Image Segmentation by Total Variation Methods

Dennis Sun
Stanford University
Dept. of Statistics
dlsun@stanford.edu

Matthew Ho
Stanford University
Dept. of Electrical Engineering
matthew.ho@stanford.edu

1. Introduction

Total-variation denoising (TVD) is a robust algorithm for reconstructing noisy images. Because it imposes an L1 penalty on differences between adjacent pixels, it tends to result in images with piecewise constant regions. The goal of image segmentation is to assign pixels to clusters such that pixels within each cluster are similar. Our goal is to see if TVD can be used to perform segmentation, by taking the constant regions of the denoised image as the segments. We hope that this project can yield insight into image segmentation, which is a central problem in computer vision and is an intermediate step in many higher-order tasks, such as object recognition.

2. Data

We would like to develop our algorithm for grayscale and color images, and if time permits, RGB-Z images. For grayscale and color images, we will use the Berkeley Segmentation Dataset. For RGB-Z images, we will use images that we will collect around campus using Kinect. (We would appreciate it if we could use images collected from past student projects.)

3. Methods

In the original paper by Rudin, Olsher, and Fatemi, TVD was applied to grayscale images. We intend to make use of existing TVD scripts in Matlab and apply them first to grayscale images, where "piecewise constant" is most clearly defined. Although TVD has also been applied to RGB images, it is less clear how we will search for "piecewise constant" regions, since it is unlikely that the reconstructed image will be piecewise constant in all three coordinates. We will implement and test a variety of methods, such as majority voting (if two of three coordinates are piecewise constant) and thresholding (if the L1 norm of the channel differences is less than some ϵ). There are also techniques to scalarize RGB data, as decribed in Ivanovska. Finally, if time permits, we would like to try TVD on RGB-Z images, combining the channels in a similar fashion.

Figure 1. Buffalo image from the Berkeley Segmentation Dataset, original and smoothed using total variation denoising.

4. Results

As expected, we found that total variation denoising gave us smooth regions while preserving edges. (See Figure 1.) However, TVD did not give us exactly smooth regions, most likely because of the numerical tolerance of the optimization algorithm and because TVD it too local (it only constrains directly adjacent pixels). Therefore, some postprocessing was necessary to obtain the final clusters.

Since the convexity of TVD is one of its primary appeals (it is guaranteed to give the same answer every time), we wanted to avoid nonconvex postprocessing algorithms such as k-means, mixture of Gaussians, and normalized cuts. We initially tried applying the watershed algorithm to the gradient, in the hopes that the smooth regions would correspond to troughs in the gradient. However, the gradient had too many local minima, so the watershed algorithm resulted in an oversegmented image. Nevertheless, applying TVD first significantly reduced the oversegmentation problem of the watershed algorithm, as shown in Figure 2. One direction we could pursue is to successively merge clusters until we obtain the desired number.

In order to have some preliminary benchmarks for testing TVD, we decided to try k-means and normalized cuts as a postprocessing step. Visually, applying TVD first before applying naive k-means (clustering based on pixel intensity and location only) gives a much more acceptable segmentation, as shown in Figure 3. This is confirmed numerically

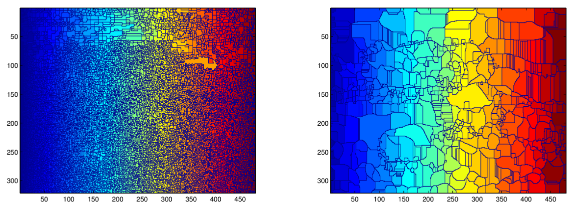

Figure 2. The watershed algorithm as applied to the original and to the TVD smoothed buffalo image.

Figure 3. The top two images show the resulting segmentation after applying k-means directly to the original image and to the TVD smoothed image. The latter is much more visually appealing. To show that arbitrary smoothing will not necessarily solve the problem, the bottom left image depicts the results from applying a Gaussian filter to the image. Finally, the bottom right image depicts the ground truth (24 segments) as labeled by a human.

in the Evaluation section below. TVD did not seem to confer any advantage when we used normalized cuts, however. The results are not displayed here.

5. Evaluation

One way of evaluating segmentations is to do visual comparisons with the human eye. However, it is also desirable to come up with quantitative measures of performance.

5.1. Region Matching

Region Matching is a simple method for quantifying how closely a generated segmentation S_1 matches a ground truth segmentation S_2. Intuitively, it measures how the consistent overlapping regions are between S_1 and S_2. It is described in Martin et al.

We calculate local refinement error for each pixel:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \backslash R(S_2, p_i)|}{|R(S_1, p_i)|}$$

then compute two quantities, GCE (Global Consistency Error) and LCE (Local Consistency Error).

$$GCE(S_1, S_2) = \frac{1}{n} \min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\}$$

$$LCE(S_1, S_2) = \frac{1}{n} \sum_i \min \left\{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \right\}$$

As both GCE and LCE account for refinement, they are only meaningful when the number of segments for S_1 and S_2 are approximately the same. Thus when we use these two measures as a benchmark, we supply the K-means algorithm with a pre-defined k.

5.2. Empirical Results

We ran our TVD segmentation on the bison image from the Berkeley Segmentation datset. We compared the results of K-means segmentation $k = 24$ on the raw image and K-means segmentation on the smoothed TVD output to the human-generated ground truth segmentation wtih 24 regions. The results are compiled below.

	GCE	LCE
raw image	0.2419	0.2053
Gaussian-filtered	0.2290	0.1760
TVD	0.2187	0.1744

We see that both measures improve after using TVD as a preprocessing step.

6. Further Directions

In the course of testing out total variation denoising, we have found that TVD is highly sensitive to changes in illumination, although it is fairly robust to affine transformations. This limitation means that TVD alone cannot achieve state-of-the-art performance. Nevertheless, we are interested in pushing the algorithm to its limits for this segmentation task and plan to continue to explore postprocessing algorithms and to extend the algorithm to RGB data.

However, TVD has the potential to achieve state-of-the-art performance as a preprocessing filter because of its ability to smooth an image while preserving edges. This would be useful in edge detection algorithms such as the Canny edge detector, which typically uses a Gaussian filter for noise reduction. However, this also dulls the edges. TVD allows one to achieve the best of both worlds.

We also want to look into other ways of quantitatively evaluating segmentation results, as Region Matching, although simple and intuitive, has its weaknesses.

7. Background

L. Rudin, S. Osher, E. Fatemi. "Nonlinear Total Variation based noise removal algorithms". *Physica D*. 60: 259-268. 1992.

P. Blomgren, T. Chan. "Color TV:Total Variation Methods for Restoration of Vector Valued Images (1996)". IEEE Trans. Image Processing (1996)

T. Chan, S. Kang, J. Shen, Total Variation Denoising and Enhancement of Color Images Based on the CB and HSV Color Models, Journal of Visual Communication and Image Representation, Volume 12, Issue 4, December 2001,

T. Ivanovska, L. Linson. "Converting RGB Volume Data to Scalar Fields for Segmentation Purposes".

Berkeley Vision Group. "The Berkeley Segmentation Dataset and Benchmark".