INSTITUTO
DE INFORMÁTICA
UFRGS

c-M2DP: A FAST POINT CLOUD DESCRIPTOR WITH COLOR INFORMATION TO PERFORM LOOP CLOSURE DETECTION

Leonardo Perdomo
Edson Prestes e Silva Junior

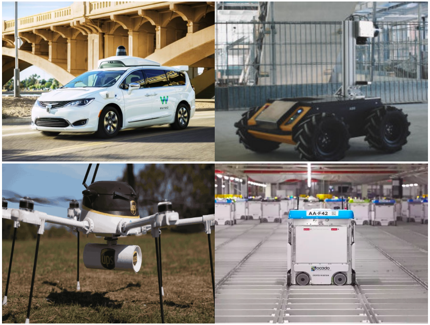
November, 2019

OUTLINE

1. Introduction
2. Proposal
3. Related Work
4. c-M2DP
5. Loop Closure Detection
6. Dataset Sequences
7. Experiments
8. Results

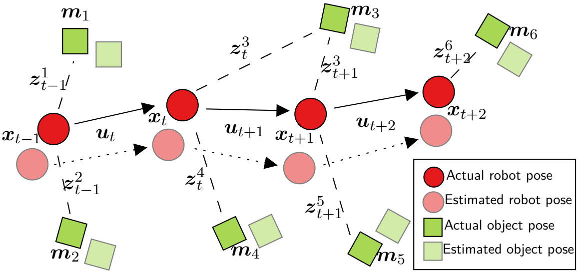
AUTONOMOUS ROBOTS

- Several examples of recent autonomous robots applications;
- Perform tasks in distinct real-world environments:
  - Simultaneous Localization and Mapping (SLAM).



Self-driving, inspection, delivery, retail, among others.

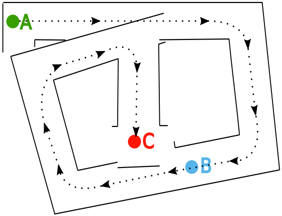
- Estimates its own pose and incrementally builds an map, using sensor measurements and odometry while moving in the environment.



Example of the SLAM problem (DURRANT-WHYTE; BAILEY, 2006)

LOOP CLOSURE DETECTION

- Estimation errors accumulated during movement increase the pose uncertainty and drifting;
- Recognize previously visited places, reducing uncertainty and updating the map.



Map built using only odometry (CADENA et al., 2016).

HOW TO DETECT LOOP CLOSURES?

- Correspondence search using appearance signatures from places visited during the trajectory;
- Signatures can be built using shape, color and other data available from sensors:
  - Several visual-based methods[1] for cameras developed over the past years;
  - 3D LIDAR-based methods are considered less mature[2].

[1] (CUMMINS; NEWMAN, 2008; MILFORD; WYETH, 2012; LOWRY et al., 2016)
[2] (HE; WANG; ZHANG, 2016; DUBÉ et al., 2017)

POINT CLOUD DESCRIPTORS

- Typically, loop closure detection with 3D LIDARs employ point cloud matching approaches[3] using feature descriptors:
  - Global descriptors represent the entire cloud geometry into a single descriptor;
  - Local descriptors compute the characteristics around multiple keypoints:
    - Quality and performance issues with keypoint detection techniques[4].

[3] (BOSSE; ZLOT, 2013; CIESLEWSKI et al., 2016; HE; WANG; ZHANG, 2016)
[4] (DUBÉ et al., 2017)

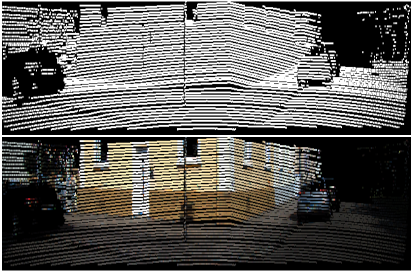
MULTIVIEW 2D PROJECTION (M2DP)

- Recently, the M2DP[5] descriptor presented significant results when applied to loop closure detection:
  - Outperforms other descriptors, such as SHOT[6]'s global variant;
  - Avoids using normals, which can be costly to estimate for large clouds;
  - Spatial density distributions are computed from multiple 2D projections of a point cloud;
- We noticed that it could be extended to compute additional information from each projection.

