BA Thesis

Author: Matthias Moosburger

Title: Colour Labelling of Art Images Using Colour Palette Recognition

---

Colour Extraction Tool

The colour extraction algorithm is implemented using Python 3.5 and Anaconda open data science platform. Anaconda provides a handy package of all required dependencies. Without using Anaconda you will need to install at least the following dependencies: PIL, skimage, scipy.cluster, webcolors, matplotlib, numpy, pandas.

Colour extraction can be started with `colorextract.py` Python script on console.
The console script requires the following parameters:

- `input`: (mandatory) input csv file or path to directory with image
- `terms`: (mandatory) path to terms csv file
- `output`: (mandatory) path to output folder
- `-j` or `--json`: produces JSON output if set
- `-v` or `--verbose`: verbose output on STOUT

Example: `colorextract.py ./input_files.csv ./terms/ral_colors.csv ./output --json --verbose`

The extraction algorithm produces output CSV files containing the colour values per image (palette.csv), a metric table containing volume and maximum colour distance (metrics.csv), a list of error (errors.csv), a list of greyscaled images (bw.csv) and the list of images with colour term ids (terms.csv). The latter list references to IDs of colour terms list.

An example CSV file of images is given in `artigo`-folder. Example color term lists in CIELAB colour space are given in `color_terms` folder.

Results

The results on colour extraction and labelling can be found in `results` folder.

There you'll find the following files:

- `ral_basic_color_results_ordered`: An ordered list of images with basic color IDs from RAL color table (in `terms/ral_colors.csv`)
- `ral_color_results_ordered`: same as previous file but with accurate RAL colors
- `metrics_results.csv`: volume of convex hull and maximum

color distance per image (to compare chromaticy of images)

- `greyscale_results.csv`: IDs of images with only 1 greyscale channel
- `errors_results.csv`: errors occured (predominantly FileNotFound Errors)
- `color_palette_results.csv`: list of all colour values per image in CIELAB colour space (127 MB)

SQL and ARTigo

SQL queries are listed in `sql` folder tested on PostgreSQL.

A image title, artist and artwork_id Dictionary can be found in `artigo` folder with a `images.csv` file listing all images in ARTigo database.

Web

Files for web-interface are located in `web` folder.

To generate output for web-interface, the `--json` parameter has to be set on image processing. The web-interface eather allows to select images from the `file_list.json` and then needs a file list with files in `images` folder which can be created with Python-Script `make_file_list.py`. Or it can load images and JSON from `images` folder with URL parameter `file=<ID/Filename>`. The Web-Interface requires Internet-Connection for CDN Script loading (can be downloaded separately).

Misc.

In the `utils` folder, you'll find suitable tools for splitting and concating CSV files for Bash.

Contact

For any questions, suggestions, critique contact me:

mail@matthiasmoosburger.de