Abhishek Manoj Kumar, Student ID: 01675536

Date: 12/20/2016

Paper Information:

Title: Nonparametric Scene Parsing via Label Transfer

Authors: Ce Liu, Jenny Yuen, Antonio Torralba

Publisher: IEEE

Year: 2011

DOI: 10.1109/TPAMI.2011.131

Paper Summary:

The paper proposes a nonparametric approach to perform scene parsing and to recognize objects in an image using a new technique called label transfer, along with a combination of other techniques to label any input image. The key difference which differentiates this technique is the nonparametric approach and what this means is that there is no need to train a model for each object like in other parametric approaches. Here the input image is compared with the images in the existing database and similar images are found. The images in the Database are annotated i.e. each object (tree, house, car etc.) is labelled and when the closest matches are found for the input image, the labels are just transferred to the input image and the image is labelled. The images in the database are labelled using the LabelMe online annotation tool. There are 3 key tasks that are performed after an input image is obtained to label it. First, scene retrieval of the closest neighbouring images is done to obtain a set of images that closely match with the input image. The nearest neighbour is defined by combination of (K-NN) and (i+ε) which gives a (K , ε)- NN closest neighbours. To find the distance between any two images, two approaches are used; Euclidean distance of GIST and spatial pyramid histogram intersection of HOG is used, also they have made use of the combination of both. Second we find the Dense Scene Alignment using the Sift flow algorithm. The Sift flow algorithm is able to establish semantically meaningful matches between two images. By finding the dense alignment, the nearest neighbours obtained in the first set are reduced to M voting candidates that will be used for label transfer. To improve the performance, the Shift algorithm is modified into a coarse-to-fine pyramid SIFT flow matching algorithm. The final step is to parse the input image by transferring the labels obtained from the M voting candidates to the input image. The M voting images which are chosen by the SIFT flow algorithm are combined and the input image is parsed and labelled accordingly.

There are Two database LabelMe Outdoor(LMO) and SUN database on which the experiments are performed. The LMO dataset consists of only outdoor scenes and it is a subset of SUN dataset which consists of both indoor and outdoor scenes. LMO dataset consists of 2688 images out of which 2466 images are annotated and in the database and other 200 are used to test the system. There are 4 key parameters on which the recognition rate for the 200 images depend on, which are K, M, α and β. Here K corresponds to the number of nearest neighbours which are found for any image. M is the number of voting candidates selected from the nearest neighbour. The α and β represent the Markov random field representing the prior and spatial smoothness. Various results are obtained by changing these parameters and are visualized. Overall the Nonparametric approach has several advantages over the standard parametric approach such as; its openness i.e. if we are needed to add more objects categories then we can simply add annotated images of that object into the database without having to train for each object and also other algorithms can be used here instead of SIFT or GIST or HOG which makes it easier to make the system more efficient in the future without changing the whole experiment. Some future improvements of the system are also proposed such as taking the synonyms, co-occurrence and occlusion relationship between the labels of the pixels.

Reproducibility Overview:

Title: Nonparametric Scene Parsing via Label Transfer.

Code: The code is implemented using Matlab and is made available at http://people.csail.mit.edu/celiu/LabelTransfer/code.html

Data: Datasets are made available along with the code in the same link. There are 2 datasets on which the experiments are run, datasetLMO which consists of 2688 images all outdoor images, divided into 2488 for training and 200 for testing and datasetSUN which consists of 9556 images both indoor and outdoor, divided into 8556 for training and 1000 for testing. The testing datasets have to be created for each method of finding nearest neighbour (GIST, HOG, ALL) and for finding the voting candidates by calculating the dense scene alignment (SIFT flow).

VM's and Containers: No VM or Containers are made available.

