：検出した特徴点全ての追跡.
   - 特徴量を利用せずに追跡 → 高速化の達成.



4.SLAM

VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points

SLAM：問題設定 (コスト関数)



$$\underset{R_j, t_j, p_i}{\operatorname{argmin}} E = \sum_{i}^{N} \sum_{j}^{M} \rho\left(\| \boldsymbol{u}_{ij} - \phi\left(R_j^T \left(\boldsymbol{p}_i - \boldsymbol{t}_j \right) \right) \|_2 \right)$$

$\phi(\boldsymbol{x})$: 射影関数(ピンホールカメラモデル)　　$\rho(x) = \dfrac{x^2}{x^2 + \sigma^2}$: M推定カーネル

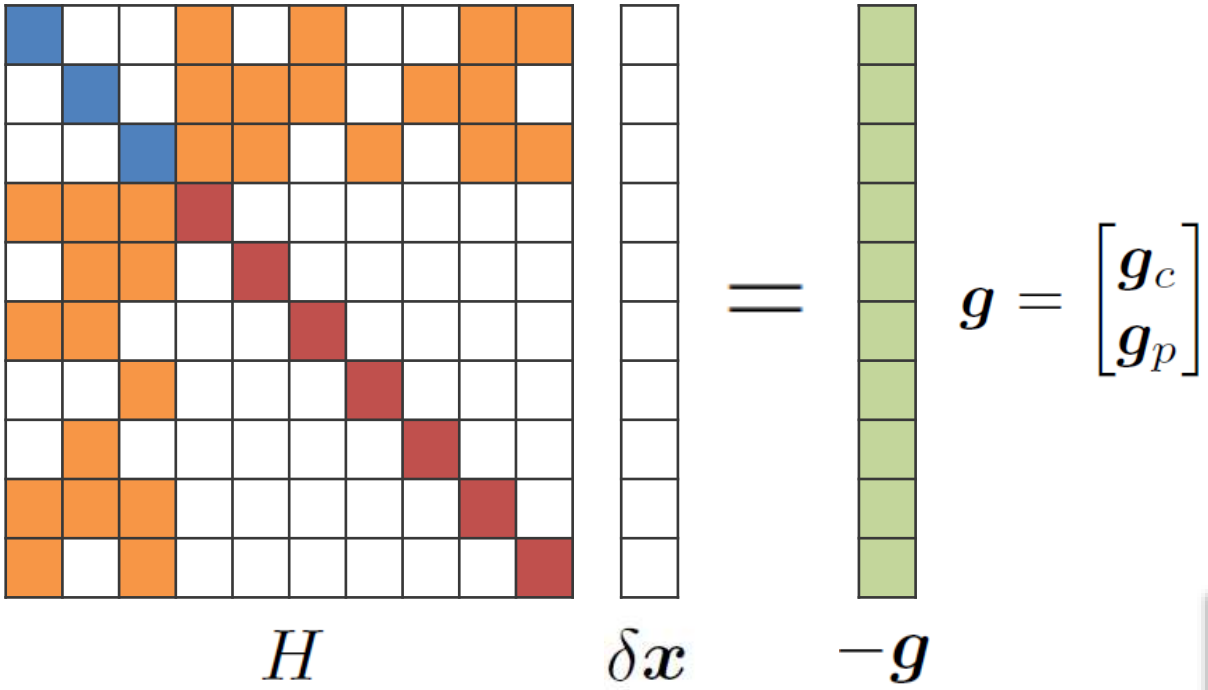
SLAM：ガウス・ニュートン法(従来法)

コスト関数 $E = \sum_{i}^{N} \sum_{j}^{M} \rho \left(\| \boldsymbol{u}_{ij} - \phi \left(R_j^T \left(\boldsymbol{p}_i - \boldsymbol{t}_j \right) \right) \|_2 \right)$

