Robust Regression on Image Manifolds for Ordered Label Denoising

Hui Wu, Richard Souvenir
Department of Computer Science, University of North Carolina at Charlotte.

A recent trend for acquiring labels for large image sets is via crowdsourcing or co-located sensors, which effectively automates the label collection process, allowing for the rapid creation of large labeled data sets. However, label accuracy often suffers. The goal of our work is to correct mislabeled examples for image sets with ordered labels. We take advantage of the fact that these data sets contain semantically-related images whose relationship can be exploited to learn a smooth function of the labels with respect to the images. From this point of view, the problem can be framed as robust regression in the high-dimensional domain of images. Unlike traditional robust regression methods, our method incorporates the observation that many natural image sets, although embedded in high-dimensional spaces, have only a few underlying causes of change. The contributions of this paper are:

- introducing the problem of ordered label denosing;
- an efficient, data-driven algorithm, based on the Hessian regularizer, for high-dimensional robust regression; and
- providing more accurate labels for widely-used image sets.

For a set of images, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]^\intercal$, each example, $\mathbf{x}_i \in \mathcal{R}^D$, corresponds to the D-dimensional feature representation (e.g., raw pixel values, bag of words, HOG) of image i. We assume that the images are samples drawn from (or near) a low-dimensional manifold, $\mathcal{M}$, embedded in $\mathcal{R}^D$; the labels, $\mathbf{y} = [y_1, y_2, \cdots, y_N]^\intercal$, are samples of a function defined on the manifold; and the set of labels is contaminated by outliers. The goal is to learn a smooth function on the image manifold that recovers denoised labels, $\hat{\mathbf{y}}$. Our formulation includes the Hessian regularizer and an $L1$ loss term:

$$\underset{\hat{\mathbf{y}}}{\operatorname{argmin}} \quad \hat{\mathbf{y}}^\intercal \mathbf{B} \hat{\mathbf{y}} + \lambda \|\hat{\mathbf{y}} - \mathbf{y}\|_1 \tag{1}$$

where $\mathbf{B}$ regularizes the function on the manifold defined by the samples, $\mathbf{X}$, and λ is the trade-off parameter. This convex optimization can be solved efficiently using solvers specialized for large-scale, sparse $L1$-regularized least squares problems.

We compare our method, H3R against the following regression methods: K**-NN**, radial basis function network **(RBFN)**, **RANSAC**, ε-support vector regression **(SVR)**, and kernel supervised principal component analysis **(KSPCA)**. Each method is provided the (corrupted) labeled data as input. Figure 1 shows the ground truth, corrupted input, and regression results from each method for a trial with 50% corruption. Across all of the experiments, H3R returns the closest predicted values, even at corruption rates as high as 80%.

Figure 2 shows images with associated weather metadata. Cloud okta is a measure of cloudiness from clear (0) to cloudy (8). These weather values are estimated from the closest weather stations, which may be far enough to be under different weather conditions from where the image is captured. Those boxed in red are examples where the original label does not appear to match the cloud level depicted in the scene. H3R provided predictions that most closely matched visual appearance of the scene.

Figure 3 shows faces with an estimate of the face pan angle. This parameter would be used to, for example, retain only front-facing subjects. The first row shows sample faces with the associated pose estimate. Each of the subsequent rows show a subset of images sorted by the denoised head pose estimate. The red boxes indicate examples where the pose estimate does not visually match the direction the subject is facing. H3R outperformed each of the competing approaches, resulting in no grossly mislabeled examples.

We presented an algorithm for robust regression on image manifolds and applied it to the problem of ordered label denoising for natural image sets. While the bulk of the algorithms and data sets for supervised learning in computer vision address classification, or categorization problems, there are important problems that rely on ordered output, and our work is one of the first to address this underserved area.

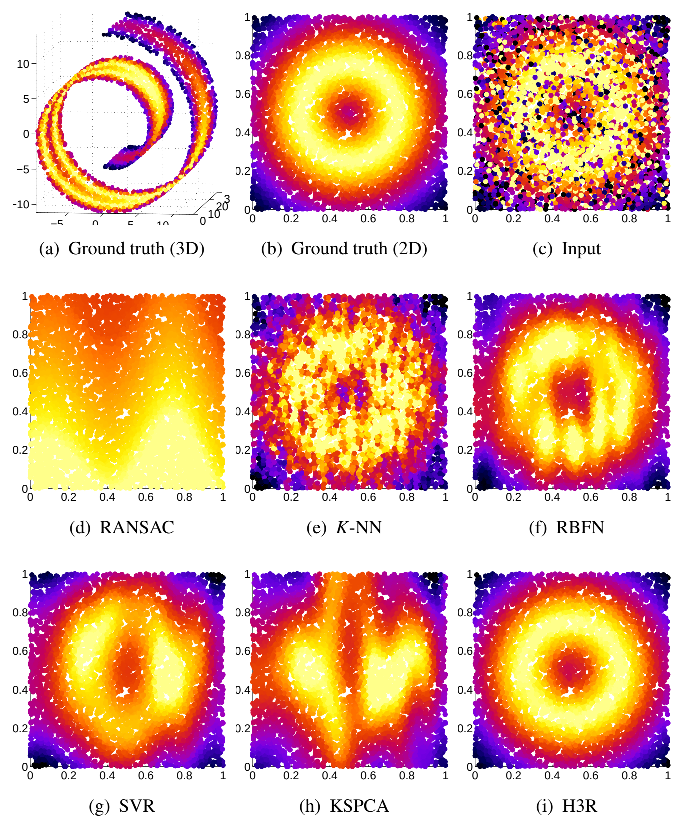

Figure 1: For the Swiss Roll (50% label corruption), the color in each plot indicates the manifold function value. For clarity, (b) to (i) are plotted using 2D manifold coordinates.

	2.4	1.3	7.0	2.8	8.0
RBFN	0.3	5.9	3.2	7.0	8.0
SVR	0.8	3.9	2.2	7.0	7.8
KSPCA	0.4	5.3	2.4	6.3	8.0
H3R	0.8	1.3	4.7	6.4	8.0

Figure 2: Each image shows the original cloudiness label, which ranges, from 0 (clear) to 8 (cloudy). For each method, the predicted value is shown. Clearly mislabeled (input or predicted) values are indicated by the red text and boxes.

Figure 3: For each row, the images are shown with the (input or predicted) head pose estimate. Clearly mislabeled examples are highlighted by red boxes.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.