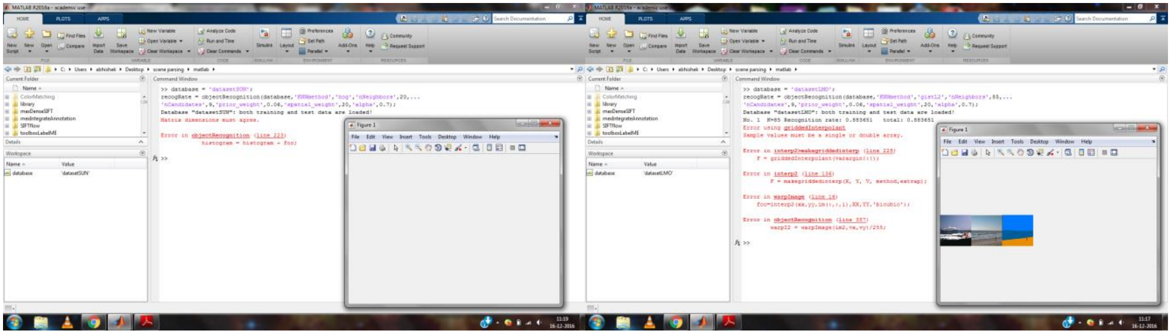
Workflows: No workflow is provided but Details of how to implement the code and datasets is made available and also the system configuration is specified. There is also a system pipeline given which gives a high level illustration of the whole system.

Provenance: No provenance information is provided, but intermediate results and the parameters to obtain those results are provided.

Reproducible rating: 4 out of 5 since most of the data, code and implementation details are provided. Most of the figures in the paper are reproducible if all the various test datasets were available. The figures 9 and few in 12 which were comparing other algorithm could not be reproduced because the code for those algorithms was not made available.

Experiments:

Initially when I fallowed the documentation and tried to execute the code, I got a few errors shown in the figure.



```
>> database = 'datasetSUN';
recogRate = objectRecognition(database,'KNNmethod','hog','nNeighbors',20,...
'nCandidates',9,'prior_weight',0.06,'spatial_weight',20,'alpha',0.7);
Database "datasetSUN": both training and test data are loaded!
Matrix dimensions must agree.

Error in objectRecognition (line 223)
            histogram = histogram + foo;

>>
```

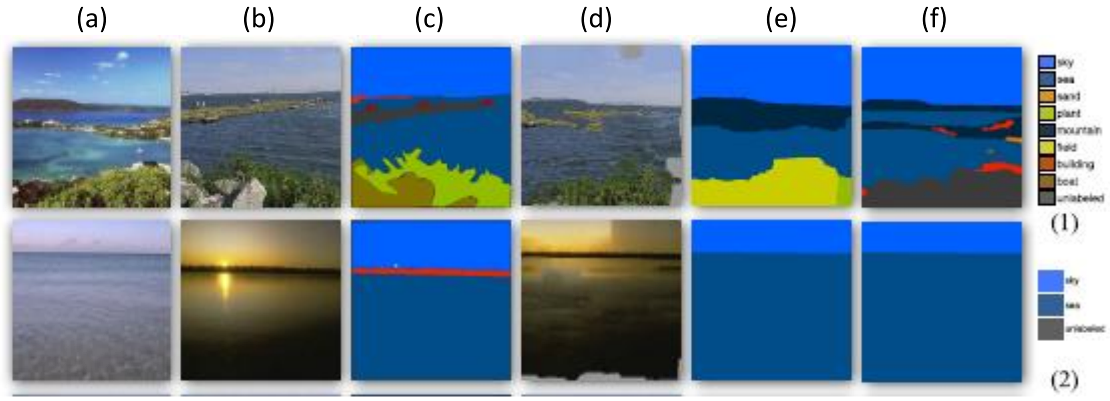
```
>> database = 'datasetLMO';
recogRate = objectRecognition(database,'KNNmethod','gist12','nNeighbors',85,...
'nCandidates',9,'prior_weight',0.06,'spatial_weight',20,'alpha',0.7);
Database "datasetLMO": both training and test data are loaded!
No. 1  K=85 Recognition rate: 0.883651   total: 0.883651
Error using griddedInterpolant
Sample values must be a single or double array.

Error in interp2>makegriddedinterp (line 225)
    F = griddedInterpolant(varargin{:});

Error in interp2 (line 136)
        F = makegriddedinterp(X, Y, V, method,extrap);

Error in warpImage (line 16)
    foo=interp2(xx,yy,im(:,:,1),XX,YY,'bicubic');

Error in objectRecognition (line 357)
        warpI2 = warpImage(im2,vx,vy)/255;

>>
```

Figure 1: Errors while running the code on MATLAB 2016 a.

There was information about the system configuration provided but no information about the Matlab version was mentioned. I later ran the code on MATLAB 2011 a, since the date of publish of the article was 2011 and the code ran.