[5] (HE; WANG; ZHANG, 2016)
[6] (TOMBARI; SALTI; STEFANO, 2011)

COMBINING COLOR AND SHAPE

- Alongside 3D spatial data, color can provide more descriptive scenes:
  - Object recognition works[7] report increase in descriptiveness;
  - Insufficiently investigated approach for loop closure detection.



Colored point cloud generated using LIDAR and camera.

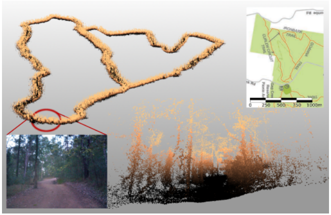
[7] (TOMBARI; SALTI; STEFANO, 2011; FENG; LIU; LIAO, 2015; LOGOGLU; KALKAN; TEMIZEL, 2016)

Color M2DP (c-M2DP)

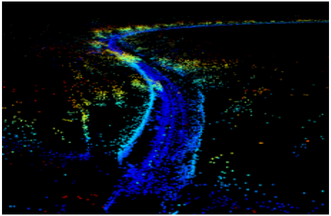
- A global descriptor comprising of color and shape data computed from the point cloud;
- An improved loop closure detection, using the c-M2DP descriptor on point cloud sequences generated through camera-LIDAR fusion, or stereo depth estimation.

LOOP CLOSURE DETECTION USING POINT CLOUD DESCRIPTORS

- Both avoid using normals, measuring point distributions from the point clouds;
- However, both are local descriptors, and compute only shape data from the cloud.



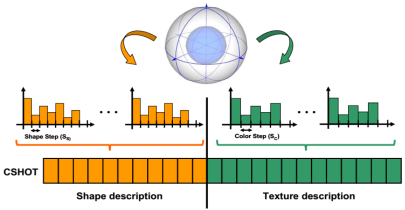
3D Gestalt (BOSSE; ZLOT, 2013).



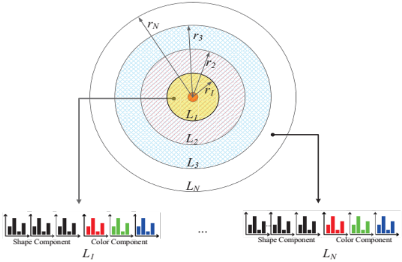
NBLD (CIESLEWSKI et al., 2016).

COLOR AND SHAPE POINT CLOUD DESCRIPTORS

- Local descriptors designed for object recognition applications:
  - Histograms of normals and color characteristics are computed from a local support split in concentric spheres;
- Additionally, CSHOT have a global variant that uses the whole cloud as support.



CSHOT (TOMBARI; SALTI; STEFANO, 2011).



CoSPAIR (LOGOGLU; KALKAN; TEMIZEL, 2016).

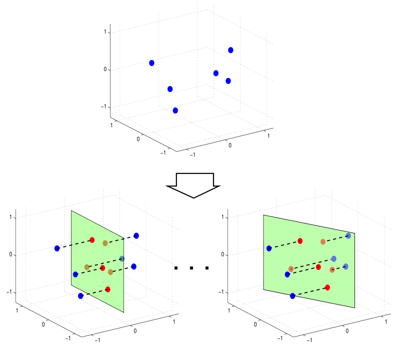
EXTENDING M2DP

- Our proposal takes advantage of M2DP's existing structure:
  - Reference frame, shape signatures, and dimensionality reduction steps remains unchanged;
  - Color signatures are computed alongside shape, from the multiple 2D projections;
  - Increased length of signature matrix and descriptor vector.

REFERENCE FRAME

- Point cloud $\boldsymbol{P}$ centroid is computed and used as the reference frame origin;
- PCA is performed on $\boldsymbol{P}$, with the 1st and 2nd PCs defined as the x-axis and y-axis, respectively.

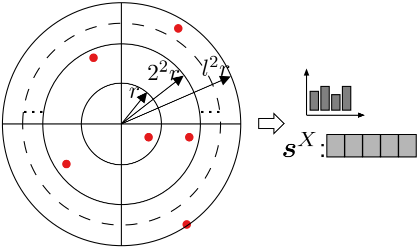
- Distinct 2D planes are generated by varying $[\theta, \phi]$;
- $\boldsymbol{P}$ is projected onto each 2D plane, in order to compute shape and color signatures from each 2D projection.