ヤコビアン $J = \frac{dE}{d\boldsymbol{x}}$　　ヘッセ行列 $H = J^T J$　　勾配 $\boldsymbol{g} = \boldsymbol{e}^T J$

$$H \delta \boldsymbol{x} = -\boldsymbol{g}, \quad \boldsymbol{x} = \boldsymbol{x} + \delta \boldsymbol{x}$$

$$H = \begin{bmatrix} H_{cc} & H_{cp} \\ H_{cp}^T & H_{pp} \end{bmatrix}$$



$$\boldsymbol{g} = \begin{bmatrix} \boldsymbol{g}_c \\ \boldsymbol{g}_p \end{bmatrix}$$

SLAM：部分空間ニュートン法(提案手法)

$$H\delta\boldsymbol{x} = -\boldsymbol{g}, \quad \boldsymbol{x} = \boldsymbol{x} + \delta\boldsymbol{x} \quad H = \begin{bmatrix} H_{cc} & H_{cp} \\ H_{cp}^T & H_{pp} \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} \boldsymbol{g}_c \\ \boldsymbol{g}_p \end{bmatrix}$$

$$H_{c_i c_i}\delta\boldsymbol{x}_{c_i} = -\left(\boldsymbol{g}_{c_i} + \sum_{l=1}^{i-1} H_{c_l c_i}\delta\boldsymbol{x}_{c_l} + \sum_{r=i+1}^{M} H_{c_i c_r}\delta\boldsymbol{x}_{c_r} + \sum_{j=1}^{N} H_{c_i p_j}\delta\boldsymbol{x}_{p_j}\right),$$

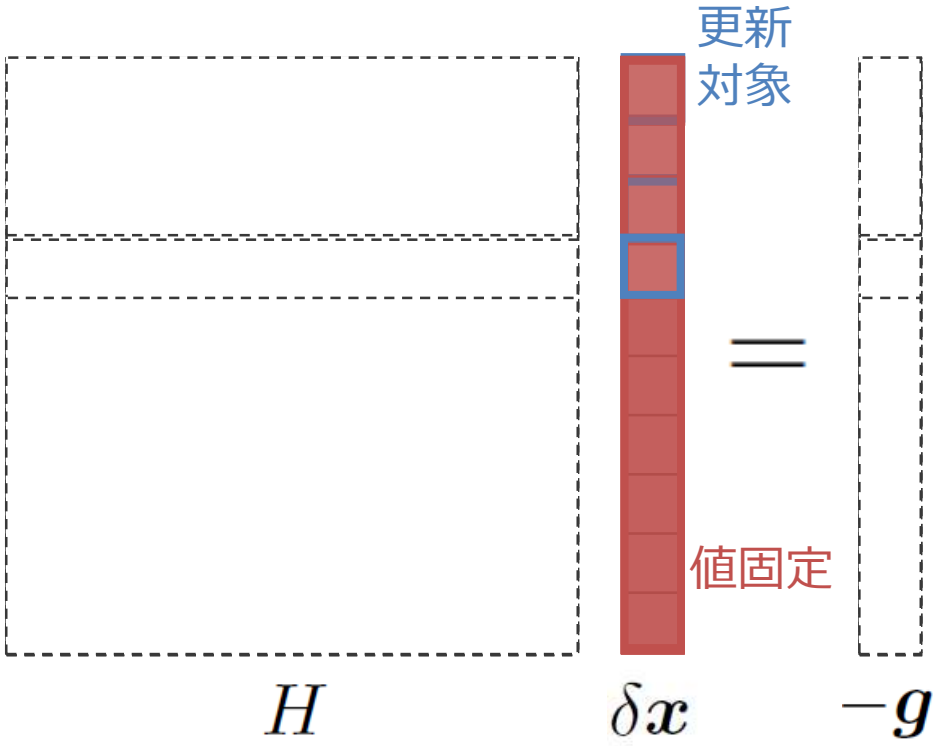
$$H_{p_j p_j}\delta\boldsymbol{x}_{p_j} = -\left(\boldsymbol{g}_{p_j} + \sum_{l=1}^{j-1} H_{p_l p_j}\delta\boldsymbol{x}_{p_l} + \sum_{r=j+1}^{N} H_{p_i p_r}\delta\boldsymbol{x}_{p_r} + \sum_{i=1}^{M} H_{c_i p_j}^T\delta\boldsymbol{x}_{c_i}\right).$$

