Learning Natural Parameters of Variation

Erik Miller, Chris Stauffer, Kinh Tieu, Michael Ross & Raquel Romano

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

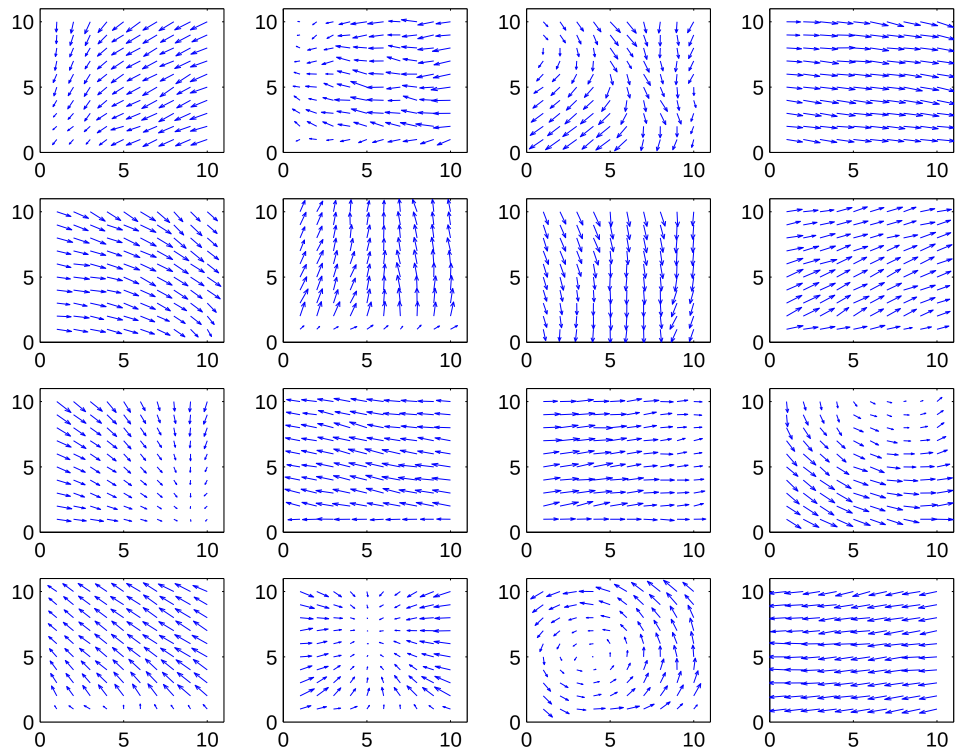

Figure 1: Flow fields learned from a video stream of a static scene with a moving camera.

The Problem: A pervasive problem in machine vision is how to compare images that are produced by the same underlying physical process but under varying imaging conditions. We start by defining what we call the parameterization problem. This is the problem of automatically determining the "natural" parameters of variation in a learning problem, by which we mean the physical parameters that cause variation in the observed image, such as a rotation or a brightness change.

Motivation: Many machine learning and computer vision algorithms rely on a notion of distance between points or feature sets representing the observed data. This depends highly upon the parameters which have been chosen to represent the variation in points. In particular the natural parameters should directly correspond to the causes of variability. Often in machine vision, the parameters are either not addressed or given by the experimenter.

Previous Work: Previous work on modeling variation either assumed a generic ("unnatural") parametric model [1] or specified a priori the variations [2].

Approach: Solving the parameterization problem is difficult given only point samples, since we have in general no way of inferring neighborhood relationships among points. The solution is to use "curve samples" of the image manifold rather than point samples, since they contain explicit information about smooth parameter changes. We define the *image curve process*, a model of variations in natural video streams that allows us to solve the parameterization problem. We begin by capturing continuous video streams of a static scene with various camera motion such as translation and rotation. The assumption is that the scene is constant in almost all parameters. This enables us to obtain curve samples for use in learning the natural parameters of camera motion. As preliminary research, we have learned a set of flow fields by clustering curve samples from our video streams (see Figure 1). These learned flows were successfully used as prior transforms (scalings, translations, rotations) in a "congealing" algorithm that produces "latent image"/transform density estimates [3, 5, 6] from a data set of handwritten digits.

Difficulty: There is a great deal of research in optimizing the values of the parameters of a problem once a specific parameterization is chosen. What parameterization to chose is much less clear. Images of objects typically undergo many different types and large amounts of variation. Unlike, traditional supervised learning, there is no "teacher" providing samples of natural parameterizations.

Impact: By learning the natural parameters of a domain, we can avoid ever having to specify the parameters of variation we wish to model. This retains maximum flexibility for solving real world problems, and at the same time controls the capacity of learned models. It also allows parameters to be shared between models [3, 4]. As a result, current "hard" learning problems may be simplified.

Future Work: We hope to construct a parametric hierarchy differentiating between broadly-applicable transformations and those specific to sub-classes of objects. We will investigate methods for using confidence-rated data in clustering curve samples, and experiments on longer video streams should provide a more complete set of curve samples to learn from. We plan to extend these techniques to other modalities such as audio and to non-transform parameters. We are currently working on learning the natural variations of the colors of objects caused by lighting changes in the scene.

Research Support: This work was supported by a fellowship from Microsoft Corp.

References:

[1] M. Turk and A. Pentland Eigenfaces for recognition. *J. Cog. Neuro.*, 3(1), 1991.

[2] P. Simard and Y. LeCun and J. Denker. Efficient pattern recognition using a new transformation distance. *Adv. Neur. Info. Proc. Sys.*, 5, 1993.

[3] E. Miller and N. Matsakis and P. Viola Learning from one example through shared densities on transforms. *IEEE Conf. on Comp. Vis. and Patt. Recog.*, 2000.

[4] J. Baxter Learning Internal Representations. *Proc. of the 8th Inter. ACM Workshop on Computational Learning Theory*, 1995.

[5] B. Frey and N. Jojic Estimating Mixture Models of Images and Inferring Spatial Transformations Using the EM Algorithm. *IEEE Conf. on Comp. Vis. and Patt. Recog.*, 1999.

[6] T. Vetter and M. Jones and T. Poggio A bootstrapping algorithm for learning linear models of object classes. *IEEE Conf. on Comp. Vis. and Patt. Recog.*, 1997.