Projecting $\boldsymbol{P}$ on multiple 2D planes.

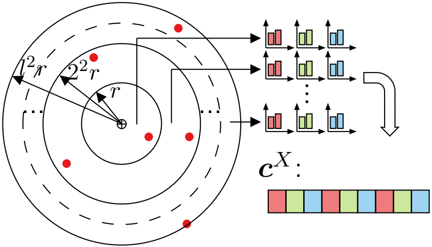
SHAPE SIGNATURE

- Each plane is split into l concentric circles;
- Each concentric circle is divided in h shape bins, indexed by the x-axis;
- Shape signature $\boldsymbol{s}^X$ is computed by counting the points within each bin.



Computing the shape signature $\boldsymbol{s}^X$.

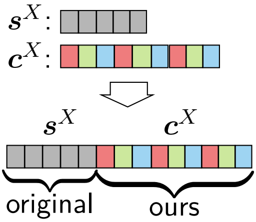
- We build color histograms, in which each channel is divided in g bins;
- Histograms are computed for every concentric circle, and are concatenated into a single color signature vector $\boldsymbol{c}^X$;



Computing the color signature $\boldsymbol{c}^X$.

SIGNATURE VECTOR

- Both $\boldsymbol{s}^X$ and $\boldsymbol{c}^X$ are normalized and concatenated into a single signature vector;
- The signature matrix $\boldsymbol{A}$ is augmented by a row with the concatenated vector.



Concatenated shape and color signatures.

DIMENSIONALITY REDUCTION

- For every 2D projection, both shape and color signatures are computed, concatenated and included into $\boldsymbol{A}$;
- SVD of $\boldsymbol{A}$ is computed, with the resulting 1st left and right singular vectors being concatenated and used as the final descriptor.

PIPELINE

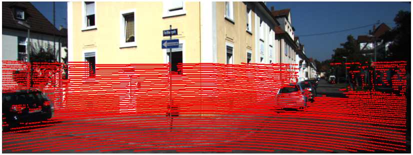
- Compute a descriptor for each point cloud and query them against the database:
  - Brute-force matching approach using the $L2$ norm;
  - Detection comes down to finding the most similar descriptor under a predefined threshold (later used for PR curves).

POINT CLOUDS

- KITTI[8] sequences 00, 05, 06 and 07 were used:
  - 3D LIDAR with 360° FoV, and a forward facing stereo color camera system, providing synchronized frames and rectified images;
- For each sequence, we generated semi-dense and dense point clouds offline, using sensors readings and public available tools;

[8] (GEIGER; LENZ; URTASUN, 2012)

- kitti_lidar_camera[9] package (ROS) was used:
    - LIDAR limited to forward facing FoV;
    - 3D points were projected onto 2D image, associating color values.



3D LIDAR points projected on 2D image. Frame from the KITTI odometry dataset.

[9] https://github.com/LidarPerception/kitti_lidar_camera

STEREO DEPTH ESTIMATION

- image_undistort[10] package (ROS) was used:
  - Employs block matching technique from OpenCV[11];
  - Point clouds generated using default parameters for KITTI sequences.

Depth estimated from stereo camera. Frame from the KITTI odometry dataset.

[10] https://github.com/ethz-asl/image_undistort

[11] https://opencv.org/

OUR PLATFORM

- Laptop Intel i7 quad-core 2.00 GHz CPU and 8 GB RAM;
- Both M2DP and c-M2DP were implemented in C++, using PCL[12] and Eigen[13];
- In order to compare our results, we used the global variant of the CSHOT descriptor provided by PCL.

[12] http://pointclouds.org
[13] http://eigen.tuxfamily.org/

PARAMETERS

- M2DP and c-M2DP parameters were the same from original work;
- c-M2DP color bins parameter was set as $g = h$;

M2DP and c-M2DP Parameters

Parameter	M2DP	c-M2DP
Azim. angles (b)	4	4
Elev. angles (q)	16	16
Conc. circles (l)	8	8
Shape bins (h)	16	16
Color bins (g)	-	16
Vector length	192	576

PARAMETERS

- CSHOT default parameters from PCL (vector length: 1344);
- Normals are estimated for CSHOT using the neighborhood around each point, requiring a radius parameter:
  - It can be insufficient, generating invalid results, or be a costly process due to the amount of points;
  - Before each sequence, using the 1st frame:
    - Radius was set as 5 times the average distance of the nearest point.

