Since our last update, we have fully launched a GPU version of HCC, and below we highlight some useful improvements on image classification. Thank you all for your enthusiasm and important support.

**Classifier improvements**

Image features computed on the Grid are used to build an image classification system. We measure the accuracy of our classifier by comparing its output against scores assigned by human experts. Some images are difficult to interpret, causing experts to score these images differently.

While not yet fully finished and optimized, we have prepared an improved 11-way logistic regression classifier, which exceeds the HPCS\(^1\) Random Forest model on precision, recall, or both for all image classes. Recall is improved on the most common classes, clear (98.8%) and precipitate (96.0). Precision is improved on crystal (74.9) and other less-common classes.

When the 11 classes are aggregated into three super-classes (clear, has-crystal, other), the results are:

<table>
<thead>
<tr>
<th>Truth</th>
<th>clear</th>
<th>cryst</th>
<th>other</th>
<th>Total</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>clear</td>
<td>5706</td>
<td>0</td>
<td>68</td>
<td>5774</td>
<td>99%</td>
</tr>
<tr>
<td>cryst</td>
<td>240</td>
<td>1522</td>
<td>1374</td>
<td>3136</td>
<td>49%</td>
</tr>
<tr>
<td>other</td>
<td>222</td>
<td>134</td>
<td>8358</td>
<td>8714</td>
<td>96%</td>
</tr>
<tr>
<td>Total</td>
<td>6168</td>
<td>1656</td>
<td>9800</td>
<td>17624</td>
<td></td>
</tr>
</tbody>
</table>

Note that this model is more precise about images it identifies as crystal, at a cost of poor recall (<50%). We continue to work on a neural network-based model that should provide a substantial improvement.

**Other progress**

In this direction, we have created SCRPDB\(^2\), which has been accessed by 691 unique users (since January 2012). Importantly, SCRPDB is now deposited at http://pubchem.ncbi.nlm.nih.gov, and is the 7th largest source of deposited structures (out of 207; and the second largest academic deposition) - 2.5 times larger than the deposit from IBM last Fall.

While not directly linked to the HCC project, we have also co-authored several related papers:


