Unsupervised Learning of Object Deformation Models
Supplemental Material

Iasonas Kokkinos
Department of Statistics, UCLA
jkokkin@stat.ucla.edu

Alan Yuille
Department of Statistics, UCLA
yuille@stat.ucla.edu

1. AAM learning

1.1. Derivation of the 'feature transport' PDE

Non-trivial deformations result in local contractions and expansions. These naturally capture object scalings but can have a negative side effect, namely making object features disappear or inflate. For this reason we want the deformation fields to have zero acceleration in the direction perpendicular to image features; this guarantees that the features are only 'transported'.

This requirement can be phrased as follows: consider a deformation field $\mathbf{h} = (h_x, h_y) = (x + f_x, y + f_x)$; f_x and f_y are the deformation increments calculated from the linear basis synthesis. Along orientation n_x, n_y this deformation field has 'speed' $f_x n_x + f_y n_y$; a constant speed means that the motion of features in this orientation does not distort them, i.e. it is purely *translating* them. We can thus enforce our constraint by requiring that the directional derivative of this speed function equals zero, i.e.

$$\frac{\partial(f_x n_x + f_y n_y)}{\partial x} n_x + \frac{\partial(f_x n_x + f_y n_y)}{\partial y} n_y = 0, \tag{1}$$

or simply

$$\begin{aligned} \sum_i n_i \partial_i \sum_j f_j n_j = 0, \\ \sum_i \sum_j n_j \partial_j f_i n_i = 0 \end{aligned} \tag{2}$$

where in the first line we simplify notation by identifying the coordinates x, y with the indexes $1, 2$ and in the second we interchange the i and j indexes and reorder the summations.

It is convenient that linear expansions are used to synthesize any object deformation: constraining all shape basis elements to satisfy (2) automatically guarantees that this holds also for any synthesized deformation, since the constraint is also linear in f_x, f_y. We thus solve this problem by iteratively projecting the basis elements onto the space of functions satisfying (2) in alteration with performing gradient descent as in (7) in the paper.

By projecting $f = (f_1, f_2)$ we mean finding the function $g = (g_1, g_2)$ that has minimum L^2 distance from f, while satisfying the constraint (2). We therefore consider the following variational problem:

$$\begin{aligned} \min_g \iint \left[\frac{1}{2} \sum_i (f_i - g_i)^2 + \lambda \sum_i \sum_j n_j \partial_j g_i n_i \right] dx_1 dx_2 = \\ \min_g \iint \left[\frac{1}{2} \sum_i (f_i - g_i)^2 + \lambda \sum_i \sum_j n_j n_i \partial_j g_i + \lambda \sum_i \sum_j n_j g_i \partial_j n_i \right] dx_1 dx_2 = \end{aligned} \tag{3}$$

where λ is a field of Lagrange multipliers, enforcing the inner-product to equal zero.

The E-L equation writes:

$$\sum_i \left[\frac{\partial J}{\partial g_i} + \sum_j \frac{J_{\partial_j g_i}}{\partial_j} \right] = 0 \tag{4}$$

and the terms showing up above for, say, f_i are

$$\begin{aligned}\frac{\partial J}{\partial g_i} = -(f_i - g_i) + \lambda \sum_j n_j \partial_j n_i \\ J_{\partial_j g_i} = \lambda n_i n_j \\ \frac{J_{\partial_j g_i}}{\partial_j} = \partial_j(\lambda n_i n_j)\end{aligned} \tag{5}$$

By substituting we get, for each i:

$$\lambda \sum_k n_k \partial_k n_i + \sum_j \partial_j(\lambda n_i n_j) = 0 \tag{6}$$

are, for,

Taking the Euler-Lagrange derivative yields:

$$\partial_1(\lambda n_1 n_i) + \partial_2(\lambda n_2 n_i) = f_i - g_i, \quad i = 1, 2 \tag{7}$$

Since g must satisfy $\sum_{i,j=(1,2)} \partial_j g_i n_i n_j = 0$, we can eliminate the (unknown) g and end up with an equation involving only f. For this, we build the individual summands by appropriately differentiating (7) w.r.t. x_i and multiplying it with n_i, and add up the corresponding left and right hand sides. We thereby obtain the following elliptic PDE:

$$\sum_{i,j=(1,2)} n_i n_j \left[\partial_{1,j}(\lambda n_1 n_i) + \partial_{2,j}(\lambda n_2 n_i)\right] = \sum_{i,j=(1,2)} \partial_j f_i n_i n_j \tag{8}$$

from which we estimate λ by solving the corresponding linear equation system. We then use the calculated λ in (7) to estimate g_i from f_i.

