Spectral regularization in computer vision

Neil Birkbeck

March 31, 2010

1 Introduction

I attended a talk yesterday given by Robert Tibshirani, who talked mostly about the Lasso. It was an exciting talk. One part in particular discussed the use of spectral regularization for matrix completion. Both my supervisor and I thought about the related applications in computer vision. This document is meant as a rough outline of some crude experiments I performed using this technique for both appearance modeling and orthographic structure and motion. I have yet to search the literature for other methods in vision that use this technique. So these results should only be interpreted for personal interest.

2 Appearance Modeling

In our research we have often treated entire texture maps as observations and used PCA to reduce the dimensionality to a handful of basis images. In the case of solid 3D objects, not all surfaces are visible in every input image. Previously, we dealt with this by just taking the mean of the entries where a texel was visible, and then filled in the occluded regions with the mean (this was done so as to not introduce any unneeded variation).

The spectral regularization of Mazumder, Hastie and Tibshirani offered one way to get a better estimate for these values. Their SOFT-IMPUTE algorithm seeks to find a full matrix $\mathbf{Z}$ that minimizes:

$$\frac{1}{2} \sum_{(i,j)\in\Omega} |\mathbf{Z}(i,j) - \mathbf{X}(i,j)|^2 + \lambda||\mathbf{Z}||_*,$$

where Ω is the indices of the known (or visible) elements. The second term penalizes the nuclear norm of $\mathbf{Z}$, which is the sum of absolute values of the singular values of $\mathbf{Z}$. The solution is obtained for a set of λ's, and the best solution is taken (e.g., by cross-validation). In our case we know that our input observations have some constraint on the rank. So instead, we use their algorithm to find the smallest λ that gives us a matrix $\mathbf{Z}$ with our desired rank.

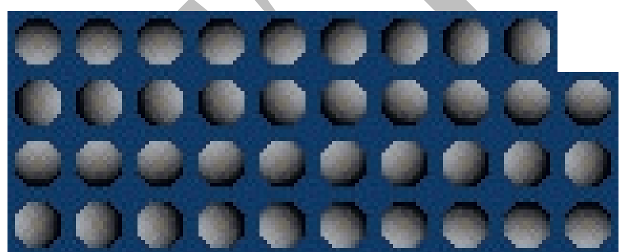

Figure 1: The input sphere sequence.

To fill in missing values, we can simply grab the corresponding elements of $\mathbf{Z}$ or even replace $\mathbf{X}$ with $\mathbf{Z}$.

2.1 Testing

To test this method, I have a number of rendered views of a lambertian sphere with different illumination (Fig. 1). Without taking into account shadows, the sphere should be representable by 3 basis images.

Figures 2 through 4 illustrate the results where 90%, 70% and 60% of the elements were randomly chosen to be visible. In each of the figures the left image illustrates what the sample input (with non visible pixels marked) looks like. The middle image shows the results if the missing values were just filled in with the mean. And the right shows the results from the spectral regularized version of the matrix. The bottom illustrates what the basis elements would look like if we were to do PCA on the images. Notice that in each case the two basis elements for the spectral reconstructed images are less noisy, and hence more like the true values.

As the random visibility is not really what would happen in the real textures, I have generated another example where the visibility is correlated to the darkness (somewhat more similar to what would happen when texturing a real object). Fig. 5 shows the results. In this case the regularized reconstruction still looks better, but does seem to have a bit more noise in both the reconstruction and the basis images.

Table 1 shows the amount of variation accounted for by the first two principle components. Notice that the first two components always better model the data in the spectral method. Furthermore, notice how the amount of variation accounted for by the spectral method is far less sensitive to the amount of visible pixels. This suggests that the spectral reconstructed data would be better to use than a simple replication of the mean.

3 Orthographic Camera Factorization

A similar problem occurs in the case of structure and motion factorization. In this case the observation matrix is $2 * ncams \times npoints$ and the missing data corresponds to occluded points. For orthographic cameras, the camera translation can be taken as the object centroid, and after this has been subtracted, the measurement matrix should be rank 3; the components are the x-y axes of the cameras and the 3D locations of the points.

To test out the spectral factorization in this context synthetic correspondences of a 3D cube were generated by a number of cameras. Additional random points were scattered throughout the cube, and the visibility of the points was varied randomly.

Before doing the Tomasi style factorization, the missing data was filled in by the spectral clustering. In this case, we again chose the λ that gave us a particular rank. And we iteratively reduced the rank from 6 to 3 using the intermediate results to update the mean of the points (for camera centers), which in turn was used to normalize measurement matrix.