Experiment 1:

The first experiment was to check if I could reproduce similar image labelling as shown in figure 10 of the paper. These are the images from the datasetLMO. The figure shows input images from the test dataset and then the labelled image is the output. Here (a) is the input image; (b) is the closest match; (c) is the annotation of the closest match; (d) is the warped version of (b); (e) is the annotated version of (a) obtained by the combination of m voting candidates; and (f) is the actual ground truth annotation of (a). This figure shows the higher recognition rates of the images. The images are picked randomly and for comparison I have put the images of the paper along with the reproduced images.



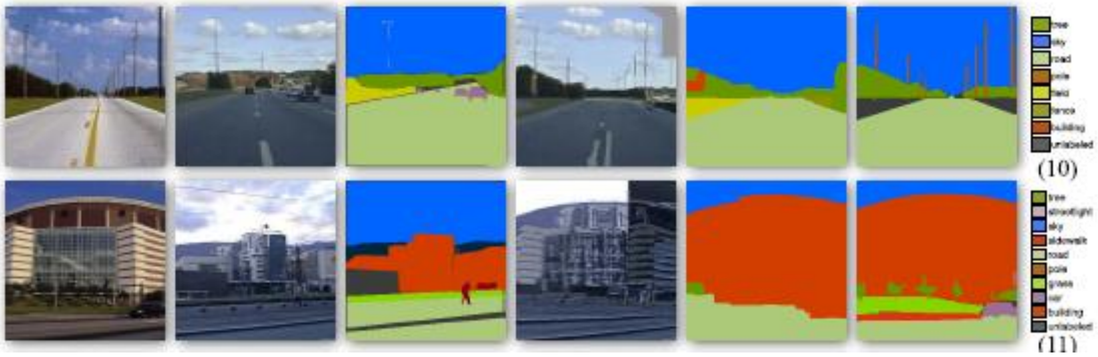


Figure 2: Images from figure 10 of the paper

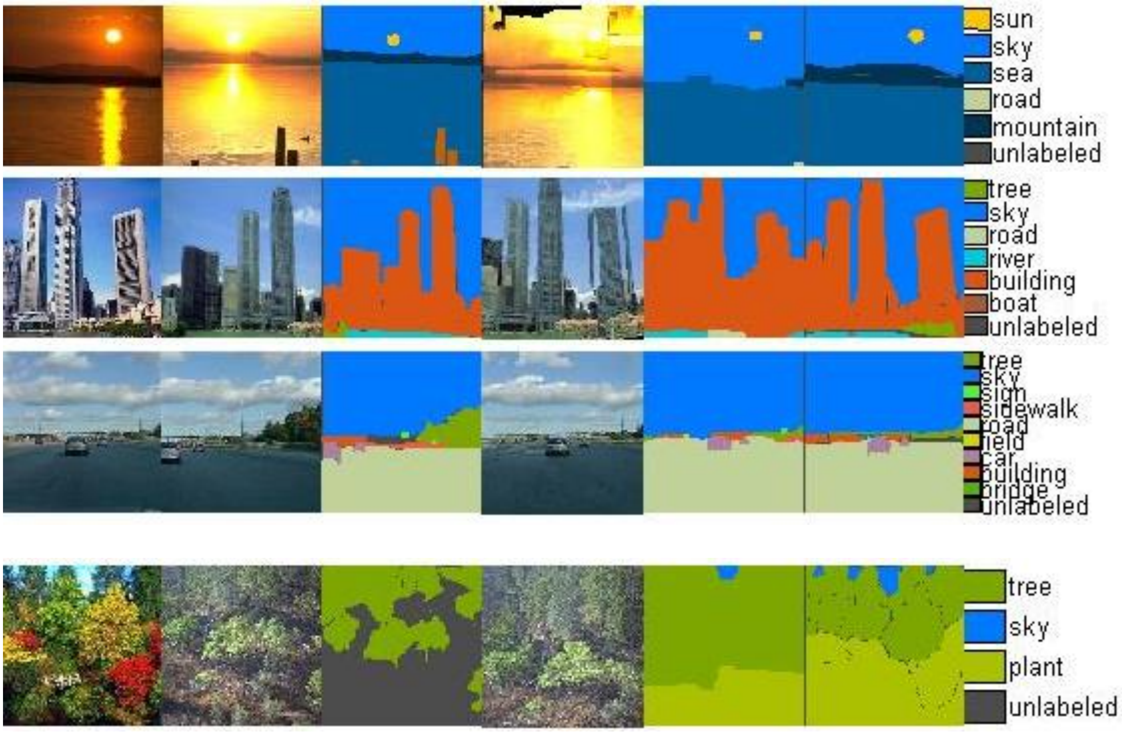


Figure 3: Images obtained on my system.
Similar results were obtained as mentioned on the paper.

