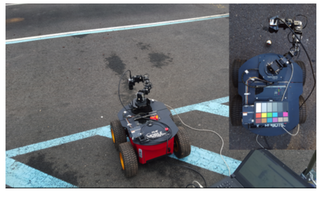
Differential Angular Imaging for Material Recognition

Jia Xue[1], Hang Zhang[1], Kristin Dana[1], Ko Nishino[2]
[1]Rutgers University, Piscataway, NJ, [2]Drexel University, Philadelphia, PA.

Overview:

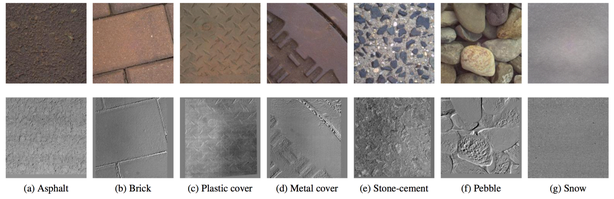
- Ground Terrain in Outdoor Scenes (GTOS) database: over 30,000 images, 40 classes, varying weather and lighting conditions.
- Differential angular imaging: middle-ground between reflectance-based and image-based material recognition.
- Differential Angular Imaging Network (DAIN): surpasses single view or coarsely quantized multiview images.

Measure equipment:



- Mobile Robots P3-AT robot
- Basler aca2040-90uc camera
- Macmaster-Carr 440C stainless steel Sphere
- Cyton gamma 300 robot arm
- Edmund Optics 25mm /F1.8 lens
- DGK 18% white balance and color reference card

Differential Angular Imaging:

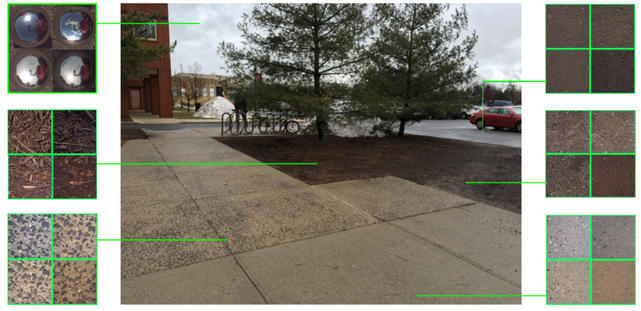


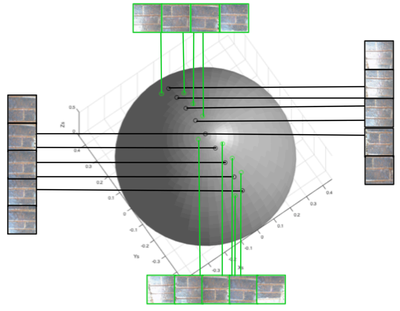
- Reveal the gradients at particular view point
- Observe relief texture
- Sparsity provides computational advantage

GTOS Dataset:

40 surface classes, 4 to 14 instances, 19 viewing directions, 4 different weathers, 3 different exposure times

Datasets	samples	classes	views	illumination	in scene	scene image	camera parameters	year
CUReT [8]	61	61	205		N	N	N	1999
KTH-TIPS [15]	11	11	27	3	N	N	N	2004
UBO2014 [42]	84	7	151	151	N	N	N	2014
Reflectance disk [43]	190	19	3	3	N	N	Y	2015
4D Light-field [40]	1200	12	1	1	Y	N	N	2016
NISAR [3]	100	100	9	12	N	N	N	2016
GTOS(ours)	**606**	**40**	**19**	**4**	**Y**	**Y**	**Y**	**2016**

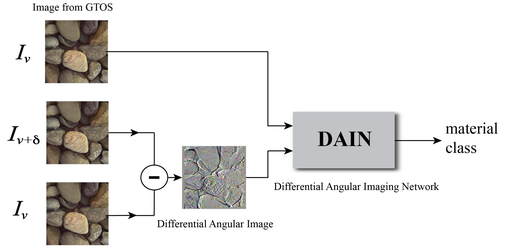




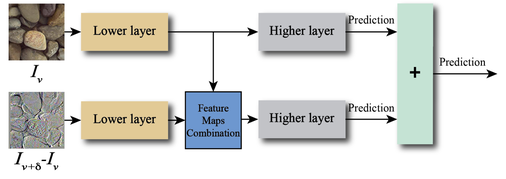
Application:

- **Robot Navigation & Automatic Driving**
  - Determining control based on current ground terrain
  - Determining road conditions by partial real time reflectance measurements
- **Photometric Stereo**
  - Estimating the surface normal of materials
- **Shape Reconstruction**
  - Capturing the shape and appearance of materials

DAIN for Material Recognition:



Method Overview



DAIN (differential angular image network)

Experiment:

Method	First input	Second input	Accuracy
single view CNN	I_v	-	$74.3_{\pm 2.8}$
multiview CNN, voting	I_v	-	$78.1_{\pm 2.4}$
multiview CNN,3D filter	I_v	-	$74.8_{\pm 3.2}$
single view DAIN (Sum)	I_v	$I_{v+\delta}$	$77.5_{\pm 2.7}$
single view DAIN (Sum)	I_v	I_δ	$79.4_{\pm 3.4}$
single view DAIN (Max)	I_v	I_δ	$79.0_{\pm 1.8}$
multiview DAIN (Sum/voting)	I_v	I_δ	$80.0_{\pm 2.1}$
multiview DAIN (Sum/pooling)	I_v	I_δ	$81.2_{\pm 1.7}$
multiview DAIN (3D filter/pooling)	I_v	I_δ	$81.1_{\pm 1.5}$

- Compare standard CNN with DAIN
- Single view DAIN achieves better recognition than multiview CNN
- Multiview DAIN (Sum/pooling) and multiview DAIN (3D filter/pooling) perform best

Architecture	Accuracy
FV+CNN [4]	75.4%
FV-N+CNN+N_{3D} [10]	58.3%
MVCNN [36]	78.1%
multiview DAIN (3D filter), pooling	**81.4%**

Compare with state of art algorithms on GTOS dataset

Acknowledgement:

This work was supported by National Science Foundation award IIS-1421134. A GPU used for this research was donated by the NVIDIA Corporation. Thanks to Di Zhu, Hansi Liu, Lingyi Xu, and Yueyang Chen for help with data collection.