EVALUATION

- Times to compute each descriptor and perform the matching process were measured;
- Precision-recall curves were generated by varying the descriptor similarity threshold:
  - Two locations are considered as the ground truth loop closure if their distance is $< 10m$;
- Recall rates at 100% precision are highlighted:
  - False loop closures are catastrophic for the map building and can be irrecoverable for SLAM.

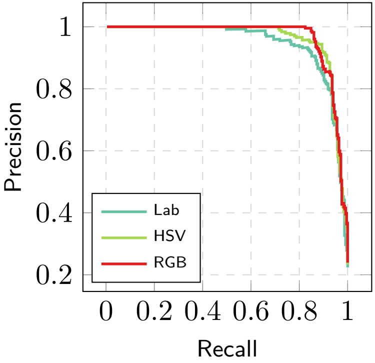
EVALUATION

- At first, we evaluated each descriptor with semi-dense clouds, generated through camera-LIDAR fusion;
- After that, we experimented with more dense clouds, generated through stereo depth estimation.

c-M2DP COLOR SPACE

- c-M2DP color space was chosen after evaluating it using RGB, HSV and CIELab.

Color Space	Recall Rates	
	Pr. 100%	Pr. 90%
RGB	**82.5%**	89.2%
HSV	71.4%	91.5%
CIELab	49.8%	86.8%



KITTI 06 camera-LIDAR.

TIME EFFICIENCY

- In camera-LIDAR sequences:
  - c-M2DP computing time is only $\mathbf{23.2}\%$ higher than M2DP;
  - c-M2DP is $\mathbf{22.6}\%$ faster to compute than CSHOT.

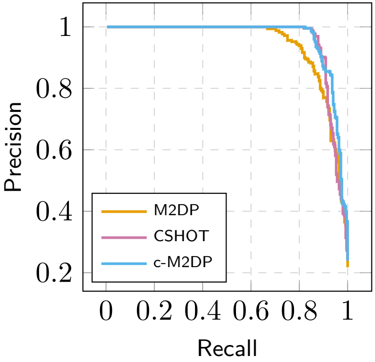
Descriptor	**Computing** (s)	**Matching** (s)
M2DP	$\mathbf{0.0674} \pm 0.0041$	$\mathbf{0.0043} \pm 0.0004$
c-M2DP*	$\mathbf{0.0830} \pm 0.0052$	$\mathbf{0.0051} \pm 0.0006$
CSHOT	0.1072 ± 0.0168	0.0059 ± 0.0005

*Ours

PRECISION-RECALL CAMERA-LIDAR

Descriptor	Recall Rates	
	Pr. 100%	Pr. 90%
c-M2DP*	**82.5%**	89.2%
M2DP	66.8%	82.1%
CSHOT	81.9%	90.6%

*Ours

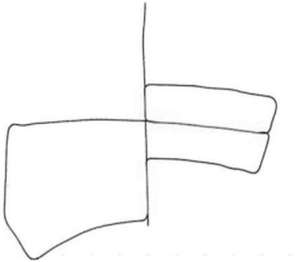


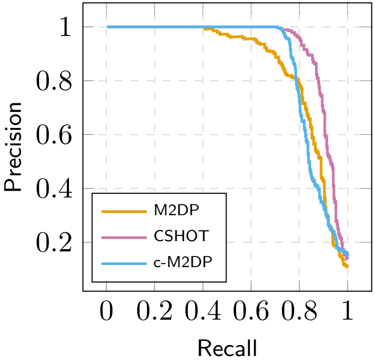
KITTI 06 camera-LIDAR.