Experiment 2:

Next I checked if the system generated mismatches and wrong labelling as shown in figure 11 of the paper. These images had bad recognition rates because the annotations were ambiguous and no good matches where found from the training dataset.

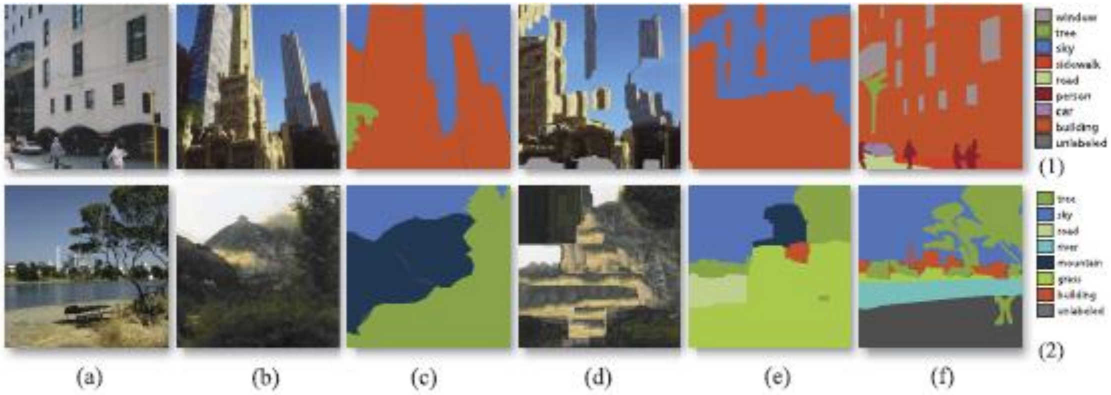
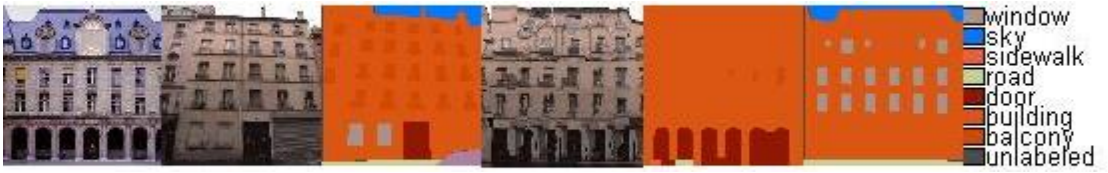


Figure 4: Images from the paper. Here (e) is the obtained labelling whereas (f) is the ground truth or actual annotation.



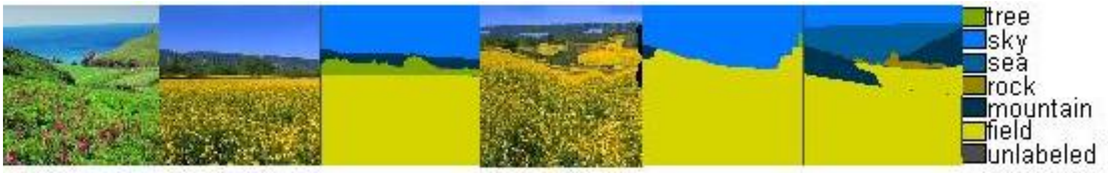


Figure 5: Images with bad Labelling on my system.

Experiment 3:

I was able to reproduce some of the figures in figure 12 of the paper which measured the performance of the system with different parameters and methods. All figures were not reproducible because the test dataset for each method had to be created and there were only 2 test datasets, one for "gistL2" and other for "all" that were made available for the dataLMO dataset. The method for creating other datasets is provided but because of the system requirements and the time required, the datasets could not be created. Figure 12a, 12d, and 12h were reproducible.