The points of the cube were brought into alignment after reconstruction and the residual was recorded. For all of the sequences only 10 cameras were used. Unfortunately, the gains of the spectral regularization are not as evident in this case. Although the technique appeared to work for a small number of occlusions, but even in the case of no noise it didn't do near as well as the subsequent bundle adjustment (e.g., Fig. 6).

The story was similar with noise (Fig. 7). The method was capable of coping with noise, but bundle adjustment still did far better. With so few points and cameras the results of the factorization started to noticeably degrade. With more random points (e.g., 100s or thousands) it was possible to increase the occlusion fraction up to about 70 %, but the results were again not as accurate as the final bundle adjustment. The good news is that the filled in data combined with factorization was good enough to use in bundle adjustment. However, notice how the factorization method performs equally well as bundle adjustment when there is little occlusions. It would be nice if the factorization was capable of doing this for larger numbers of occlusions.

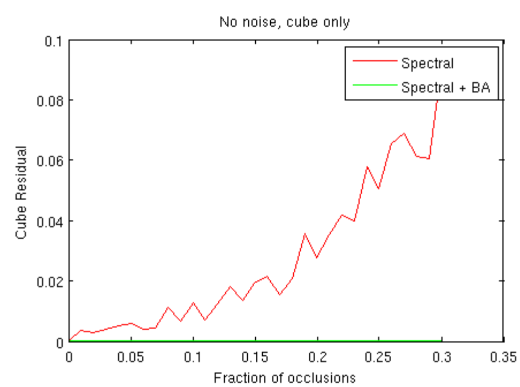

Figure 6: Even with no noise the spectral regularization didn't seem to give as good as results as bundle adjustment. The results were okay with a small number of occlusions, but steadily degraded.

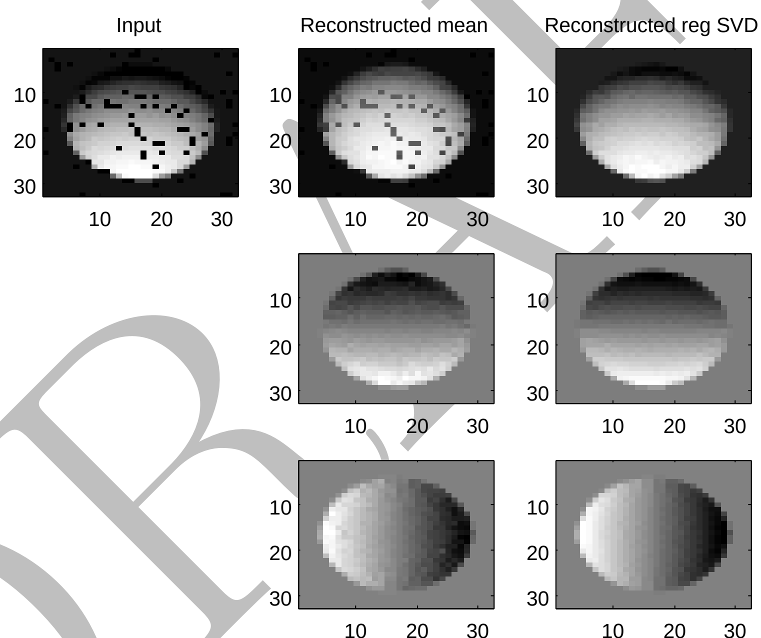

Figure 2: Reconstruction results with 90% of the elements being observed.