PRECISION-RECALL CAMERA-LIDAR

Descriptor	Recall Rates	
	Pr. 100%	Pr. 90%
c-M2DP*	**70.9%**	76.5%
M2DP	40.9%	78.2%
CSHOT	**70.8%**	83.5%

*Ours



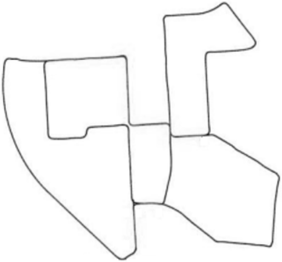


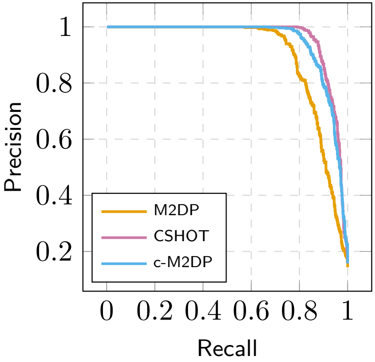
KITTI 05 camera-LIDAR.

PRECISION-RECALL CAMERA-LIDAR

Descriptor	Recall Rates	
	Pr. 100%	Pr. 90%
c-M2DP*	67.3%	85.4%
M2DP	57.4%	78.2%
CSHOT	**79.2%**	88.6%

*Ours



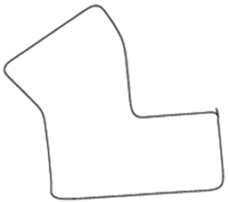


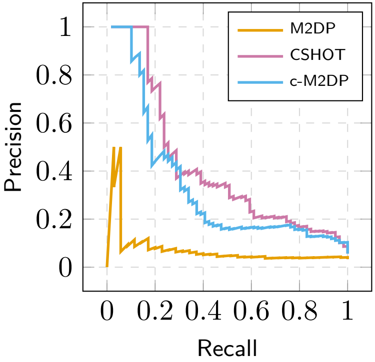
KITTI 00 camera-LIDAR.

PRECISION-RECALL CAMERA-LIDAR

Descriptor	Recall Rates Pr. 100%
c-M2DP*	10.2%
M2DP	-
CSHOT	**17%**

*Ours





KITTI 07 camera-LIDAR.

TIME EFFICIENCY

- In stereo sequences:
  - Overall increase in the average times computing the descriptors;
  - c-M2DP computing time is only $\mathbf{18.8}\%$ higher than M2DP;
  - CSHOT heavy computational burden, with an average time $\mathbf{315.9}\%$ higher than c-M2DP.

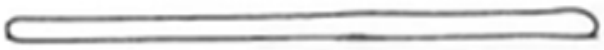
Descriptor	**Computing (s)**	**Matching (s)**
M2DP	$\mathbf{0.3584} \pm 0.0816$	$\mathbf{0.0044} \pm 0.0008$
c-M2DP*	$\mathbf{0.4259} \pm 0.0956$	$\mathbf{0.0054} \pm 0.0006$
CSHOT	1.7711 ± 1.0159	0.0061 ± 0.0005

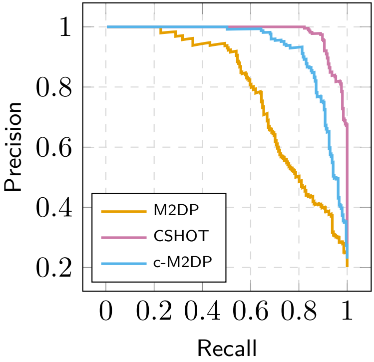
*Ours

PRECISION-RECALL STEREO

Descriptor	Recall Rates	
	Pr. 100%	**Pr. 90%**
c-M2DP*	50.2%	82.2%
M2DP	22.8%	54.5%
CSHOT	**82.3%**	91.7%

*Ours



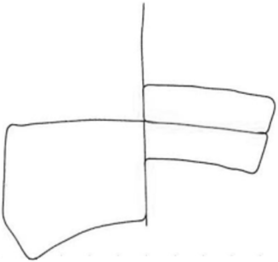


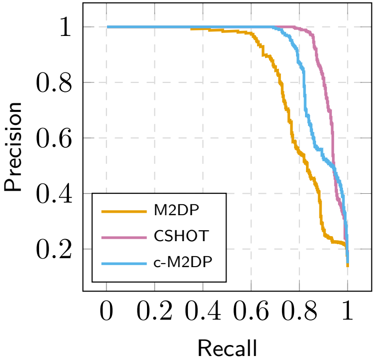
KITTI 06 stereo camera.