- カメラ変数(6次元)，特徴点変数(3次元)を個別に最適化を行う.
  - 従来法　：全てを一括で最適化 → 巨大行列 → 計算コスト大
  - 提案手法：個々に最適化 → 小行列 → 計算コスト小

SLAM：部分空間ニュートン法(提案手法)

$$H_{c_ic_i}\delta\boldsymbol{x}_{c_i} = -\left(\boldsymbol{g}_{c_i} + \sum_{l=1}^{i-1} H_{c_lc_i}\delta\boldsymbol{x}_{c_l} + \sum_{r=i+1}^{M} H_{c_ic_r}\delta\boldsymbol{x}_{c_r} + \sum_{j=1}^{N} H_{c_ip_j}\delta\boldsymbol{x}_{p_j}\right),$$

$$H_{p_jp_j}\delta\boldsymbol{x}_{p_j} = -\left(\boldsymbol{g}_{p_j} + \sum_{l=1}^{j-1} H_{p_lp_j}\delta\boldsymbol{x}_{p_l} + \sum_{r=j+1}^{N} H_{p_ip_r}\delta\boldsymbol{x}_{p_r} + \sum_{i=1}^{M} H^T_{c_ip_j}\delta\boldsymbol{x}_{c_i}\right).$$



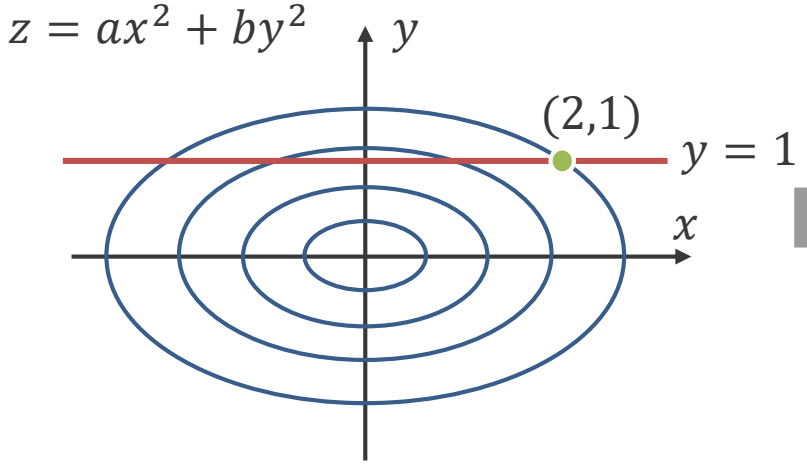
SLAM：提案手法の解釈

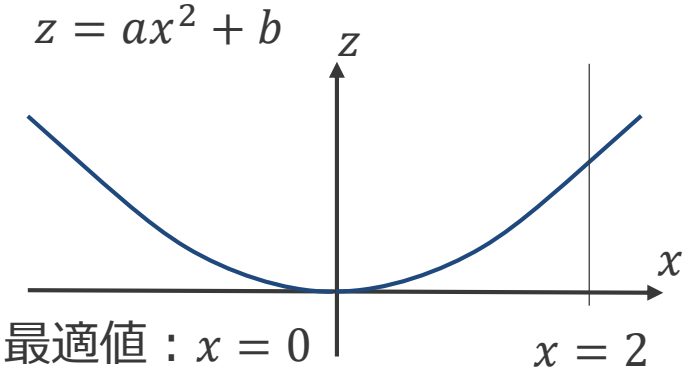
- 「多次元正規分布」としての解釈.
  1. 更新したい変数以外を固定 → 条件付き分布推定.
  2. 条件付き分布の期待値 → 変数更新.
  3. 繰り返し.