We locally estimate the orientation n_1, n_2 of the template domain by averaging the orientation of the (deformed) primal sketch contours that were used to build up the template at the previous iteration.

We note that in [3] a similar approach was used to derive divergence-free deformation fields. Herein we extend this approach to the case where we use the local structure of the image is used to dictate the direction in which there should be no change in scale. Specifically, the spatially varying signals n_1, n_2 are related to the image orientation and are used to avoid the contraction/expansion of features along that direction. The idea of [3] stemmed from fluid dynamics, while we are not aware of work related to the PDE we propose here.

1.2. Enforcing Orthonormality

Herein we derive the update used to enforce the orthonormality of the eigenvectors. The constraint can be enforced by introducing Lagrange multipliers in the initial criterion used to derive the update rule:

$$C(S, I, \mathbf{s}) = \sum_{k=1}^{N} \left[\sum_{\mathbf{x}} \left[I_k(S(\mathbf{x}; \mathbf{s}_k)) - \mathcal{T}(\mathbf{x})\right]^2 + \lambda \sum_j \frac{\mathbf{s}_{k,j}^2}{\sigma_j^2} \right] \tag{9}$$

Specifically, we consider the multiplier $\mu_{i,j}$ associated with the constraint

$$S_i \cdot S_j \quad = \quad \delta_i(j), \quad \text{where} \tag{10}$$

$$S_i \cdot S_j \quad \equiv \quad \sum_{\mathbf{x}} \left[S_{x,i}(\mathbf{x}) S_{x,j}(\mathbf{x}) + S_{y,i}(\mathbf{x}) S_{y,j}(\mathbf{x})\right], \quad \delta_i(j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{11}$$

The modified criterion writes

$$C(S, I, \mathbf{s}) + \sum_i \sum_{j \geq i} \mu_{i,j} (S_i \cdot S_j); \tag{12}$$

taking the partial derivatives of the criterion with respect to $S_{x,i}(\mathbf{x})$ and $S_{y,i}(\mathbf{x})$ gives

$$\frac{\partial E}{\partial S_{x,i}(\mathbf{x})} + \sum_j \mu^*_{i,j} S_{x,j}(\mathbf{x}) = 0 \tag{13}$$

$$\frac{\partial E}{\partial S_{y,i}(\mathbf{x})} + \sum_j \mu^*_{i,j} S_{y,j}(\mathbf{x}) = 0 \tag{14}$$

$$\text{where} \quad \mu^*_{i,j} = \begin{cases} \mu_{i,j} & i \geq j \\ \mu_{j,i} & j < i \end{cases} \tag{15}$$

$$\tag{16}$$

Multiplying (13) with $S_{x,k}(\mathbf{x})$ and (14) with $S_{y,k}(\mathbf{x})$, summing the resulting equations and then summing over $\mathbf{x}$ we get the system of equations

$$\sum_{\mathbf{x}} \left(S_{x,k} \frac{\partial E}{\partial S_{x,i}(\mathbf{x})} + S_{y,k} \frac{\partial E}{\partial S_{y,i}(\mathbf{x})} \right) + \sum_j (S_j \cdot S_k) \mu^*_{i,j} = 0 \tag{17}$$

Taking the derivative of the criterion with respect to the Lagrange multipliers we get

$$(S_j \cdot S_k) = \delta_j(k) \tag{18}$$

From this we get the relation

$$\mu^*_{i,j} = \sum_{\mathbf{x}} \left(S_{x,j} \frac{\partial E}{\partial S_{x,i}(\mathbf{x})} + S_{y,k} \frac{\partial E}{\partial S_{y,i}(\mathbf{x})} \right) \tag{19}$$

which finally gives us the optimal update of S that satisfies the introduced constraints:

$$\sum_{\mathbf{x}} \left(S_{x,k} \frac{\partial E}{\partial S_{x,i}(\mathbf{x})} + S_{y,k} \frac{\partial E}{\partial S_{y,i}(\mathbf{x})} \right) + \sum_j (S_j \cdot S_k) \mu^*_{i,j} = 0 \tag{20}$$

We can see this as a modification of the technique used in [4] to our problem; in their work the orthonormality constraints were imposed at the pixel level, herein they are imposed on functions defined over the whole image.

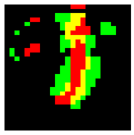
2. Details on the estimation of the observation potentials

Due to lack of space in the paper we provided a brief sketch of the method employed to estimate the observation potentials. Herein we provide some details to clarify the points mentioned in the paper.

