Deformation invariant image matching by spectrally controlled diffeomorphic alignment

Deformation Invariant Image Matching by Spectrally Controlled Diffeomorphic Alignment*

Christopher M. Yang Sai Ravela
Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology
http://ravela.net/sca.html

Abstract

We present a new approach to deformation invariant image matching. Our matcher (a) aligns templates to targets over a broad range of nonlinear deformations, (b) factors the total deformation into spectral categories, where low wavenumber deformations are smooth and global and high wavenumbers are turbulent and local, and (c) weighs the reduction in template-target misfit within each category to differentiate between relevant and irrelevant deformations.

It accomplishes this by aligning images in a scale-cascaded fashion, with more complex, local deformations following simpler, more global ones. Each step of the cascade involves finding an iterative solution to a nonlinear optimization problem using a Gabor deformation basis. Cascaded alignment makes deformation invariant matching feasible and efficient. Our approach is applied to recognize the flexible bodies of salamanders from a large database; results indicate that the method is very promising.

1. Introduction

Algorithms to recognize animals and plants from their photographs hold great promise for the study of migratory behavior and for the development of conservation plans [8]. However, a vast number of animals and plants deform in highly nonlinear ways, which makes recognizing them difficult. Indeed, the well-known trade-off between invariance and selectivity teaches us that a sparsely parameterized model for handling deformations found in nature will be impossible for object recognition. As invariance to a larger range of deformations is accomplished, perceptually unrelated objects are just as completely and easily related to one another as perceptually related ones [9].

*This material is funded in part by NSF DBI-0640529, NSF CNS-0540259, and Lincoln Labs #7000074210. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily represent those of the sponsors.

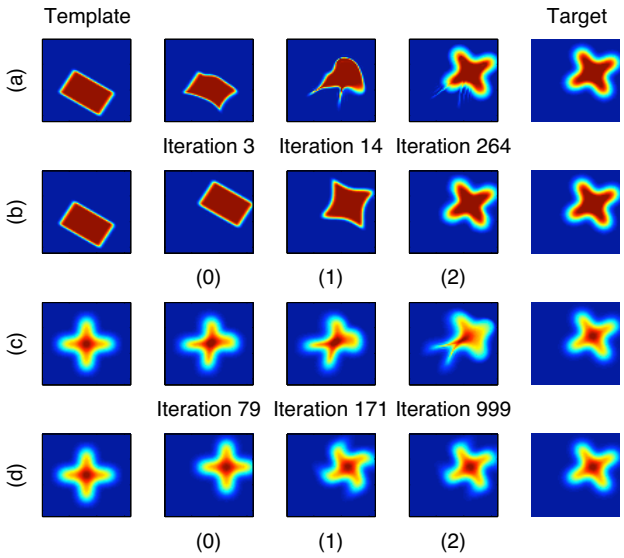

Figure 1. (a) A rectangle is nonlinearly and iteratively deformed to a cross using viscous alignment. Although powerful, such methods cannot be used for recognition. (b) Using our method, the same rectangle can be diffeomorphically aligned but with intermediate states useful for image matching and recognition. (c) A translated and rotated cross elicits a complex explanation from viscous alignment. (d) Our method prefers simpler explanations, which recovers the translation and rotation without losing the flexibility of viscous alignment. No correspondences are involved.

Let us suppose that objects can be recognized by matching their images. In the following discussion, we will also assume that a total deformation vector-field $\mathbf{q}$ nonlinearly maps a 2D template X into a target Y discretized on a domain Ω. Using $\underline{p} = (x, y)$ as a position coordinate we may define $X \circ \mathbf{q} \equiv X(\underline{p} - \mathbf{q}(\underline{p}))$ to be the deformation of the template, evaluated if necessary using interpolation. This deformation field will typically be the solution of an optimization procedure of the form: $J(\mathbf{q}) := \frac{1}{2} ||Y - X \circ \mathbf{q}||^2_{\underline{\underline{R}}} + L(\mathbf{q})$, where $\underline{\underline{R}}$ is a covariance defining