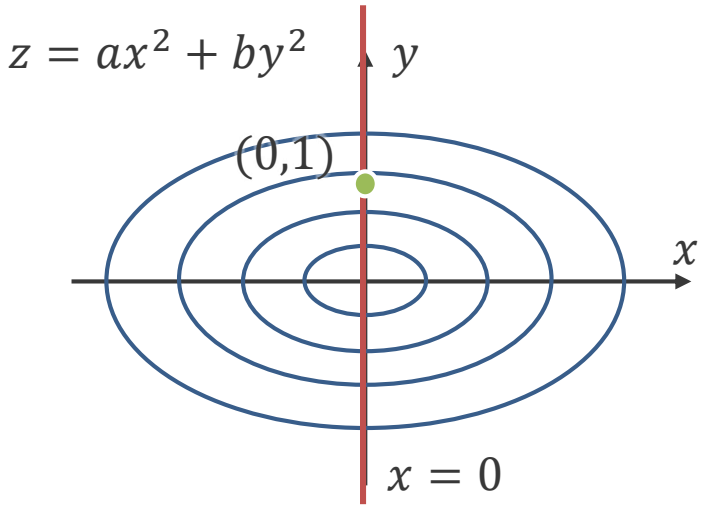
- 「Linear Solver」としての解釈.
  - Gauss-Seidel法(線形方程式の反復解法)の多変数同時更新.
  - 通常のGauss-Seidel法は1次元(＝1変数)更新.
  - 提案手法は多変数に拡張.