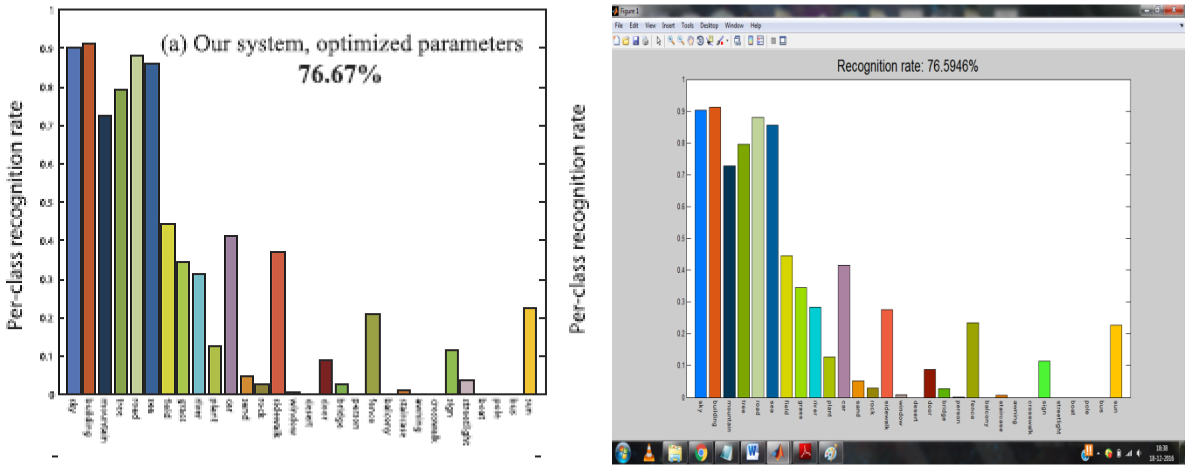
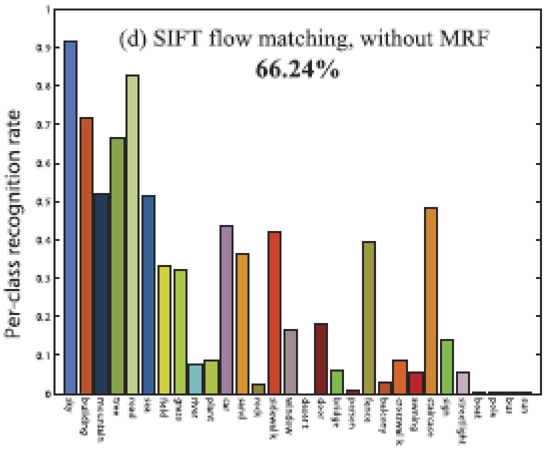


Figure 6: (12a) The highest recognition rate for optimal parameters mentioned in the paper 76.67% vs the performance on my system on the same parameters 76.5946%. The gistL2 method was applied here i.e L2 norm on GIST features. Almost similar results were obtained.



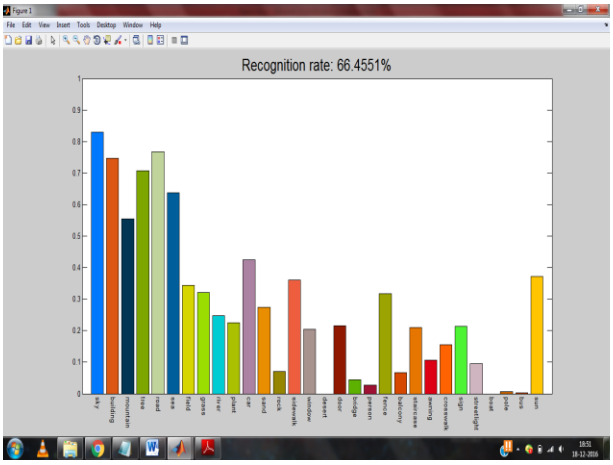
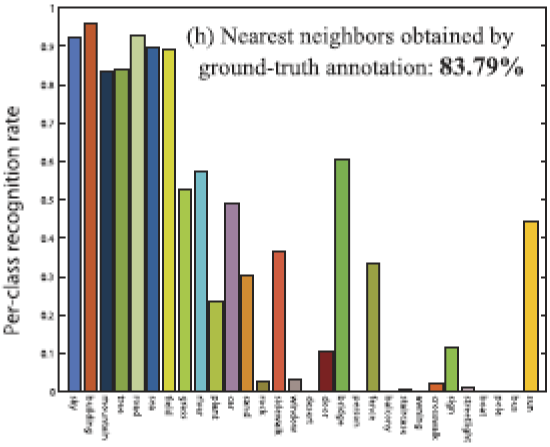


