Project Proposal

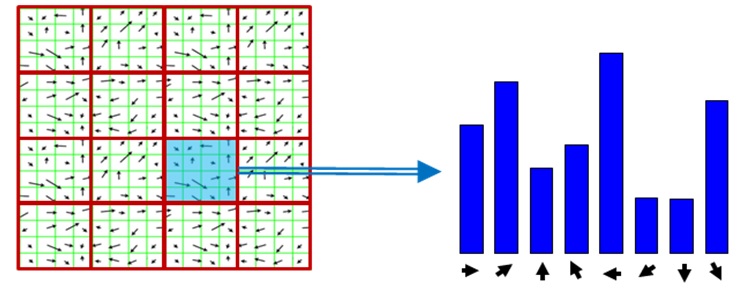
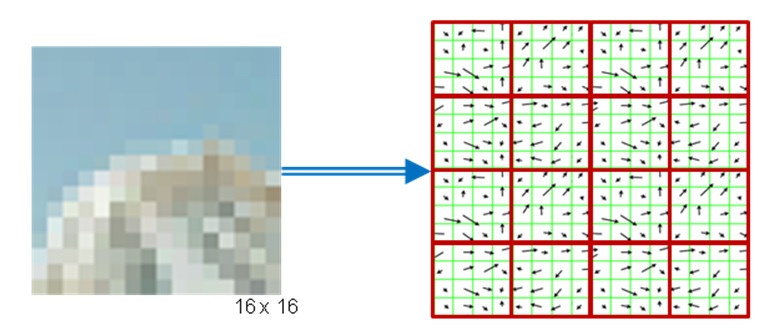
Jicheng Gong 325005650

1. Introduction and Motivation

In this course project, I want to try to use machine learning method to reduce dimension of SIFT descriptors so that I can optimize its efficiency when SIFT is implemented on some real-time measuring system.

1. SIFT algorithm

SIFT (Scale invariant feature transform) is an algorithm vastly used in field of computer vision to detect local features in images. It provides relatively robust result for object matching and thus improved robotic mapping, image stitching and 3D modeling, etc. A SIFT keypoint descriptor is created from a 16×16 neighborhood area whose size is determined by the feature scale. For example, it will create a larger window for large features and smaller window for small features. In the 16 by 16 region, we set each 4 by 4 as a sub-region and calculate its 8 bins orientation histogram. Thus we have a 4×4×8=128 dimension vector for each keypoint.



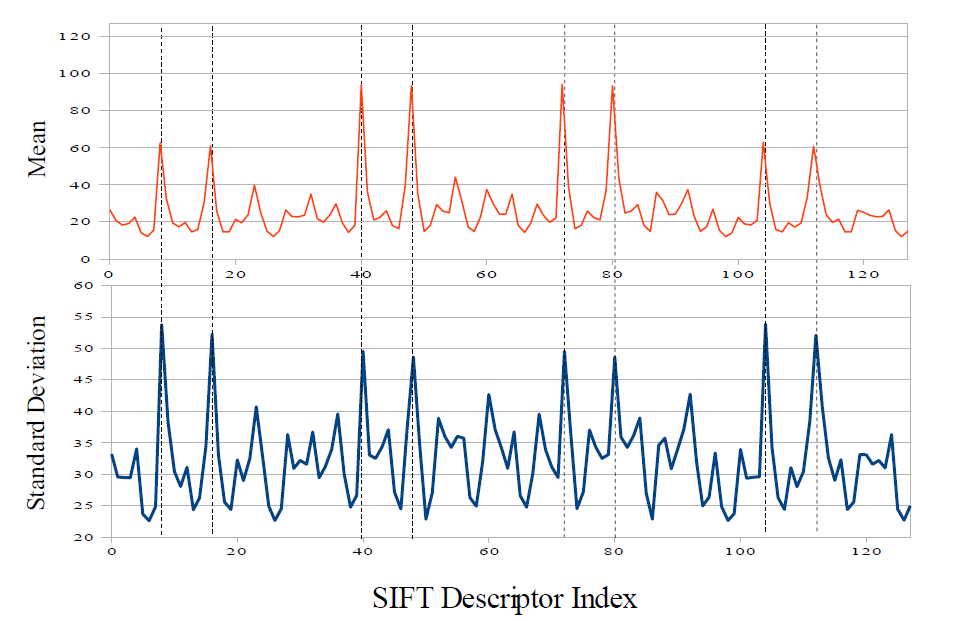
1. SIFT matching

To determine a match pair of points, the algorithm calculates the Euclidean distance between two keypoint descriptor vectors. The smallest distance will be regarded as the potential match point (actually it will be a little bit more complicated). However, since each image may contain more than 10,000 keypoints and Euclidean distance is computation expensive, the SIFT algorithm is typically very slow.

1. Improvement to be made

One improvement is PCA-SIFT which can effectively reduce the dimension. However, since PCA needs training large volume of data, it’s not appropriate in real-time system.

Some researcher found that some of the dimensions have a larger mean value and variation. If only some of the dimensions are chosen to do the matching calculation, we can still get a good accuracy.



However, according to my experiment, although some dimensions have larger mean value, different group of pictures have different large dimension component. So we may need a more general method to look for large elements or combination of them for each circumstance.

1. Deliverable

Find which dimensions or combination of dimensions have larger contribution to matching process and can retain the matching accuracy. Given the dominating dimensions, find the relationship between correspondences and try to explain why they have larger value

1. Methodology

I plan to use some neural network knowledge to see whether I can train a network to do feature selection. Then I want to test the matching accuracy for the low dimension descriptors. Finally since the calculation of neural network also takes time, I need to test the efficiency of the neural network.

1. Resources

Most of the data I need is images. Since SIFT is a well-built algorithm, it is not easy to reduce its dimension while remain good accuracy in most general cases. I may first do the training mainly on autonomous driving vehicles to get a good result then try other image data sets to see whether it can remain the accuracy. The other resource I use is the SIFT algorithm. My algorithm is an open source version called ”ezSIFT”.

1. Progress schedule

Since I don’t have too much previous knowledge of neural network, I’m only able to figure out a very general schedule

1. Generate large number of keypoint descriptors from different image sets. Try different neural network algorithm to look for a best one.
2. Try optimization of the selected model and test the efficiency (not include the training time, I may need a pre-trained network) and accuracy.
3. Find the underlying principle such as why certain dimensions are important.
4. References
5. David G. Lowe 2004 Distinctive Image Features from Scale-Invariant Keypoints.
6. Geoffrey Treen 2009 Methods for Improved SIFT Matching.
7. I.T. Joliffe. Principal Component Analysis. Springer-Verlag, 1986.
8. Herbert Bay, Tinne Tutelaars 2009 Robust Interest Point Detector and Descriptor.