Photo.net Dataset - March'09 Release

Ritendra Datta (datta [at] cse [dot] psu [dot] edu)

Description of columns of photonet_dataset.txt

Column 1: Index
Column 2: Photo-ID
Column 3: Number of aesthetics ratings received
Column 4: Mean of ratings
Column 5: Standard Deviation of ratings
Columns 6-12: Distribution (counts) of aesthetics ratings in 1-7 scale, 6 being for 1, and 12
being for 7.

NOTE: In a few cases, only the mean was available, the individual ratings were not available.
In these cases, only the mean should be used. Columns 5 to 12 will be -1 in these cases. If
your model involves distributions or std. dev. of ratings, please handle these cases as
appropriate.

How to obtain the original images?

The photo ID is unique, but the exact URL of the files may change on the Photo.net server.
Probably the best way to obtain the image URL is to download/wget the html page hosting the
image, and scrape that html file for the image URL (there is usually a pattern that can be
easily detected). The URLs for the html pages are currently of the following form:

    http://photo.net/photodb/photo?photo_id=<Photo-ID>
e.g.,
    http://photo.net/photodb/photo?photo_id=1008958

Instead of doing the above, at the time of this release, the images can be downloaded directly
from URLs of the following form:

    http://gallery.photo.net/photo/<Photo-ID>-md.jpg
e.g.,
    http://gallery.photo.net/photo/1008958-md.jpg

Copyright Issues

Please respect copyright that applies to you and to the data sources. Please do not re-
distribute the image data. This is the same reason we do not distribute the original images.