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

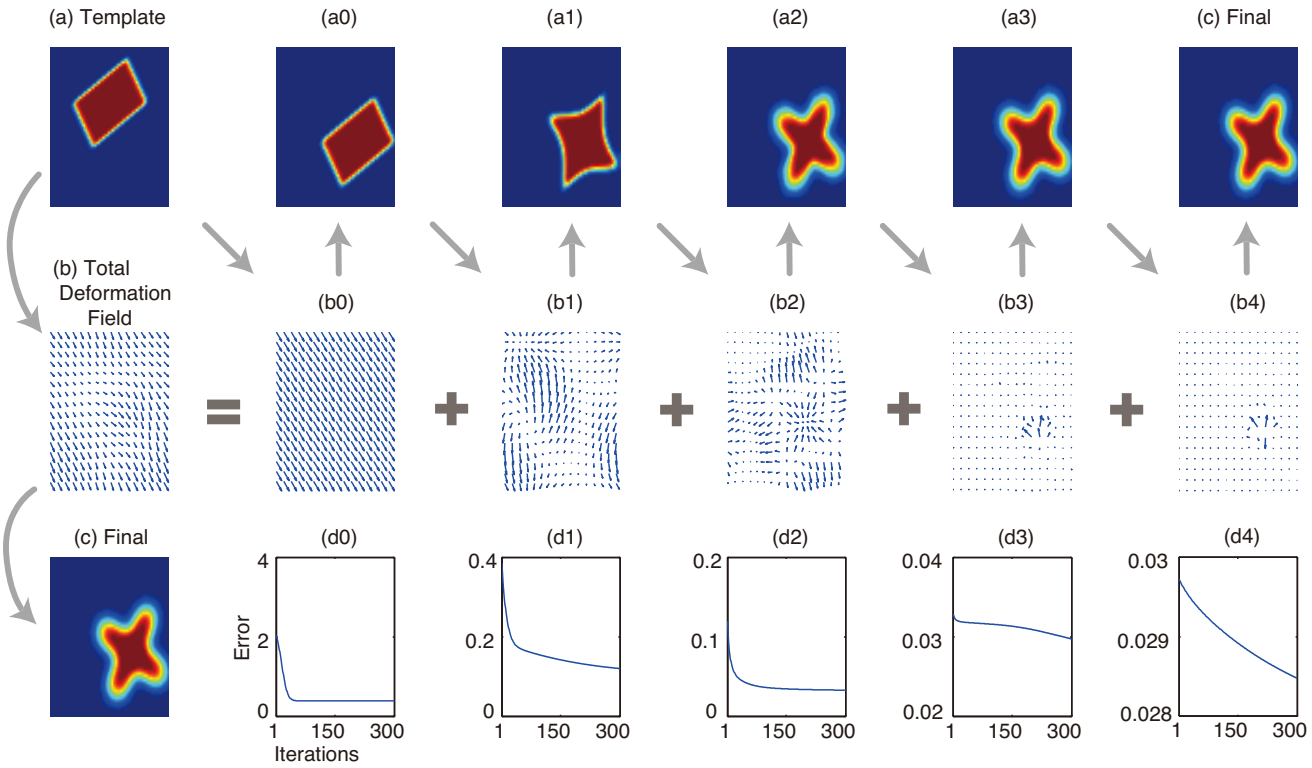

Figure 2. Summary of scale-cascaded alignment. The template (a) is completely warped into the target using viscous alignment, producing a total deformation field (b). Our method factors the total deformation into spectral categories, which start with translations (b0) and become increasingly more local (b4). The template (a) can be advected by the component categories in a cascade, generating the intermediate states shown in (a0)–(a3). The deformation within each category (b0)–(b4) is obtained by iteratively solving a nonlinear optimization problem, using a Gabor deformation basis. The corresponding error sequences (d0)–(d4) can be weighed to discriminate between categories of deformations by their relevance to object similarity.

the norm and $L(\mathbf{q})$ expresses constraints on the deformation field. In this paper, we are particularly interested in diffeomorphic alignment using smoothness, non-divergence and other low order fluid-like or higher order differential constraints [3, 4, 10, 19], that we shall refer to here as *viscous alignment* methods.