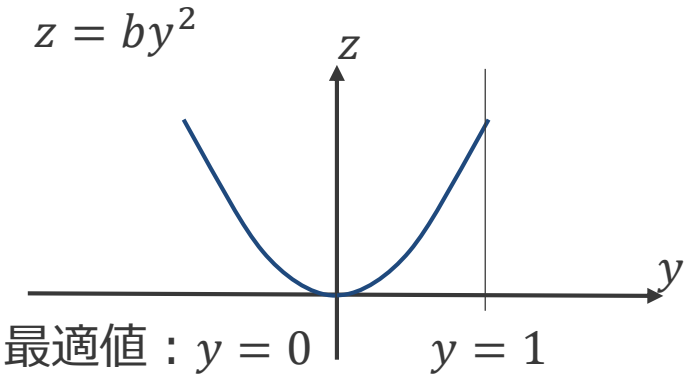
SLAM：提案手法の解釈

初期値：$x = 2, y = 1$









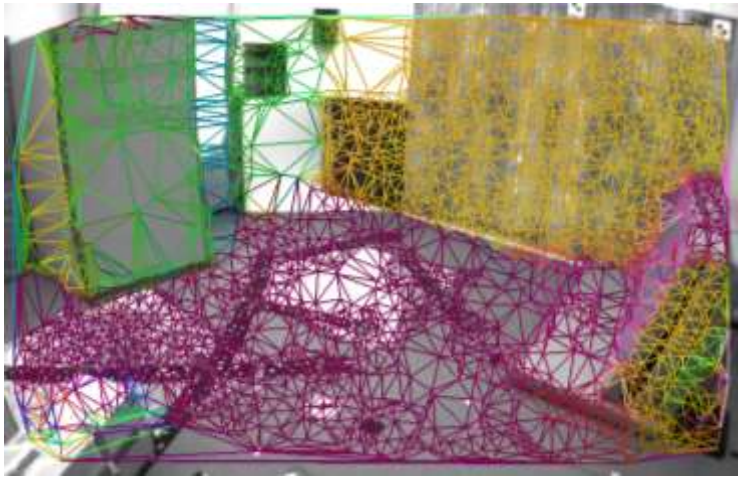
収束後：$x = 0, y = 0$

5.環境復元

VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points

メッシュ生成＆ノイズ除去

① メッシュ生成：ドロネー三角形分割

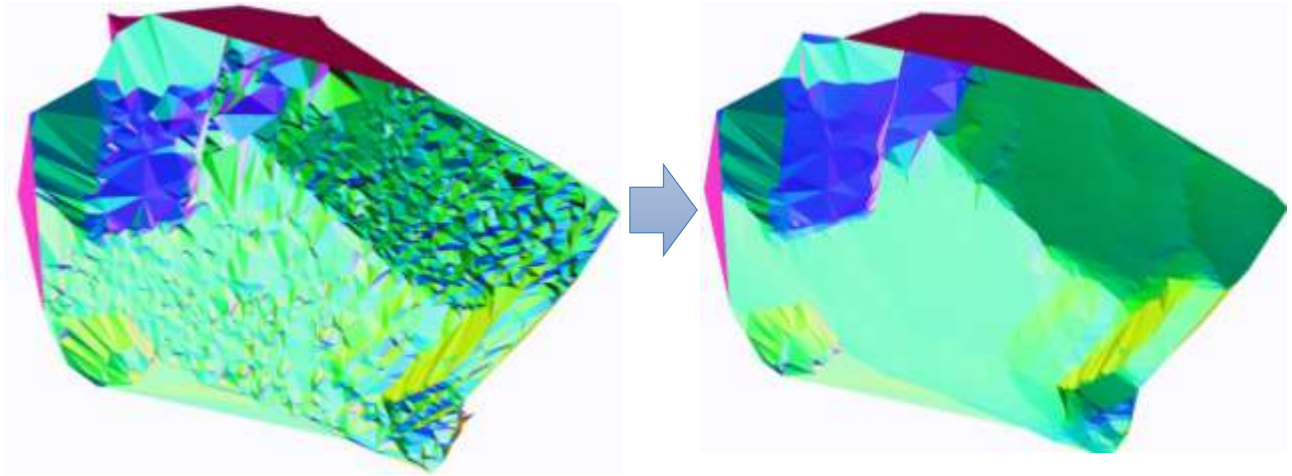


画像上の特徴点群に対して，
互いに交差しないように辺を構成.

デプス画像の代わりとして，
三角形メッシュを利用.

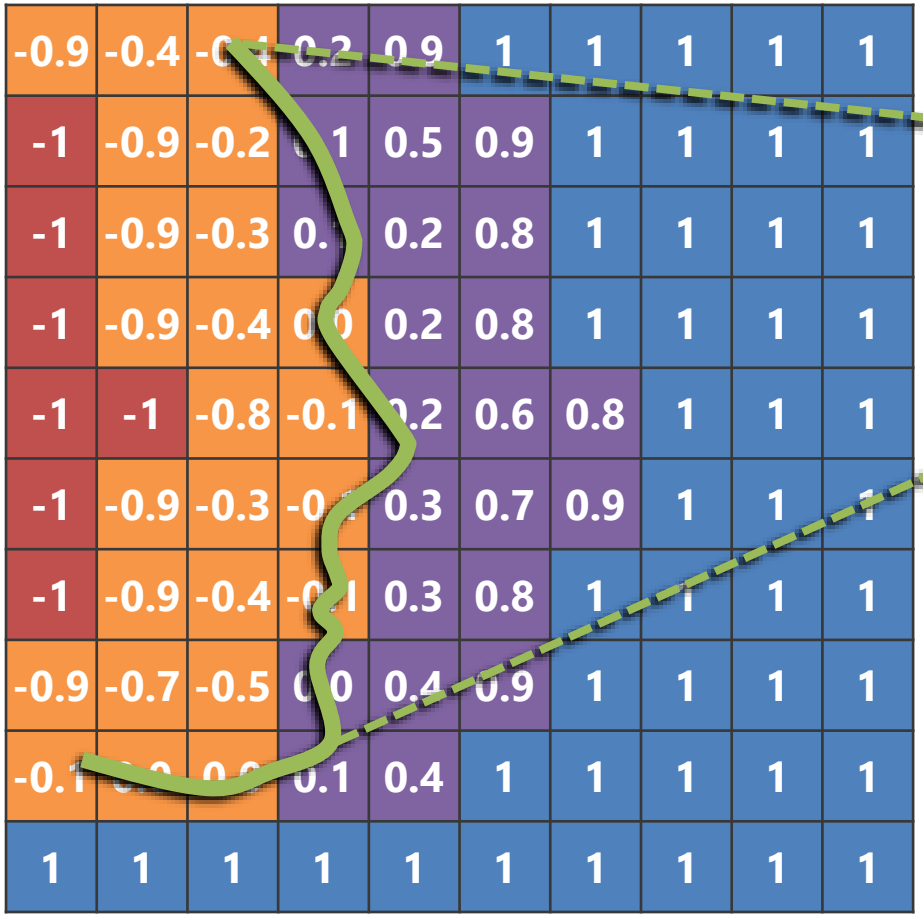
デプス画像のノイズ除去手法を
三角形メッシュに適用.

