Modeling Extent-of-Texture Information for Ground Terrain Recognition

Shuvozit Ghose\textsuperscript{a}, Pinaki Nath Chowdhury\textsuperscript{b}, Partha Pratim Roy\textsuperscript{c}, Umapada Pal\textsuperscript{b}

\textsuperscript{a}Institute of Engineering and Management, Kolkata, India. \\
\textsuperscript{b}Indian Statistical Institute, Kolkata, India. \\
\textsuperscript{c}Indian Institute of Technology, Roorkee, India.

Presenter: Pinaki Nath Chowdhury 
Venue: International Conference on Pattern Recognition, 2020
Contents:

- Challenges in Ground Terrain Recognition
- Motivation
- An Overview of our Solution
- Modeling Extent-of-Texture (EoT) Information
- EoT Guided Inter-domain Message Passing
- Results
- Conclusion
Ground Terrain Recognition is a difficult task due to various reasons.

The context information varies significantly over the regions of a ground terrain image. Like some local regions possess significant texture information, while shape information is more dominant at some other parts.
Motivation

As most real-world ground terrain images show wide variations in texture and shape information at different local regions in an image, thus the classification of such realistic ground terrain images requires a more local level modeling of texture and shape information.
An Overview of our Solution

We propose a novel approach towards ground-terrain recognition via modeling the Extent-of-Texture information to establish a balance between the order-less texture and ordered-spatial information locally.
Modeling Extent-of-Texture (EoT) Information

Given an image $I \in \mathbb{R}^{H \times W \times 3}$, a backbone CNN feature extractor network $G(\cdot)$ takes $I$ and outputs latent feature representation $Z$. Thus,

$$Z = G(I; \theta_G)$$  \hfill (1)

Patch-extraction is performed on $Z \in \mathbb{R}^{8 \times 8 \times 512}$ using a sliding window mechanism where the window size and stride is chosen as $(3 \times 3)$ and 1 respectively. The patch-extraction operation generates $\psi = \{\psi_i\}_{i=1}^{k}$, where $\psi_i \in \mathbb{R}^{3 \times 3 \times 512}$ and $k$ is the number of patches. Average pooling of $\psi$ gives $\psi^* = \{\psi_i^*\}_{i=1}^{k}$. 
Modeling Extent-of-Texture (EoT) Information (continued...)

Let $X = \{x_1, x_2, x_3, \ldots, x_k\}$, where $x_i$ denotes the central region of the $\psi_i$ patch e.g. $x_i = \psi_i[2; 2; :]$ and $x_i \in \mathbb{R}^{1 \times 1 \times 512}$.

The cosine similarity between $\psi^*$ and $X$ describes the order-less texture information $\mathcal{T}$, where $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_k\}$ and $\mathcal{T}_i$ denotes the order-less texture information of the $i^{th}$ patch. Therefore,

$$\psi_i^* = \text{AvgPool}(\psi_i, 3) \quad (2)$$

$$\mathcal{T}_i = \frac{\psi_i^* \cdot x_i}{\|\psi_i^*\|_2 \|x_i\|_2} \quad (3)$$

$$\mathcal{T}_i' = \frac{\mathcal{T}_i - \mathcal{T}_{\text{min}}}{\mathcal{T}_{\text{max}} - \mathcal{T}_{\text{min}}} \quad (4)$$
A high value of $\mathcal{T}$ indicates the presence of greater extent of the order-less texture information, whereas a small value of $\mathcal{T}$ represents higher shape information.

The ordered shape information $S$, where $S = \{S_1, S_2, S_3, \ldots, S_k\}$ and $S_i$ denotes the ordered-spatial information of the $i^{th}$ patch. Then,

$$S_i = 1 - \mathcal{T}_i$$  \hspace{1cm} (5)
The EoT Guided Inter-domain Message Passing module is used for sharing knowledge between texture and shape features to balance out the order-less texture information with ordered-spatial information.
**Results**

Table: Comparison of Deep-TEN, baseline B1, B2, B3 and B4 with the proposed methodology for single scale and multi scale training on GTOS-mobile [1] dataset using a pre-trained ResNet-18 module as the convolutional layer. Baseline B1 is similar to Deep Encoding Pooling Network (DEP) by Xue [1].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Scale</td>
<td>74.22</td>
<td>76.07</td>
<td>77.81</td>
<td>78.55</td>
<td>78.93</td>
<td><strong>80.39</strong></td>
</tr>
<tr>
<td>Multi Scale</td>
<td>76.12</td>
<td>82.18</td>
<td>83.78</td>
<td>84.31</td>
<td>84.36</td>
<td><strong>85.71</strong></td>
</tr>
</tbody>
</table>
Table: Comparing Our method with several state-of-the-art methods on Describable Textures Dataset (DTD) and Materials in Context Database (MINC)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FV-CNN [5]</td>
<td>72.3</td>
<td>63.1</td>
</tr>
<tr>
<td>Deep-TEN [2]</td>
<td>69.6</td>
<td>80.4</td>
</tr>
<tr>
<td>DEP [1]</td>
<td>73.2</td>
<td>82.0</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>75.7</td>
<td>85.3</td>
</tr>
</tbody>
</table>
Conclusion

- we have proposed a novel approach towards ground-terrain recognition via modeling the extent of texture information to establish a balance between the order-less texture component and ordered-spatial information locally.
- The driving idea of our architecture is the modeling of context information locally.
- The proposed framework is simple and easy to implement.
- We demonstrate the effectiveness of our system by conducting experiments on publicly available ground terrain datasets.
References


S. Bell, P. Upchurch, N. Snavely, and K. Bala, “Material recognition in the wild with the materials in context database,” in *CVPR*, 2015.

Thank you

Questions?

Source Code is available at:

github.com/ShuvozetGhose/Ground-Terrain-EoT