Visual Recognition
Week 3

Team 1
Eric López
Gerard Martí
Sergio Sancho
Adriana Fernández

Experiments and Implementations for week 3

For this week, we will implement a 2x2 spatial pyramid, and we will compare the results with the ones obtained in the previous week. We will also apply descriptors using an Intermediate Fusion scheme, and combine it with the pyramid approach to check whether is there any improvements.

Later, we will try to implement another layout of spatial pyramids, and see the effect that soft assignment can have to improving results.

From previous weeks, we have the following best results. We will apply the new methods on them to try to improve the results.:

Method	**Performance**
FAST+SIFT, k = 5000, nsamples = 50000	Accuracy = 0.86
DENSE + SIFT, k = 5000, nsamples = 50000	Accuracy = 0.81
Early Fusion (DENSE - SIFT+COLORHist)	Accuracy = 0.81
Early Fusion (Fast - SIFT+COLORHist)	Accuracy = 0.84

Comments

- We decided to reduce the number of samples from 5000 to 1000 due to memory problems, nevertheless the results we have obtained will not be too much affected.
- We tried different types SVM kernels of the given code:
  - Linear
  - HISVM
  - SPMKernel
- Finally we have studied the effect of the weight between the two vocabularies.

2x2 Spatial pyramid

The following table show the comparison of applying or not spatial pyramids in our system (k = 1000 samples = 50000, no PCA):

Detector	Descriptor	Classifier	Accuracy	
			No-SPM	SPM
FAST	SIFT	LinearSVM	0.8289	0.8327
		HISVM	0.8550	0.8674
		SPMKernel SVM	-	0.8624
FAST	SURF	LinearSVM	0.4957	0.5849
		HISVM	0.4944	0.5960
		SPMKernel SVM	-	0.5998

Intermediate Fusion + 2x2 Spatial pyramid

The following table show the comparison of applying or not spatial pyramids in our system (k = 1000 samples = 50000, no PCA):

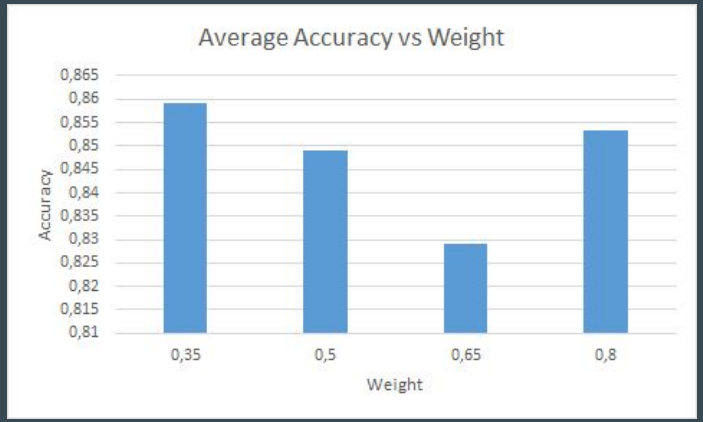
Descriptor	Color Descriptor	Classifier	Accuracy	
			No-SPM	SPM
SIFT	ColorNaming	LinearSVM	0.8228	0.8364 (0.50)
		HISVM	0.8587	0.8736 (0.50)
		SPMKernel SVM	-	0.8538(0.50)
	ColorHistogram	LinearSVM	0.7992 (0.50)	0.8401 (0.50)
		HISVM	0.8537 (0.50)	0.8785 (0.50)
		SPMKernel SVM	-	0.8562 (0.50)

Descriptor	Color Descriptor	Classifier	Accuracy	
			No-SPM	SPM
SURF	ColorNaming	LinearSVM	0.2565	0.3321
		HISVM	0.2466	0.4820
		SPMKernel SVM	-	0.5415
	ColorHistogram	LinearSVM	0.4077	0.4981
		HISVM	0.3693	0.5013
		SPMKernel SVM	-	0.3730

Importance of weighting

Spatial We are showing now the dependence of the accuracy with respect to the ratio of importance of shape/color.

Spatial Pyramid	Classifier	Accuracy			
		0.35	0.50	0.65	0.80
No	LinearSVM	0.8277	0.8228	0.7646	0.8216
	HISVM	0.8736	0.8587	0.8438	0.8761
	SPMKernel SVM	-	-	-	-
Yes	LinearSVM	0.8488	0.8364	0.8178	0.8395
	HISVM	**0.8822**	0.8736	0.8624	0.8736
	SPMKernel SVM	0.8636	0.8538	0.8575	0.8563



We have not found any strong dependence with the weight ratio. It may be caused by the fact that we did not make the study using the cross-validation, because of time constraints.

