

What to hand in for the Bag of Words Project

Due date is flexible. By the end of March would be good, earlier is even better. I'm more interested in seeing a thoughtful attempt rather than something rushed together to meet an arbitrary deadline. However, there is going to be a second project that you should start thinking about, so it's not in your best interest to drag out project 1 for too much longer.

Roughly 75% of your project grade will be based on submitting a written report discussing your program, design decisions, and experimental observations, and the other 25% on submitting "runnable" code. I'm not going to try to run your code, but will skim through by eye to see if it looks reasonable, both in terms of programming practice [documentation!] and whether it is plausible that the code has all the components needed to produce the results that you are showing.

The report should serve the purpose of helping me understand what you did. Recall that there were several design decisions to make when putting together a bag of words approach. A slide at the end of the bag of words class lecture laid out some of these design choices. It should be clear from your report what your decision was for each of those choices.

Project 1 Assignment

Choose Task: object recognition or scene recognition;
supervised vs unsupervised

Choose Dataset: if at all possible, choose a dataset that
already exists so you compare results.

Choose feature extraction method. Dense vs sparse;
what is your feature descriptor

Choose a quantization method, e.g. kmeans, EM,
mean-shift etc...

Choose parameter learning method: MLE or MAP, care
to try pLSA or LDA?

Choose a classifier method: Naïve Bayes vs SVM or
any discriminative classifier you want

Quantitative Evaluation: how will you measure how well
your method works?

A report layout something like the following would be good:

a) Make sure all group member names are on the report, so I know who is working with who!

b) Summarize in your own words what you were trying to achieve with your project. What was the task domain? What would successful performance be in that task domain? Think of this as analogous to the "abstract" section of a paper.

c) Present an outline of the procedural approach along with a flowchart showing the flow of control and subroutine structure of your code. This is also the place to present the design decisions you had to make (see slide above) and what you decided to do about them. Think of this as analogous to the “approach” section of a paper. Even if you chose to try to replicate the method and results of a previously published paper, still tell in your own words what the approach is, so that a reader who hadn’t read the original paper could still understand what you have done.

d) Run the different portions of your code and show pictures of intermediate and final results that convince me (and you) that the program does what you intended it to do. What does some sample input data look like? Can you visualize extracted features and clusters of features found? What do the bag of words feature vectors (e.g. histograms) look like for some sample classes? Do they appear discriminable? Imagine you were putting together an illustrated storyboard for how your approach works. Think of this as analogous to the figures in the “approach” section of a paper, or as material for a poster or presentation that would be used to describe the work to others.

e) Present your experimental results. What do you observe about the behavior of your program when you run it? Does it seem to work the way you think it should? Is the method sensitive to parameter values, and if so, how should one decide what values to use? Explain your quantitative evaluation criterion and show the quantitative results. Are these numbers reasonable? Have other papers reported results on the same or similar data? Is so, how do your results compare to theirs? Think of this as analogous to the “experimental results” section of a paper.

f) Discussion. In hindsight, was your approach to the problem successful? Were there some problem areas that arose which would lead you to make different design choices if you had to do this all over again from scratch? This is the place where you get to talk about what you learned and to give advice to some hypothetical grad student who might be interested in working on this problem in the future. Think of this as analogous to the “discussion and future work” section of a paper.

g) Finally, document what each team member did to contribute to the project. It is expected that you are dividing up the labor into different tasks, and it is OK if not everyone contributes precisely equal amounts of time/effort. However, if two people did all the work because a third member was slacking off, this is where you get to vent about that, so the grading can reflect the inequity. This is almost never an issue in grad-level classes like this one, but the option is here in case it is needed.

FAQ: How long should the report be? There is no strict number of pages as long as you are adequately addressing all of the above.