To match images, the first apparent measure of similarity is the residual after deforming the template into the target. But this measure is problematic. As shown in Figure 1(a), viscous alignment can be used to easily, automatically, and unambiguously warp the rectangle into the cross. But now there is a total loss of selectivity; the two cannot be distinguished. A second measure of similarity can be constructed when an object category exhibits characteristic deformations. But this is also problematic. Characteristic-deformations change between categories or they may be sufficiently varied within a single object category. Thus, a more universal deformation model may be warranted. Viscous alignment is a powerful approach in this regard, but its "invariance," the ability to align under highly nonlinear deformations, is obtained at the cost of selectivity, as the following example shows.

Let us merely choose $L(\mathbf{q})$ to prefer smooth and non-divergent deformations [3, 4, 10, 19], which are the simplest constraints for diffeomorphic alignment (see Figure 1(a)). Applying this approach to a simple problem can be disastrous, as shown in Figure 1(c). The translated and rotated cross elicits a lot more than a translation and a rotation as an explanation, evident from the complicated intermediate states as the optimizer iteratively aligns the template to the target. As we shall see, this problem does not really arise from the optimizer but is more fundamentally related to the constraints. And we cannot, obviously, terminate the optimization midway in a naïve attempt to "control" the deformation because intermediate states may bear no meaningful resemblance to the template or the target, see Figure 1(a,c).

Is there a way to align flexibly, as powerfully as viscous alignment, and yet have the ability to discriminate and recognize? The principal contribution of this paper is to address this dichotomy by converting diffeomorphic viscous alignment into an image matching tool appropriate for object recognition. Our deformation invariant matcher has the following properties: (a) It aligns images over a broad range of nonlinear deformations. (b) It factors the total deforma-

tion into "universal" categories ranging from translation to turbulence. (c) It weighs the reduction in template-target misfit within each category to differentiate between relevant and irrelevant deformations. In this way, an image matching tool for whole images or image patches around distinguished locations is proposed.

The key elements of our approach are outlined in Figure 2, illustrated here with the sequence shown in Figure 1(a,b). The template (a) is aligned with the target by a total deformation (b) obtained as a solution to an optimization problem with smooth, non-divergent regularizing constraints. The result is a deformed template (c). However, as argued, neither the output (c) nor the deforming field (b) are useful as is for recognition.

In our approach, the total deformation (b) is factored into spectral categories that start with translations (b0) and become progressively more local and turbulent (b4). Each deformation (b0 through b4) is the result of iteratively solving an optimization problem using Gabor filters. The error between template and target within each category is reduced over several iterations, and error sequences are shown in d0 through d4. The deformations b0 through b4 are produced and applied in a cascade, depicted by the sequences shown from a0 through a3 (and in the sequences in Figure 1(b,d)). We call this scale-cascaded alignment (SCA).

Although SCA also produces a diffeomorphically aligned template, it produces better explanations for the error between template and target (Figure 1(d)) and generates more meaningful intermediate states (Figure 1(b)). This makes it a robust alignment tool which retains invariance to a broad range of nonlinear deformations. Additionally, the error sequences (d0 through d4) can be weighed to discriminate between categories of deformations by their relevance to object similarity. We can, for example, easily discard or accept the contributions of deformations (b1) and beyond, depending on whether such higher order deformations are natural modes of variability of the rectangle or not.

The resulting matcher is simply parameterized, powerful and can be used for matching without detecting corresponding features and possibly with sparse measurements. To the best of our knowledge, such an approach to deformation invariant matching has not been shown before. To demonstrate the utility of our approach, we apply it to recognize the marbled salamander, a flexible animal, using over 6000 photographs collected by field biologists over six years. The results improve existing work and are very promising.

2. Related Work

Deformable templates were introduced with Widrow's rubber masks [24] and Fischler and Elschlager's spring-loaded templates [7] (also see [13]). Of particular interest are free-form deformable models [5, 12, 14, 17, 20]. These models all tend to be low-dimensional in the types of deformations they can produce or describe. Our scale-cascaded alignment, in contrast, is easily parameterized to encompass deformations from simple translation to turbulence.

Correspondence-based deformation warps images by corresponding a small set of features and constructing a global warp based on that correspondence [2, 22]. Our method uses no feature correspondence and thus is applicable for whole image matching. Alternatively, this method can be adapted by matching the neighborhood of potential correspondences.

