Cognitive Computing – Fall 2018

HW#2 (DUE: 11pm, Sunday, 4 November, 2018)

Note:

1) If you have any questions regarding the homework, send e-mail to the TA at <d06944009@ntu.edu.tw>.

2) Submit a soft copy of your write-up (in PDF) and programs to TA **before the due** to CEIBA. Name your zip file with your student ID. i.e. d06944009.zip. The images are not required.

3) The dataset is available at the CEIBA page (with password).

4) You are highly encouraged to write the homework in English.

5) Please DO write appropriate comments along with your codes.

The goal of the homework is to design and evaluate the visual features across different image categories measured by image matching (or retrieval) performance. For this work, we will use the parts of the Stanford Mobile Visual Search Dataset (SMVS) [1] for the evaluation. We will use 6 categories among them include "dvd_covers" (dvd), "book_covers" (book), "cd covers" (cd), "video frames" (video), "museum paints" (museum), and "business_cards" (card). In this dataset, each category contains around 100 reference (ground truth) images and 4 types of query images (each having 100). The number in each category associates the query with the ground truth. You can go through the dataset for more information. For the evaluation, we are to use all the queries (~400 for each category) for the image match (retrieval) among all the reference images (totally ~600 across 6 categories). The other reference images are used as the distractor for comparison. The evaluation metric is introduced below.

Also remember to visualize your query results (at your own and NOT needed to be submitted) and see if it matches the performance evaluation.

You will do **color, texture/shape, and local features** for the homework, and decide the proper similarity or distance metrics (e.g., L1, L2 distance, cosine similarity, etc.). Please brief the features and the similarity (distance) metric you adopt.

1) *Color similarity*. You can refer to the color features introduced in the lectures, or other relevant papers. Suggestions for color features are (but not limited to): *global color histogram, regional color histogram, grid color moments, means in each color channel, color (auto-) correlogram*. You are encouraged to implement on your own or can find other open source tools; however, you have to acknowledge the source of the tools.
2) *Texture/shape similarity*. You can choose any texture or shape features. Suggestions for such features are (but not limited to): *Fourier features, Laws' texture measures, co-occurrence matrix metrics, Tamura's textures, Gabor texture, PHOG, gradient histogram, edge histogram, etc*. You are encouraged to implement on your own or can find other open source tools for texture or shape features; however, you have to acknowledge the source of the tools.
3) *Local feature similarity*. We suggest DoG + SIFT for implementing local features. You are encouraged to use the visual word representations (as introduced in the lecture) or some advanced methods for pooling local features (e.g., Fisher, VLAD, etc.). It will incur some technical issues when doing clustering with large codebook size (e.g., K = 100k). Will suggest using smaller codebook size.

4) We are keen to different modalities and the performance. Here we are to evaluate the S@K (success at top K or hit at top K) – meaning that how many (%) of queries (in the specific category) can retrieve the ground truth (reference) in the top K ranked list. Note that when you do the query, please use the ALL reference images (~600) across categories as the candidates for ranking. You can try to fill in the table below for the submission. A graph comparison is optional and can be the alternative to show the results. Note that the fusion part is optional.

5) The homework is NOT graded by the performance. Try your best to see what you can achieve. However, we will invite few students having a short sharing for what you been doing for the work.

6) When doing the experiments, will suggest extracting the image features offline and save them as files first. It will speed up the experiments.

categories vs. methods	dvd		cd		book		painting		video		card	
	S@1	S@5	S@1	S@5	S@1	S@5	S@1	S@5	S@1	S@5	S@1	S@5
Color												
Texture (or Edge)												
Local Feature												
Fusion *(Optional)*												

References:

[1] Chandrasekhar et al., The Stanford Mobile Visual Search Dataset. ACM MMSys 2011. http://web.cs.wpi.edu/~claypool/mmsys-dataset/2011/stanford/
[2] Bekele et al., Evaluation of binary keypoint descriptors. ICIP 2013