Web Accessible Image Similarity Measurements

Antoine Vandecremer, Paul Khouri Saba, Ben Long, Joe Chalfoun, Peter Bajcsy, Mary Brady

Background

There is a need to understand the variability of biological conclusions due to the choice of a similarity metric, and due to the software quality and parameters of similarity computations. The goal of our effort is to advance high throughput and high confidence image comparisons accessed from any stationary or mobile computer device. We built a web accessible and computationally scalable system composed of image similarity metrics. The similarity metrics are validated regularly by pre-configured tests. We have also added a workflow editor that allows access from a variety of platforms.

Approach

Our approach is based on organizing and evaluating image similarity metrics first according to several existing surveys of image similarities. The similarity metrics are represented by a triplet consisting of image loaders and color space representations, image descriptors, and proximity measures. The proximity measures are grouped into those that can operate on histogram descriptors, contiguous image segments, clusters of image pixels or raw pixel values. This classification of individual computation and their sub-categories allows us to build a simple tree taxonomy encapsulating image loading/representation, image characterization and comparison, and to map the taxonomy into intuitive web interfaces.

The Web application is based on the Google Web Toolkit (GWT) client and NCSA’s MediC Multimedia Content Management System. It provides functionalities for image upload, storage and annotations. The GWT client connects to the REST API and starts computations on it. It can start comparisons on collections and save the computations results for later use.

Access to Measurements via Web Services

The computation system implementation is based on a RESTful API which allows to:

- Get the list of available adapters, extractors & measures
- Submit new comparisons and gather the results
- Connect multiple slaves

Slaves can provide:

- Computational resources
- New algorithms

Web Access to Image Similarity Measurements

Catalog of Image Similarity Metrics

<table>
<thead>
<tr>
<th>Family</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Squared Family</td>
<td>AdditiveSymmetricChiSquared, Clark, Divergence, NeymanChiSquared, PearsonChiSquared, ProbabilisticSymmetricChiSquared, SquaredChiSquared, SquaredEuclidean</td>
</tr>
<tr>
<td>Combinations Family</td>
<td>AvgDifference, KumarJohnsonDifference, TanakaDifference</td>
</tr>
<tr>
<td>Fidelity Family</td>
<td>Bhattacharyya, Fidelity, Hellinger, Matusita, SquaredChord</td>
</tr>
<tr>
<td>Inner Product Family</td>
<td>Cosine, HarmonicMean, InnerProduct</td>
</tr>
<tr>
<td>Intersection Family</td>
<td>Czekanowski, Dice, IntersectionJaccard, Kullback-Leibler, KolmogorovSmirnov, Kulczynski, KuhaHassebrookPES, Meijia, Ruzicka, Tanimoto, WaveHedges</td>
</tr>
<tr>
<td>L1 Family</td>
<td>Canberra, Gower, Kulczyński, Lorentzian, Soergel, Sorensen</td>
</tr>
<tr>
<td>Lp Minkowski Family</td>
<td>ChebyshevLInf, CityBlock,1, Euclidian,2, Minkowski</td>
</tr>
<tr>
<td>Pixel-based Family</td>
<td>AdjustedRandIndex, totalNumOfImageValuation, TotalErrorRateTest</td>
</tr>
<tr>
<td>Shannon’s Entropy Family</td>
<td>J-divergence, JensenDifference, JensenShannon, K-divergence, KulbackLeibler, Topsoe</td>
</tr>
</tbody>
</table>

Workflow decomposition of image similarity computations

- Data collection
- Comparison method selection
- Result viewing

Access to Measurements from Mobile Devices

The purpose of the Web-based Workflow editor is to build a WF with scientific data such as medical images and scientific computations (Accessed using XMLHttpRequest API), submit it to Taverna engine & retrieve the computations results. The tool is mainly written in HTML5 in order to take advantage of the cross-browser/platform capabilities, and support execution on mobile devices, desktops and laptops. Other HTML5 components like SVG (Scalable Vector Graphics) and Local Storage (aka Web Storage) are also used to draw shapes, and store and retrieve objects locally.

Image Similarity Testing and Validation

Data-driven tests, one component of the overall quality testing strategy. In this case, exercise unit and failure-mode test conditions to identify errors corresponding to specified fault categories. These tests demonstrate expected failures due to image incompatibilities across different image parameters as well as cross-platform consistency of results. They detect errors triggered in 1 of 5 primary fault categories: hardware (HW), software (SW), image compatibility (Image), consistency with mathematical definitions (Math), and consistent treatment of failure conditions (Singularity). Each line in the graph depicts an error distribution for a specific test on a given platform. Together they show a cross-platform consistency of response for the implemented model and differ. Each was performed with image collections across multiple modalities (fluorescent and phase-contrast), image parameters (pixel size, dimension, type, color model) and differ platforms (Windows and Linux, with different implementations of Java).

The largest test (last 2 lines) combined all modalities and dimensions into one test. 56 synthetic images were used representing single/multi-band, multiple pixel data types (byte, double), pixel sizes (8-32), RGB and grayscale image variations. Executed as 1,166,592 comparisons across both platforms, yielding consistent results.