② ノイズ除去：Nonlocal Total Generalized Variation (NLTGV) 最小化



メッシュ統合：TSDFの利用

TSDF = Truncated Signed Distance Function



各ボクセルが保持する値

↓

中心から面までの距離

面復元＝距離が０になる場所を抽出

複数メッシュの統合 ＝ 距離の平均

6.実験

VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points

SLAM ベンチマーク

- EuRoCデータセット.
  - M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart. The EuRoC micro aerial vehicle datasets. *International Journal of Robotics Research*, 35(10):1157–1163, 2016.

評価方法

1. 推定軌跡の平均誤差 [cm]：精度評価
   - 単眼SLAMなので推定結果を定数倍して真値と比較.
2. 自己位置推定の成功率 [%]：ロバスト性評価
   - 画像シーケンスの内，自己位置推定が行えた割合.
3. 初期化のリトライ数 [times]
   - 単眼SLAMの初期地図作成は難しいので評価.

実験結果：MH01～MH05

Sequence name (no. of images)	Our method			DSO[6]			ORB-SLAM[21] w/o loop closure			LSD-SLAM[5] w/o loop closure		
MH01 easy (3682)	✓	12.9 ± 0.5	cm	✓	6.0 ± 0.8	cm	✓	**5.2 ± 1.1**	cm	×	(44.9 ± 7.2)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		**97.7 ± 1.6**	%		28.9 ± 23.6	%
		0 ± 0	times		0 ± 0	times		**19 ± 11**	times		–	times
MH02 easy (3040)	✓	8.8 ± 0.5	cm	✓	4.2 ± 0.2	cm	✓	**4.1 ± 0.4**	cm	×	(58.3 ± 6.9)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		**92.4 ± 1.1**	%		73.0 ± 1.5	%
		0 ± 0	times		0 ± 0	times		**56 ± 6**	times		–	times
MH03 medium (2700)	✓	**10.6 ± 1.3**	cm	✓	21.1 ± 0.9	cm	×	(4.5 ± 0.4)	cm	×	(266.2 ± 61.3)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		48.9 ± 0.8	%		28.4 ± 20.7	%
		0 ± 0	times		0 ± 0	times		0 ± 0	times		–	times
MH04 difficult (2033)	✓	**19.3 ± 1.6**	cm	✓	20.3 ± 1.0	cm	✓	33.6 ± 9.4	cm	×	(136.4 ± 114.3)	cm
		100.0 ± 0.0	%		95.7 ± 0.0	%		95.2 ± 0.8	%		27.2 ± 7.0	%
		0 ± 0	times		5 ± 0	times		6 ± 1	times		–	times
MH05 difficult (2273)	✓	14.7 ± 1.1	cm	✓	**10.2 ± 0.6**	cm	✓	14.9 ± 4.6	cm	×	(27.4 ± 16.4)	cm
		100.0 ± 0.0	%		**95.5 ± 0.0**	%		90.0 ± 4.0	%		22.7 ± 0.5	%
		0 ± 0	times		**2 ± 0**	times		18 ± 5	times		–	times