Using the best configuration obtained we have applied 5 fold cross validation to improve our results. With FAST as detector and SIFT (0.35) + COLOR(0.65) as descriptors we have obtained **0.8909** of accuracy. Then, using cross validation we have a gain of 0.98% of accuracy.

Soft Assignment

Until now, we were creating the codebooks by using hard-assignment. This is, assigning each data point to exactly one cluster. We did this by using k-means. Now, we are going to try soft assignment, where we assign a score to a data point for each cluster. This method makes that each data point could be assigned to more than one cluster.

GMM (Gaussian mixture models) clustering makes that each score is a posterior probability. Given their flexibility, we have tried to use GMM clustering in our previous methods with best performance to see if this clustering method is able to improve our results. We will implement GMM in simple Bag of words, Spatial Pyramids and in Intermediate Fusion, and see if the results are better or worse.

For each of the methods described before with best performance, we tried, with a nsamples of 60000, different sizes for the clustering (k), with the same parameters as before, and using SVM cross-validation. The results obtained can be found on the following slide, on the table.

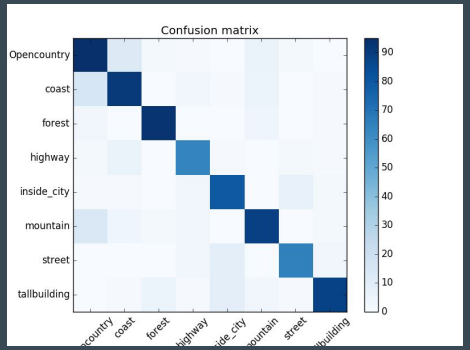
Soft Assignment - week 2 experiments

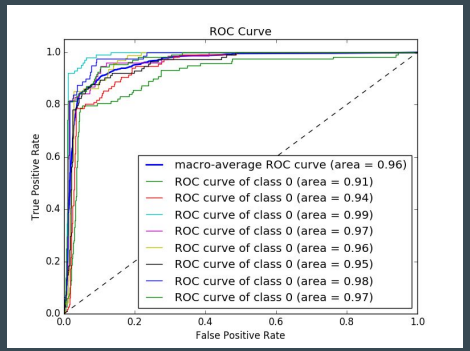
Method	k = 16	k = 32	k = 64	k = 80	k = 128	Using kmeans
FAST+SIFT, Linear SVM	0.71	0.802	0.825	0.824	0.832	0.86
DENSE + SIFT, Linear SVM	0.645	0.695	0.732	0.734	0.738	0.81

Table of accuracies for different values of k.

We obtain better performances with k = 80 in both methods, but have not been able to achieve a better performance than with using hard assignment. Despite that, results show that with more testing (different combinations of k and nsamples) we could be able to achieve at least a similar accuracy.

We have to note that computation time increases rapidly with higher number of k, but after a concrete number, the accuracy does not seem to improve.





CN and ROC curves for FAST+SIFT - k = 80

Soft Assignment - week 3 experiments

		Using k-means		Using GMM	
		No-SPM	SPM	No-SPM	SPM
FAST + SIFT + ColorHistogram	LinearSVM	0.7992 (0.35)	0.8401 (0.35)	0.8364 (0.35)	0.8748 (0.35)
	HISVM	0.8537 (0.35)	0.8785 (0.35)	0.8389 (0.35)	0.8686 (0.35)
	SPMKernel SVM	-	0.8562 (0.35)	-	0.85501 (0.35)

When we tried the best results of the Spatial Pyramid / Intermediate Fusion, we observe that the results do not change signfinicantly (we have worst results for almost all cases, except for LinearSVM using SPM). The similarity of the results leads us to believe, just as before, that it could be possible to find better parameters of the GMM clustering to get higher accuracy than k-means.

We also have to note that GMM with a high k is more computationally expensive.

Conclusions

- The spatial pyramid had improved the results in every case.
- SURF as a descriptor seems to be not adequate for this kind of application, it always show worse results.
- The SVM kernel that offer a better performance is HI-SVM.
- Applying differents weights to the combination of the vocabularies differents improvements are obtained, we cannot conclude with any tendence. In this case, the best result was weighing more color than shape.
- Soft assignment with GMM do not alter the results significantly, more testing could be needed to have a final answer.
- PCA was discarded because although the computational cost was lower, we could not find a good number of components to represent our data. The results obtained applying PCA always were worse.
- The system works pretty well, offering an accuracy of 89.09% in our best configuration.
- We could not try the 3x1 pyramid, but seeing that 2x2 improves our results. It is a task that we save for following tasks.