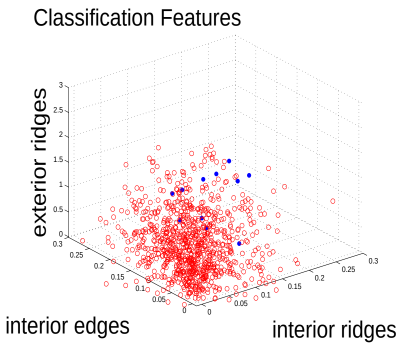
Our approach was driven primarily by the need to estimate the terms involved in the observation potentials efficiently. As described in the paper this can be accomplished using features that are expressed as summations of feature fields over domains, which can then be efficiently accomplished using curvilinear integrals instead of area summations.

Specifically, each particle used in the NBP inference scheme amounts to a specific estimate of the deformation. After deforming the template accordingly, we extract the sums of ridge/edge strengths in the interior and exterior of the areas defined by the template, and treat them as a four-dimensional feature vector that we use to evaluate the observation potentials. This deviates from the typical MRF formulation, where the observation potentials express the likelihood of the observations instead of some features; still as in [2] we argue that with good enough features the difference in performance will not be substantially different.

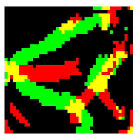
The expression for the observation potentials is based on the output of a classifier that compares the likelihood of the features under a foreground and a background model. The foreground model is built based on the previous E-step results, by extracting the features at the estimated part locations and constructing a nonparametric Gaussian distribution -we use the methods and code in [1] for efficiency. For the background model the part locations are randomly perturbed, and the process is repeated. The features used to construct the classifiers for two different parts of cows can be seen in Fig. 1.



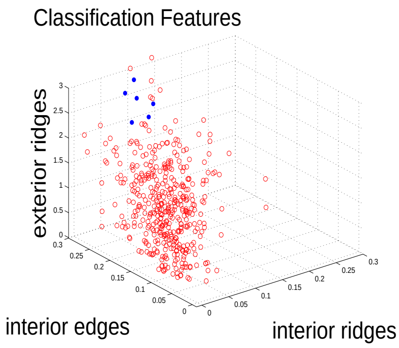
(a) Template for front foot



(b) Features for classifier construction



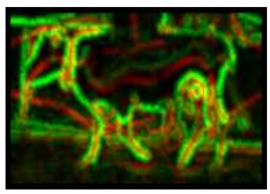
(c) Template for head



(d) Features for classifier construction

Figure 1: Features used to construct the fore- and background likelihood terms used for the estimation of the observation potentials. Blue solid dots correspond to foreground and red to background features.

(a) Input Image



(b) Edges and Ridges

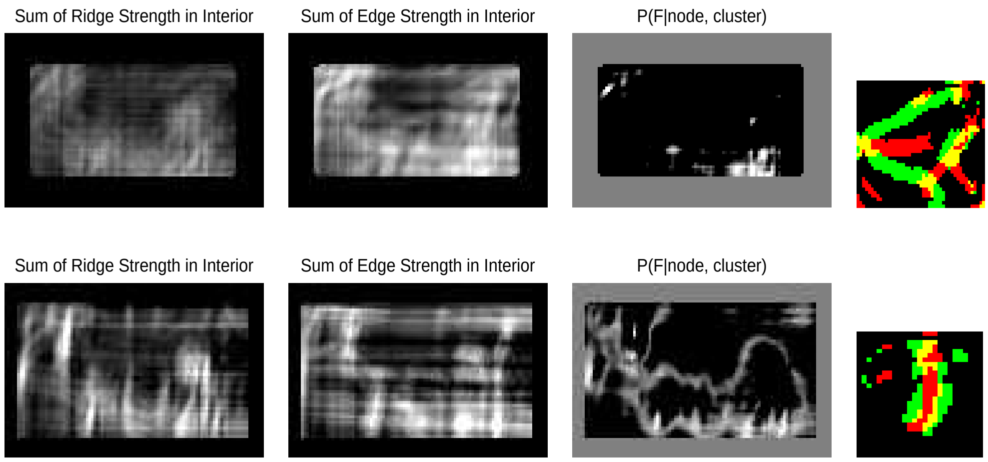


Figure 2: Observation potentials corresponding to two object parts.

References

[1] A. Ihler, E. Sudderth, W. Freeman, and A. Willsky. Efficient Sampling of Gaussian Distributions. In *NIPS*, 2004.
[2] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Extending pictorial structures for object recognition. In *BMVC*, 2004.
[3] P. Simard and G. Mailloux. A Projection Operator for the Restoration of Divergence-Free Vector Fields. *IEEE Trans. PAMI*, 10(2), 1988.
[4] D. Tschumperle and R. Deriche. Orthonormal Vector Sets Regularization with PDEs and Applications. *Int.'l Journal of Computer Vision*, 50(3), 2002.