Describing Textures in the Wild

Mircea Cimpoi1, Subhransu Maji2, Iasonas Kokkinos3, Sammy Mohamed4, Andrea Vedaldi1

1Visual Geometry Group, University of Oxford (mircea, vedaldi)@robots.ox.ac.uk 2Toyota Technological Institute smaji@ttic.edu 3Ecole Centrale Paris/INRIA-Sadiy.as.kokkinos@ecp.fr 4Stony Brook sammy.mohamed@stonybrook.edu

This work is partially supported by NSF grant #1005411, ODNI via the TIPS2 grant VisRec no. 109786, and the EU-projects TIPS2, and TIPS2-Joint Subjective Attributes. It was partially supported by ANR – Project ArtiVis

Conclusions
• Introduced a large texture dataset, exhaustively labelled with joint subjective attributes.
• Proposed a low dimensionality, meaningful, texture descriptor based on descriptive texture attributes.
• Set new state-of-the-art on challenging material datasets.

References

Acknowledgements
This research is based on work of the 2012 CLSP Summer Workshop. It was partially supported by NSF grant #0904911, ONR via the JPL HATCoE and Google Research. Mircea Cimpoi was supported by ERC grant VeReNo. 228181 and Iasonas Kokkinos by ANR-10-JCJC-0035.

Describing Textures

Data Collection
• Texture vocabulary:
  • Starting point: list of 98 words in [Bhushan 97]
  • Discarded non-visual words (e.g. "jumbled" or "rhythmic")
  • Merged similar words (e.g. "corkscrewed" + "coiled" + "spiralled")
  • Example images:
    • Consider each word as key attribute
    • Query Google (e.g. "conscrumed textures", "coiled pattern")
    • Discard or crop images covered by less than 90% with content representing the query

Coarse-to-Fine Joint Annotation
Annotations using Amazon MTurk Stage 1
Verify key attributes.
Stage 2
• Sequentially collect joint annotations based on co-occurrence probability;
• Avoid labelling low probability attributes, given key attribute;
• Using classifier scores to further reduce the number of annotations;
• Seek for consensus of multiple annotations (5 per image).

Local Descriptor Comparison on DTD
• Bag of Visual Words approach
• 470 dimensional vocabularies, built using K-means
10 visual words per texture
• Filter banks, SIFT, LBP and image patches as local descriptors
• SVM with several kernels: linear, Hellinger, exponential χ²

Descriptive Attributes as Representation
• Use the scores from the 4 classifiers trained on DTD as a meaningful, low dimensionality descriptor.
• Low dimensionality allows to apply an RBF kernel.
• DTD descriptor learned on IFV + DeCAF, alone, exceeds previous state-of-the-art on FMD and KTH-TIPS2-b.
• Combined with IFV and DeCAF results in more than 10% above previous best.

Normalization

<table>
<thead>
<tr>
<th>Dataset</th>
<th>IFV</th>
<th>BOYW</th>
<th>VLAD</th>
<th>DeCAF</th>
<th>IFV + DeCAF</th>
<th>Previous best</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURTeT</td>
<td>99.6+0.4</td>
<td>98.110.9</td>
<td>99.1+0.6</td>
<td>98.6+0.4</td>
<td>99.8+0.2</td>
<td>99.4</td>
</tr>
<tr>
<td>UMD</td>
<td>99.2+0.4</td>
<td>98.113.0</td>
<td>99.3+0.7</td>
<td>97.4+0.7</td>
<td>99.0+0.3</td>
<td>99.7+0.3</td>
</tr>
<tr>
<td>UIUC</td>
<td>97.2+0.8</td>
<td>94.6+1.3</td>
<td>97.3+0.9</td>
<td>95.5+0.9</td>
<td>99.0+0.3</td>
<td>99.4+0.4</td>
</tr>
<tr>
<td>KTH</td>
<td>96.2+0.0</td>
<td>98.1+0.8</td>
<td>97.8+0.8</td>
<td>96.4+0.8</td>
<td>99.2+0.2</td>
<td>99.4+0.4</td>
</tr>
<tr>
<td>KTH-TIPS2a</td>
<td>82.5+1.3</td>
<td>78.4+5.4</td>
<td>77.6+4.3</td>
<td>77.7+2.0</td>
<td>84.6+3.6</td>
<td>73.0+4.7</td>
</tr>
<tr>
<td>KTH-TIPS2b</td>
<td>69.3+0.9</td>
<td>58.4+2.2</td>
<td>61.7+2.2</td>
<td>70.4+1.8</td>
<td>76.0+2.9</td>
<td>66.3</td>
</tr>
<tr>
<td>FMD</td>
<td>58.1+1.7</td>
<td>49.5+1.9</td>
<td>54.8+1.2</td>
<td>57.6+1.2</td>
<td>65.6+1.4</td>
<td>57.1</td>
</tr>
<tr>
<td>DTD</td>
<td>58.6+2.0</td>
<td>53.6+1.5</td>
<td>57.3+1.5</td>
<td>52.5+1.3</td>
<td>64.7+1.6</td>
<td>--</td>
</tr>
<tr>
<td>DTD(J-API)</td>
<td>59.2+1.8</td>
<td>52.2+2.2</td>
<td>58.5+2.4</td>
<td>55.4+2.1</td>
<td>66.7+2.2</td>
<td>--</td>
</tr>
</tbody>
</table>

Feature set: KTH-TIPS2-b, FMD

State of the Art on Texture Datasets
• Experiment with various encodings on top of best performing local descriptor (SIFT)
• Improved Fisher Vector (IFV) and Deep Convolutional Feature (DeCAF) are tuned for object recognition,
but perform very well on textures
• Combined, lead to state-of-the-art results on all datasets

Conclusions
• Introduced a large texture dataset, exhaustively labelled with joint subjective attributes.
• Proposed a low dimensionality, meaningful, texture descriptor based on descriptive texture attributes.
• Set new state-of-the-art on challenging material datasets.