Figure 1
File Edit View Insert Tools Desktop Window Help
Recognition rate: 66.4551%

Figure 7: (12d) Here the same SIFTflow method is applied but without MRF i.e. the parameters α and β are 0. Here α and β represent the prior and spatial terms which are the Markov random field components. Results mentioned in the paper are 66.24% and results I obtained were 66.4551%. Slightly better results are obtained on my system.



8(a)

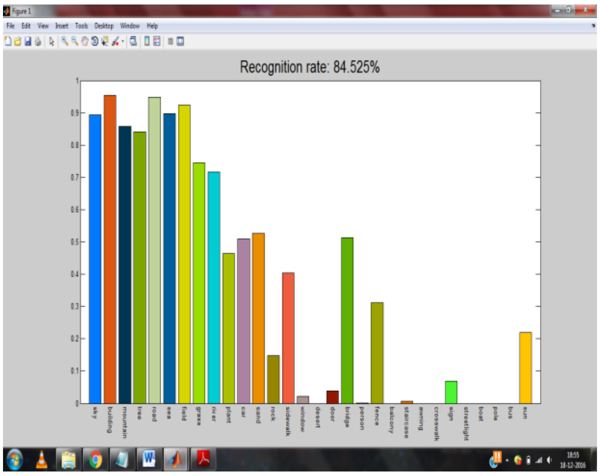


Figure 1
File Edit View Insert Tools Desktop Window Help
Recognition rate: 84.525%

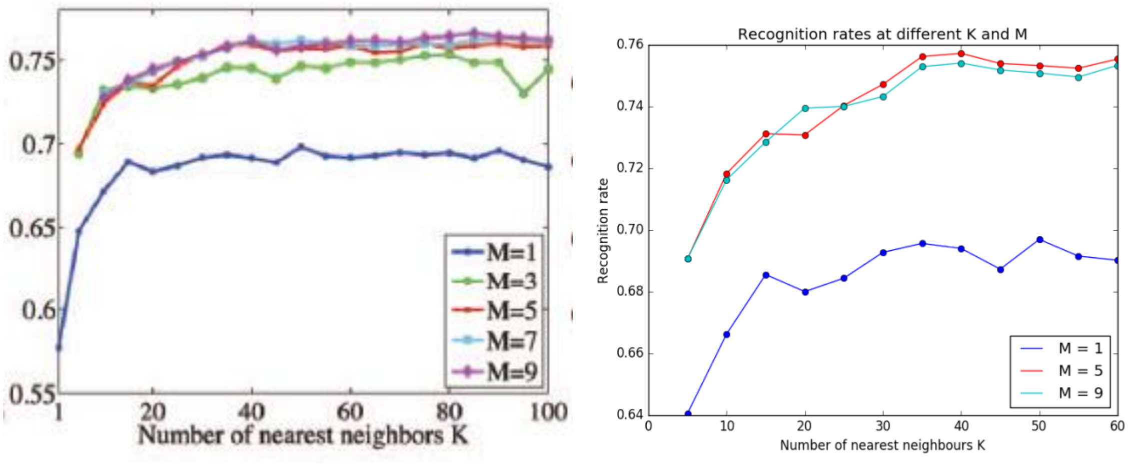
8(b)

Figure 8: Here 8(a) the upper limit of the system is shown i.e. the ground truth annotation and 8(b) shows the recognition rate obtained when KNNmethod 'all' is used, which is a combination of all the KNNmethods.

