EE368 Project Proposal

Stereo Depth Maps

Matt Stevens (mslf@stanford.edu)
Zuozhen Liu (zliu2@stanford.edu)

Introduction

Given two stereo images of a scene, it is possible to recover a 3D understanding of the scene. This is the primary way that the human visual system estimates depth. This process is useful in applications like robotics, where depth sensors may be expensive but a pair of cameras is relatively cheap. Computing depth maps from stereo images is an old problem in image processing and computer vision, but it is still an area of active research.

Previous Work

Algorithms for computing depth maps typically follow the same pipeline. First, there is an error function to determine how different two potential matching pixels are. Then, there is an aggregation function that looks at these differences over a local neighborhood. Lastly, there is some kind of optimization routine that looks for the best pixel matches given these error values [2]. Based on the optimization and aggregation routines, an algorithm can perform local optimization over a region, or global optimization on the whole image. There is a tradeoff between the two, with global methods giving cleaner results, but being more computationally expensive and ill-suited for real-time applications. There have been dozens of different approaches to these subproblems over the years. We will be focusing on one of these methods as a promising

Approach

We will implement a naive baseline algorithm as a point of comparison for more sophisticated algorithms. This approach will calculate the normalized cross-correlation with a sliding window, and take the maximum correlation for each pixel in a “winner-take-all” approach [2]. This local approach is highly prone to noise but gives very fast results. These results can be filtered with simple methods such as a median filter to remove noise and comparing left-to-right and right-to-left matches to check for consistency.

We will also implement a global method by using a graph-cut based optimization algorithm[3]. This approach minimizes a global energy function that encodes both the quality of matches between pixels and the regularity of depth in a local neighborhood. By minimizing the global energy function, this algorithm is able to grasp better context of the image and is, therefore, less

prone to local noise. However, as we increase the level of depth granularity, the algorithm may take quite long to converge.

If time permits, we may also work on some extensions to this project. We will potentially investigate dynamic programming solutions [4] as a middle ground between speed and accuracy, which could possibly be used for applications in mobile/embedded environments. We will also investigate performance optimizations in the graph cut procedure such as the LogCut algorithm, which has orders of magnitude of speed improvement over standard graph cut algorithms when using many labels [1].

We will be using a stereo image dataset from Middlebury College that has stereo pairs along with ground truth disparity maps [5]. This will allow us to directly evaluate and compare the performance of our different algorithms. For each algorithm we will evaluate both the runtime and the quality of the disparity maps produced.

References

[1] Lempitsky V, Rother C, Blake A (2007) Logcut—efficient graph cut optimization for Markov random fields. In: ICCV

[2] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” Int’l J. Computer Vision, 2002.

[3] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In ICCV, volume II, pages 508–515, 2001.

[4] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum likelihood stereo algorithm. CVIU, 63(3):542–567, 1996.

[5] Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., & Westling, P. (n.d.). High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. Lecture Notes in Computer Science Pattern Recognition, 31-42.