Introduction

- The intention of the article is to find the semantic category of the images with the approach of bag of features.
- Kernel based recognition,a new approach is also introduced by the authors, which is an approximate technique of pyramid matching scheme.
- The new approach categorizes the images on global scale, which means that it analyses all the surrounding and background objects to define the category of the image.
- Histograms are done on the local features of the innumerably divided sub regions of each image and are compared to different classes of the images.

Previous works

- Many of the works on object recognition had been done through the Histograms. This method of approach can be used as alternative to the former.
- To be precise, some of the works like multi-resolution histograms had focused on varying the feature computation resolution by fixing the intensity scale. The authors worked antithesis, to increase the dimensionsional representation such that they store more information of the image.
- While in the histograms dividing the images into resolutions and disordering the objects had been done continuously, they have not given the optimal way of disordering and subdividing.
- So, the authors also noted the best way to subdivide and disorder through combination of different principles.

Content

- The procedure starts with dividing the each image into number of divisions, called as grids. The number of division of grids is based on the level of division.
- Pyramid matching, a comparison and approximation technique is adapted to make kernel-based recognition on the whole image.
- The authors mentioned that they ought to define the category of the image by even considering the neighbouring elements of the main object, by giving an example of probability of a car being on a highway is more than the toaster on the highway.

Certain methods are followed to attain the results through their approach.

- **Spatial Pyramid Matching:**
    - In this process, initially two sets of feature vectors are considered of two images.
    - If the cell results of different grids in different images are approximately equal, then the two grids are said to be matched.
    - Considering the different levels defined by 'l', $HX_l(i)$ & $HY_l(i)$ are the histograms for number of points that fall into X & Y sets at ith level.
- Spatial Matching scheme is a process of calculating the weighted histograms to evaluate the stability of matching of different images at the different levels.
- After calculating the histograms of the images, normalization is done by the total weights of all the features, for achieving maximum efficiency.
- **Feature Extraction:**

- Two types of features are considered in this context named as 'weak features' and 'strong features.'
  - If in any direction, the minimum threshold of the image is exceeded or the oriented edge points are considered, then they are considered as the 'weak features'
  - The SIFT descriptors of the image can be defined as the 'strong features'
- The experiments are repeated on various vocabulary sizes (16, 200, 400) for around ten times with differently selected random images.
- From all the individual runs the mean and the standard deviation are taken as the final results.
- Different datasets such as Caltech-101, Graz datasets are considered to evaluate the results of diverse images.
- By using a technique called LDA, they have decreased the dimensions of the feature sets.
- Finally on the reduced dimension set, SVM is trained and then they have classified the tests.

Results

- It can be observed that the results have been improved on the whole from the levels of zero to the multi-level divisions.
- The strong features have descended their accuracy along the levels two and three, because of the most fine divisions resulted in the very few matches.
- Comparatively the weak features had ascended through the increase in the levels because of their density and spatial expansion.
- It is also noted by the authors that the increase in the vocabulary had not resulted in much accuracy changes, whereas the increase in number of divisions had. So, they concluded that the grained geometric cues had more power of categorization than the increase in training vocabulary set.

Discussion and thoughts

- The testing is done by the batches of fifty to attain the effectiveness.
- Recognizing very thin objects and the textureless animals from the Caltech-101 dataset had not been very efficient by this approach.
- Infact had been much worse than many of the approaches, from which we can derive that the approach is not applicable to those type of objects.
- This approach can also be useful to assess the new datasets ad evaluating more sophisticated recognition approaches.

Conclusion

- It had been very elegant approach for image recognition by the approach of pyramid match kernels.
- It can be easily adaptable due to the fact that it worked fine with diverse datasets and can outperform many of the previous approaches such as bag of features with pairwise relationships.
- It can also surpass more practical approaches based on the parts and the relations.