PRECISION-RECALL STEREO

Descriptor	Recall Rates	
	Pr. 100%	**Pr. 90%**
c-M2DP*	69.2%	79.2%
M2DP	35.3%	64.9%
CSHOT	**77.9%**	87%

*Ours



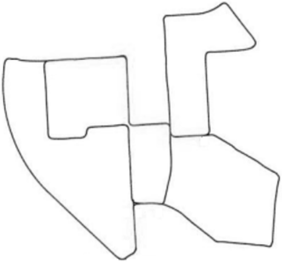


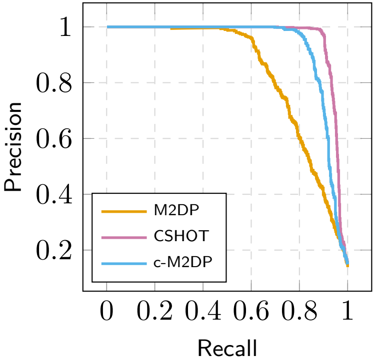
KITTI 05 stereo camera.

PRECISION-RECALL STEREO

Descriptor	Recall Rates	
	Pr. 100%	Pr. 90%
c-M2DP*	69.8%	85.3%
M2DP	27%	63.2%
CSHOT	**70.9%**	92.2%

*Ours



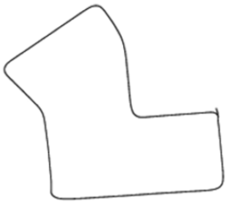


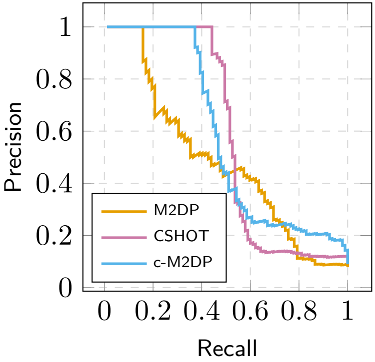
KITTI 00 stereo camera.

PRECISION-RECALL STEREO

Descriptor	Recall Rates Pr. 100%
c-M2DP*	37.2%
M2DP	15.9%
CSHOT	**44.2%**

*Ours





KITTI 07 stereo camera.

CONCLUSION

- Our proposal successfully incorporates color along shape data, extending the M2DP descriptor;
- We performed loop closure detection using c-M2DP:
  - Accuracy improvement over M2DP, while avoiding a large increase in time consumption;
  - Smaller, faster to compute, and shows competitive results against CSHOT in semi-dense point clouds;
  - Although dense point clouds were challenging for M2DP and c-M2DP, CSHOT higher accuracy comes at the cost of being several times slower than our proposal.

CONCLUSION

- Our paper was accepted in IEEE CASE 2019.
- In future works:
  - Evaluate performance using 360° colored point clouds sequences[14], and on different environments[15];
  - Investigate potential improvements for signatures, such as image pre-processing and point cloud sampling techniques.

[14] (PANDEY; MCBRIDE; EUSTICE, 2011)
[15] (BLANCO-CLARACO; MORENO-DUEÑAS; GONZÁLEZ-JIMÉNEZ, 2014)

OBRIGADO!

Leonardo Perdomo
Edson Prestes e Silva Junior
Instituto de Informática — UFRGS

- Precision is the proportion of correctly detected loop closures (TP) among the total of detected loop closures (TP+FP);
- Recall is the proportion of correctly detected loop closures (TP) among the actual loop closures in the sequence (TP+FN);

- Azimuth angles progression starts from 0 with a stride of $\frac{\pi}{b}$;
- Elevation angles progression starts from 0 with a stride of $\frac{\pi}{2q}$;
- Concentric circles are generated with varying radii $[r, 2^2 r, \ldots, l^2 r]$, where r is derived from the maximum radius, which is the distance between the farthest point of the cloud and the centroid.

- PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called PCs:
    - The 1st PC has the largest possible variance, and each succeeding PC has the highest variance possible under the constraint that it is orthogonal to the preceding PC.
- SVD is a factorization of a real or complex matrix:
    - In $\boldsymbol{A} = USV^T$, the columns of U and the columns of V are called the left-singular vectors and right-singular vectors of $\boldsymbol{A}$.