Experiment 4:

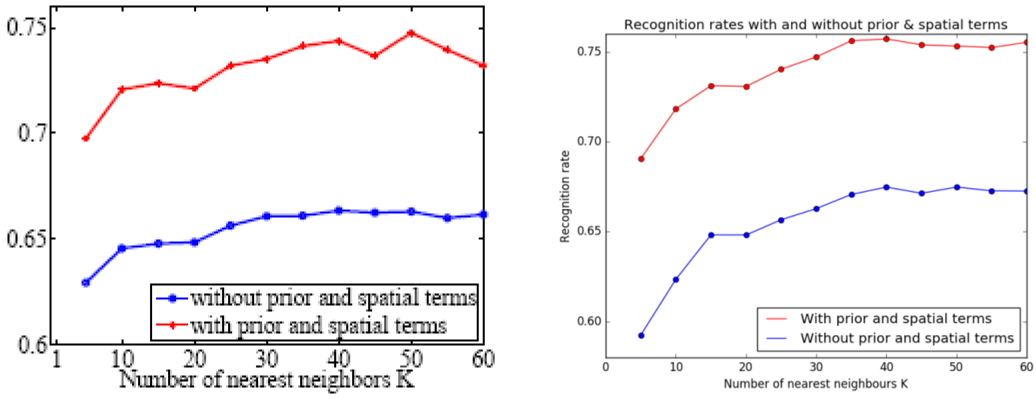
Here I tried to reproduce figure 12 from the paper. The figures here are trying to compare the recognition rates by varying the values of K and M i.e. the number of nearest neighbours and the number of voting candidates. There is another figure from their older paper which I was able to reproduce, which compared the recognition rates with and without MRF i.e. the prior and spatial terms.

Some other figures of figure 12 could not be reproduced because all the test datasets and code of other algorithms were not made available.



9(a) 9(b)

Figure 9: Each line in 9(a) shows the recognition rate by varying K at fixed values of M. I only tried for M = 1, 5, 9 and upto K=60 seen in 9(b) because of time constraints. Each iteration took more than 15 minutes. Similar results are seen. The highest recognition rate mentioned in the paper is for K = 50 and M = 5 and the highest I got was for K = 40 and M = 5.



10(a) 10(b)

Figure 10: Figure 10(a) is the one from the older paper comparing the recognition rates with and without the prior and spatial terms. And 10(b) is the figure I could reproduce. Overall the results are similar. There is a decline in the recognition rates by approximately 9 percent without the prior and spatial terms.

Figure which I could not reproduce:

This figure was not reproducible because there was no code available for [2] and I didn’t get any reply from the authors of either of the papers after emailing them for the code.

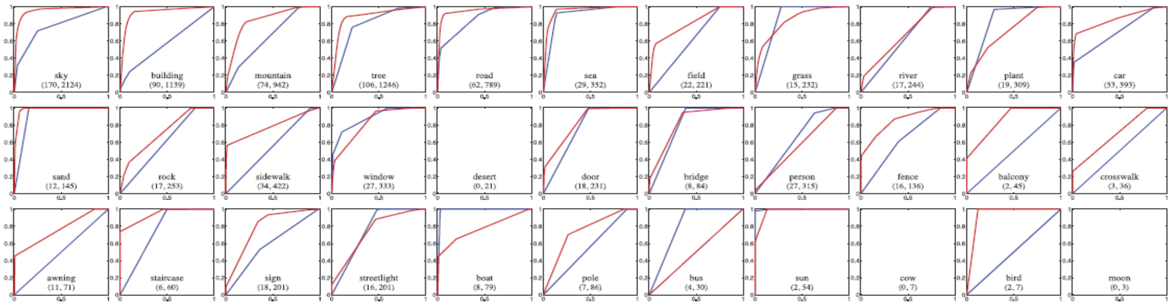


Figure 11: This figure compares each object recognition rates of the non-parametric approach of the paper vs the classier based system of human detection [2].

Conclusions:

I think proper documentation and availability of code and data contributed most towards reproducibility. Given the time and the required hardware, most of the datasets could have been created and tested for accuracy. After reproducing the figures whichever where possible for the time being, I think the paper has all the right results as most of the results were similar.

References:

- C. Liu, J. Yuen, A. Torralba, "Nonparametric scene parsing via label transfer", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2368, 2011
- C. Liu, J. Yuen, and A. Torralba, “Nonparametric Scene Parsing: Label Transfer via Dense Scene Alignment,” Proc. IEEE Conf.Computer Vision and Pattern Recognition, 2009.
- N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005.