Paper Summary: “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories”

This paper introduces a procedure for recognizing scene classified images using an approximate global geometric correspondence. Bag-of-features show quality performance in such tasks but disregard spatial information of the features making them incapable of separating foreground from background. Robust geometric correspondence based search approaches come with significant computational overhead, and methods based on increasing invariance levels of local features does not pay off either. The absence of efficient methods for this task is a major motivation for the development discussed in this paper. The proposed method is a computationally efficient extension of an order-less bag-of-features representation where the image is repeatedly subdivided into regions and local feature histograms are computed for every region. It is then followed by a kernel-based recognition which uses techniques adapted from an existing pyramid matching scheme. This approach can also classify images as containing specific objects in the presence of clutter and varying pose.

Early literatures show histograms as an important image descriptor which are generalized to locally order-less images. Multiresolution histograms computes histograms of fixed intensity scales over varied resolutions of image features in contrast to the proposed method here which fixes the feature resolutions and spatial aggregation is varied, thereby being a higher-dimensional representation that preserves more information. Pyramid matching then works by placing this sequence of grids, in increasing order of coarseness, over the feature space and taking the weighted sum of the number of matches that occur in every level. Matches at finer levels are weighted more than at coarser levels. A match happens if two points fall in the same cell of the grid. This is the method of Grauman and Darrell and is given an orthogonal perspective in this paper by first quantizing the feature space into discrete types and assuming that only features of the same type can be matched to each other. The resulting approach has the upper hand of maintaining continuity with the visual vocabulary paradigm. Finally, all histograms are normalized by the total weight of features in the image for maximum computational efficiency.

The experiments for evaluation of the system has been done on two kinds of features namely weak features, which are oriented edge points, and strong features, which are SIFT descriptors. The grid is dense with a spacing of 8 pixels because of Fei-Fei and Perona’s evaluation that dense features work better than sparse interest points for scene classification. A visual vocabulary is trained by k-means clustering. Three different datasets have been used to test the performance. All evaluations are repeated ten times and the results reported are the mean and standard deviations from these individual evaluations. For the first dataset, performance on strong features results in a classification rate of 72.2% which drops down to 63.36% if latent factor analysis techniques are applied. Results of the spatial pyramid matching improves as grid levels are increased. It was also observed that this method is robust to failures at individual levels as even when a higher level is too finely subdivided, its performance essentially remains similar to its previous level. However, this method has a disadvantage when it comes to classes characterized by high geometric variability as finding significant global features become difficult then.

The paper is short, crisp and has conveyed its idea aptly. It has cleverly made use of results and techniques of previous literatures to identify advantages and disadvantages of its own proposed method. Figures demonstrating the orthogonal pyramid matching scheme could have helped. With a very simple and powerful approach, this work has achieved a great deal in showing the significance of global non-invariant representations for scene classification of images.