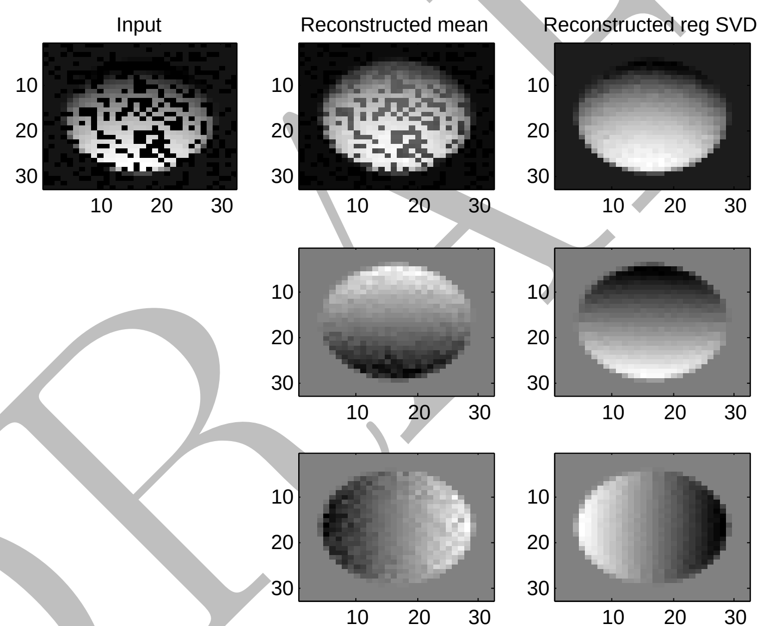

Figure 3: Reconstruction results with 70% of the elements being observed.

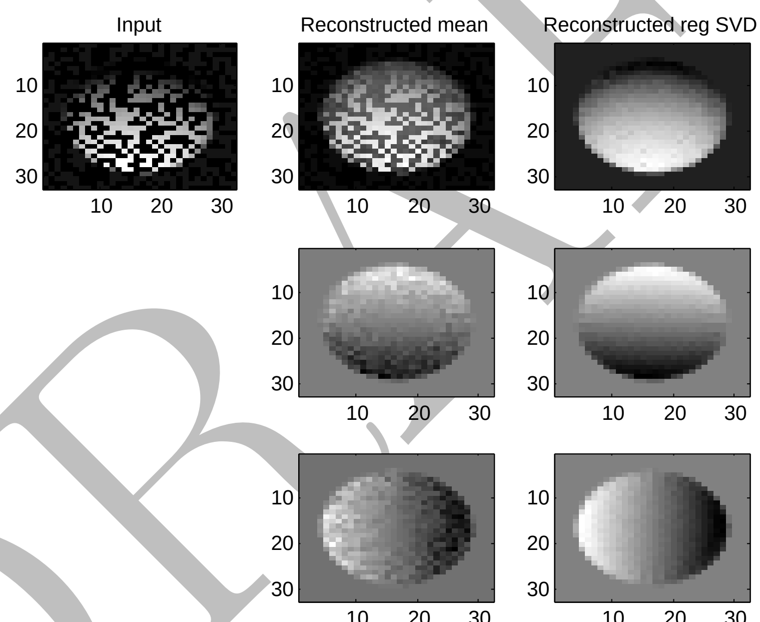

Figure 4: Reconstruction results with 60% of the elements being observed.

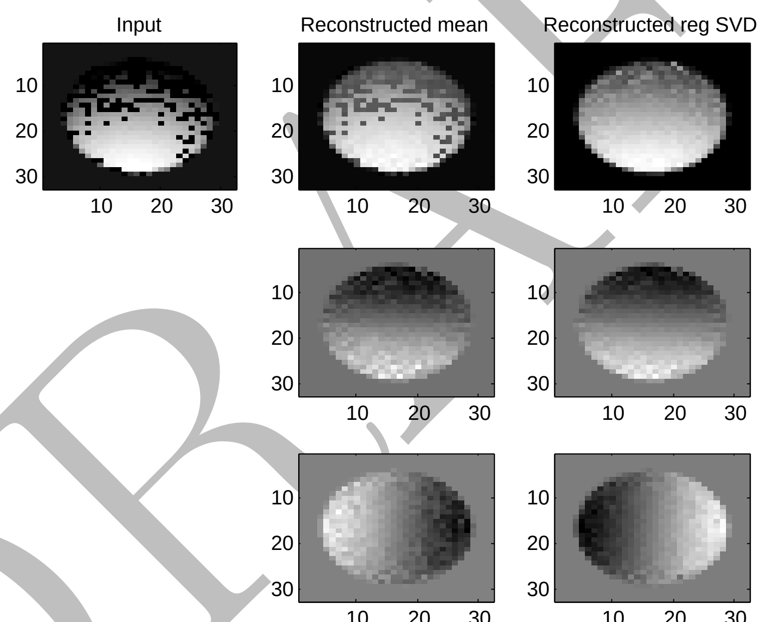

Figure 5: Reconstruction with the visibility simulation occlusion (mostly dark regions are not visible).

	90%		70%		60%		Sim occ.	
Mean inputed	0.499	0.403	0.39	0.31	0.29	0.22	0.33	0.29
PCA inputed	0.548	0.46	0.547	0.44	0.54	0.44	0.478	0.40

Table 1: Amount of variation accounted by the first two principle components (for each of the two input types). Notice that the spectral regularized version (e.g., the PCA inputed) accounts for more variation, especially in the presence of more noise.