- 各シーケンスで5回実験 → 平均＆標準偏差.
- ✓ or × : 自己位置推定 の 成功 or 失敗 (成功率90%以上で ✓).

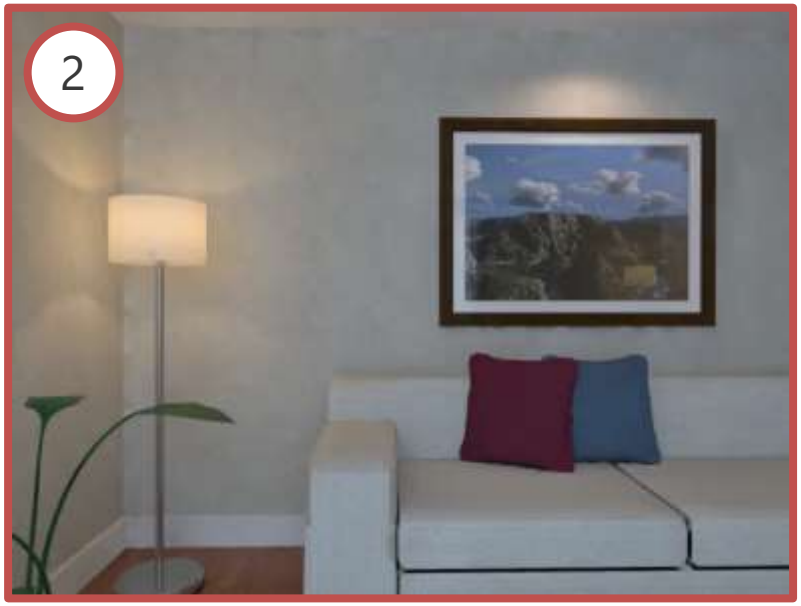
実験結果：V101～V203

Sequence name (no. of images)	Our method			DSO[6]			ORB-SLAM[21] w/o loop closure			LSD-SLAM[5] w/o loop closure		
V101 easy (2911)	✓	9.7 ± 0.2	cm	✓	13.4 ± 5.8	cm	✓	**8.8 ± 0.1**	cm	×	(20.0 ± 22.8)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		**96.6 ± 0.0**	%		11.6 ± 11.2	%
		0 ± 0	times		0 ± 0	times		**1 ± 0**	times		–	times
V102 medium (1710)	✓	**9.3 ± 0.6**	cm	✓	53.0 ± 5.5	cm	×	(14.5 ± 11.7)	cm	×	(67.0 ± 14.0)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		52.0 ± 3.3	%		15.2 ± 0.1	%
		0 ± 0	times		0 ± 0	times		17 ± 4	times		–	times
V103 difficult (2149)	✓	**11.3 ± 0.5**	cm	✓	85.0 ± 36.4	cm	×	(37.2 ± 20.7)	cm	×	(29.3 ± 2.0)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		65.5 ± 8.8	%		11.0 ± 0.1	%
		0 ± 0	times		0 ± 0	times		56 ± 26	times		–	times
V201 easy (2280)	✓	7.5 ± 0.4	cm	✓	7.6 ± 0.5	cm	✓	**6.0 ± 0.1**	cm	×	(131.3 ± 20.4)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		**95.2 ± 0.0**	%		74.1 ± 8.9	%
		0 ± 0	times		0 ± 0	times		**0 ± 0**	times		–	times
V202 medium (2348)	✓	**8.6 ± 0.7**	cm	✓	11.8 ± 1.4	cm	✓	12.3 ± 2.7	cm	×	(42.1 ± 9.2)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		99.5 ± 1.2	%		11.3 ± 0.2	%
		0 ± 0	times		0 ± 0	times		0 ± 0	times		–	times
V203 difficult (1922)	✓	**140.0 ± 5.2**	cm	✓	147.5 ± 6.6	cm	×	(104.3 ± 64.0)	cm	×	(17.7 ± 1.6)	cm
		100.0 ± 0.0	%		100.0 ± 0.0	%		16.8 ± 15.9	%		11.9 ± 0.2	%
		0 ± 0	times		0 ± 0	times		233 ± 123	times		–	times

