Computer Vision
Homework 09: Visual Feature Descriptors
CS 670, Fall 2019

Name: Kunjal Panchal
Student ID: 32126469
Email: kpanchal@umass.edu

November 16, 2019

Contents

1 Problem 1

HISTOGRAM OF ORIENTED GRADIENTS is a feature description technique which counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

1.1 Tasks where HoG can be useful

1. The application which made HoG technique mainstream is Human(Pedestrian) Detection in static images; this task was expanded to common animals and vehicles detection in static images too.

2. Recognizing three-dimensional objects in complex scenes like, higher-level representations in STRONG GEOMETRIC MODELS are computed from the 3D data in the form of surface features, parametric surface patches, or representations of parts, such as generalized cylinders and are then compared to similar representations for the shape models. Here, shape recognition techniques like HoG can accurately extract features while preserving overall relative structure of the image.

3. In mobile robotic manipulation, it need to reliably recognize objects over hundreds of images, and to compute their positions accurately enough. In addition, because this system is to be used by workers unfamiliar with the underlying technology, a high level of recognition accuracy is needed, the HoG technique can be used here. Recognition is performed on outdoor scenery with little control over the illumination and viewpoints. HoG is invariant to lights and small deformation, making it an ideal candidate.

1.2 Tasks where HoG cannot be useful

HoG doesn't recognize textures. As its implementation depends on gradients which aren't much use in detecting texture, but only for the detection of edges.

1. In medical and/or biological images; segmentation of specific anatomical structures and the detection of lesions, to differentiation between pathological and healthy tissue in different organs. HoG doesn't work for object recognition where recognition depends on different textures.

2. Many applications such as image compression, pre-processing or segmentation require some information from the regions composing an image. Each region is characterized in terms of homogeneity (region composed with the same grey-level or a single texture) and its type (textured or uniform). In tasks where the decision criterion is based on the use of classical texture attributes (co-occurrence matrix and gray-levels moments), HoG cannot be used.

2 Problem 2

The first need of Pedestrian/Human Detection is a robust feature set that allows the human form to be discriminated cleanly, even in cluttered backgrounds under DIFFICULT ILLUMINATION.

2.1 Invariance to Lighting

For better invariance to illumination, shadowing, etc., HoG is uses contrast-normalization of the local responses before using them. This can be done by accumulating a measure of local histogram "energy" over somewhat larger spatial regions ("blocks") and using the results to normalize all of the cells in the block.

It takes logarithm of intensity values of illuminated scene, it normalizes gamma and colors. Then it contrast-normalizes the overlapping spatial blocks.

Gradient strengths vary over a wide range owing to local variations in illumination and foreground-background contrast, so effective local contrast normalization turns out to be essential for good performance.

2.2 Invariance to Small Deformations

The HOG descriptor is a concatenated vector of the components of the normalized cell histograms from all of the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor. Two main block geometries exist: RECTANGULAR R-HOG BLOCKS and CIRCULAR C-HOG BLOCKS.

Shape Contexts can be found with binary edge-presence voting into log-polar spaced bins, irrespective of edge orientation. The C-HOG descriptor simulates this edge and shape orientation invariance with just 1 orientation bin.

For good performance, one should use fine scale derivatives (essentially no smoothing), many orientation bins, and moderately sized, strongly normalized, overlapping descriptor blocks.

For SMALLER BINS, each pixel calculates a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centered on it, and the votes are accumulated into orientation bins over local spatial regions. Thus, it gives us better spatial resolution.

But LARGER BINS, give us more information about the gradient direction and magnitude; which is useful when we are observing a non-rigid, shape-shifting, deformable object as the the large number of gradients will capture the deformation knowledge.

So, a moving object or deformed object will still be detected by its same feature with large bins as the bins will capture more variance of gradients.

Thus, HoG achieve invariance small deformations with larger bin size in spatial/orientation binning.

References

[1] Subhransu Maji, *Recognition* https://www.dropbox.com/s/cjl79dzn1cavfzn/lec15_recognition.pdf?dl=0