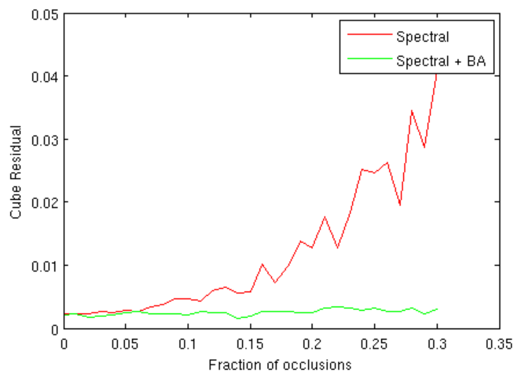

Figure 7: Using a few extra points (4), with noise at 2%, the reconstruction steadily degraded with more occlusions. The residual measures the 3D distance on the unit cube (after it was aligned with an affine transform). However, the solution from factorization was still good enough to use in bundle adjustment

In order to debug why the factorization wasn't working, I thought maybe either the mean normalization or that the true observation matrix didn't have a minimal nuclear norm. To test these hypothesis, I checked the singular values of both the initial observation matrix (with no noise and 30% occlusions), and the values of the spectral factored matrix.

For the rank 4 approximation, the singular values for a particular example were:

Observation matrix	9.888	8.139	6.677	3.80
Spectral factored	9.210	7.523	5.953	2.67

And after the rank 3 (mean normalized approximation), using both the soft spectral regularization and the hard regularization (using the soft regularization as initialization):

Observation matrix	9.811	7.04	6.676	0
Soft factored	8.891	5.97	5.755	0
Hard factored	9.811	7.04	6.676	1.03

It seems that the reconstructed rank typically closely resembled the true rank. This didn't seem like the problem, yet the reconstruction results (e.g. Fig. 8)

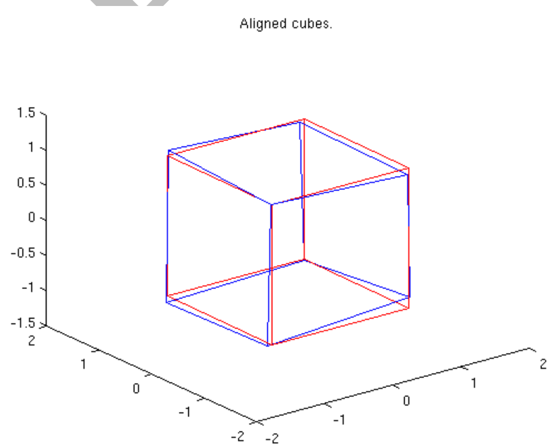

Figure 8: The results just using the spectral factorization didn't fully agree with the ground truth when there were occlusions, even in the case of 0 noise.

Instead of using only soft or hard thresholds used by the spectral projection algorithms, I experimented with using the orthographic camera factorization routine to restrict the rank. Starting with the results from the above spectral factorization, through this iterative approach it was possible to get results that approached bundle adjustment. Again, comparing the reconstruction residuals of the cube using 4 extra points, no noise, and 10 cameras, Figure 9 shows the results of this technique compare to bundle adjustment. The results are similar, with both methods having some failure in the larger percentage of occlusions. For the most part this seemed less sensitive to the number of occlusions, especially when there were more points. Again, this result is more what is expected for filling in the data. Unfortunately, for the most part several iterations were required, and there is no guarantee of convergence.

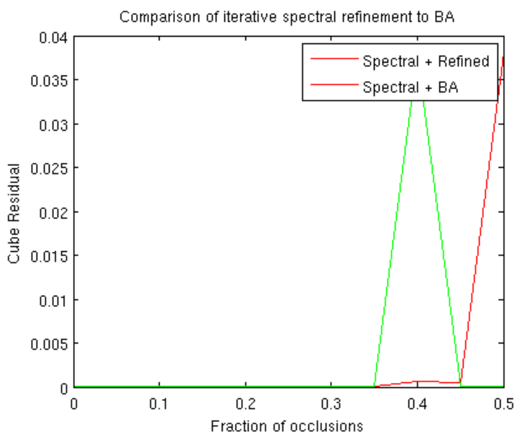

Figure 9: When doing an iterative orthographic camera factorization, which has similar properties to the spectral regularization, the results approached that of bundle adjustment.