Num. of Wins **Our method: 6** DSO: 1 ORB-SLAM: 4 LSD-SLAM: 0

実験：他データセット(TUM-RGBD ICL-NUIM)

データセット名	① TUM-RGBD
② ICL-NUIM Living Room	③ ICL-NUIM Office Room



実験：他データセット(TUM-RGBD ICL-NUIM)

Table 1. Trajectory errors [cm] in the additional datasets.

		Our method	DSO	SVO	ORB-SLAM (with loop closure)	ORB-SLAM (w/o loop closure)	LSD-SLAM (with loop closure)	LSD-SLAM (w/o loop closure)
TUM RGB-D	fr2_desk	1.7	-	6.7	**0.9**	-	4.5	-
	fr2_xyz	0.4	-	0.8	**0.3**	-	1.5	-
ICL-NUIM	Living Room 0	**0.9**	1.0	2.0	-	1.0	-	12.0
	Living Room 1	11.0	**2.0**	7.0	-	2.0	-	5.0
	Living Room 2	3.1	6.0	10.0	-	7.0	-	**3.0**
	Living Room 3	**2.4**	3.0	7.0	-	3.0	-	12.0
ICL-NUIM	Office Room 0	31.6	21.0	34.0	-	**20.0**	-	26.0
	Office Room 1	40.1	83.0	28.0	-	89.0	-	**8.0**
	Office Room 2	**3.8**	36.0	14.0	-	30.0	-	31.0
	Office Room 3	**5.5**	64.0	8.0	-	64.0	-	56.0

考察

- EuRoC：高速運動＆照明変化大.
  - 従来手法に比べて高精度・高ロバスト
    - 特に難しいシーケンスに対して.
- TUM-RGBD：容易なケース.
  - ループクロージングなしでも十分な精度
- ICL-NUIM：低テクスチャ環境.
  - 従来法に比べれば改善してはいる.
  - 失敗しているケースは概ね以下になる.
    - その場回転のみの運動： 単眼では原理的に不可能.
    - 直線が一つしかない環境：エッジベースSLAMでも不可能.

7.まとめ・今後の課題

VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points

まとめ

- 提案手法は，精度・ロバスト性について従来手法を超えることができた.
  - 既存ベンチマークで公平に評価して実証した.

- GPUを用いずに実時間3次元環境復元を行えた.
  - ドロネー三角形分割によるメッシュ生成.
  - NLTGV最小化によるメッシュノイズ除去.
  - TSDFによるメッシュ統合.

今後の課題

- 精度・密度の追求 → 継続.
  - 特徴点抽出の改善：曲率ではなく，機械学習の利用.
  - Loop Closureの実装：ICPベースの実装(BoFは非利用).
- ロバスト性の追求 → IMUとのセンサ統合.
  - EKF(Loose Coupling)，Graph Opt.(Tight Coupling)でない，Semi-tight CouplingなIMU統合.
- 形状復元 → 四面体空間分割によるDirect Meshing.
  - TSDFはメモリ効率が悪い：空間を均一なGrid分割.
  - ３次元ドロネー三角形分割