Fluid alignment model [4, 19] are related to this work. Our work shares the viscous fluid model, but develops a cascaded approach to jump the gap between alignment and matching. Our spectral interpretation is related to Heeger's spatio-temporal filters [11], but the deformable model formulation is different. Gramkow and Bro-Nielsen's [10] use convolution filters to improve Christensen's computations, and Thirion [21] uses a Gaussian to regularize. There is no connection to image matching in these papers. Amit et al. [1] present a cascaded alignment, but our method does not use a Green's function approach. Our filters are localized, and the solution uses no stochastic sampling or KL expansions. Most significantly, we cascade completely and weigh the error reductions, rather than truncate the cascading by truncating the terms in the KL expansion.

Our work draws motivation from Ling and Jacobs [16], who use the *geodesic-intensity histogram* (GIH) as a local descriptor that is deformation invariant. There are similarities to the effect of their invariance parameter α and our spectral parameterization, but both the representations and problem formulations are entirely different.

3. Deformation Invariant Matching

In order to use diffeomorphic alignment for matching, we will start by reviewing "viscous" alignment. Using the notation in Section 1, we may formulate the alignment problem as a search for a deformation field $\mathbf{q}$ that maximizes the *a posteriori* probability $P(\mathbf{q}|X,Y)$. Using Bayes' rule we write: $P(\mathbf{q}|X,Y) \propto P(Y|X,\mathbf{q})P(X)P(\mathbf{q})$. The RHS has a data likelihood, an amplitude prior (assumed independent of displacements), and a displacement prior. We will suppose that the component densities are Gaussian and thus produce a quadratic objective:

$$\begin{aligned} J(\mathbf{q}) \;=\; & \frac{1}{2}\sum_{\underline{r}\in\Omega}\sum_{\underline{s}\in\Omega}\{\,[Y(\underline{r}) - X(\underline{r}-\mathbf{q}(\underline{r}))]\,C(\underline{r},\underline{s}) \\ & \times [Y(\underline{s}) - X(\underline{s}-\mathbf{q}(\underline{s}))]\} \\ & + L(\mathbf{q}) \end{aligned} \tag{1}$$

Here C is field associated with the inverse covariance $\underline{\underline{C}} = \underline{\underline{R}}^{-1}$, which we assume in this paper to be static for

the optimization. The displacement prior is based on an energy function $L(\mathbf{q})$ modeled with divergence and non-smoothness penalties [19]:

$$L(\mathbf{q}) = \frac{w_1}{2}\sum_{\underline{z}\in\Omega}[\nabla\mathbf{q}(\underline{z})^T\nabla\mathbf{q}(\underline{z})] + \frac{w_2}{2}\sum_{\underline{o}\in\Omega}[\nabla\cdot\mathbf{q}(\underline{o})]^2 \quad (2)$$

The Euler-Lagrange equation of the objective (Equation 1) is a highly nonlinear PDE, so we solve iteratively. In particular,

$$\begin{aligned} X_i(\underline{p}) &\leftarrow X(\underline{p} - \mathbf{q_{0:i-1}}(\underline{p})) && (3)\\ \delta\underline{X}_i &\doteq \underline{\underline{C}}\,[\underline{Y} - \underline{X}_i] && (4)\\ \frac{\partial L(\mathbf{q_i}(\underline{r}))}{\partial \mathbf{q_i}(\underline{r})} &= \nabla X_i(\underline{r})\delta X_i(\underline{r}) && (5)\\ &\doteq \mathbf{f_i}(\underline{r}) && (6)\end{aligned}$$

Here $\mathbf{q_{0:i-1}}$ is the total deformation field at the start of iteration i and $\mathbf{q_i}$ is the instantaneous displacement at the end of iteration i. The image X_i is obtained by applying the total deformation to the original (template) image. The vectors $\underline{Y}$ and $\underline{X}_i$ are the target and the evolving template respectively, rasterized to vector form. The field $\mathbf{q_{0:i-1}}$ is advected by $\mathbf{q_i}$ to obtain $\mathbf{q_{0:i}}$. Thus, at iteration i, by fixing $\mathbf{f_i}$, we have a linear system

$$\begin{aligned} w_1\nabla^2\mathbf{q_i}(\underline{p}) + w_2\nabla(\nabla\cdot\mathbf{q_i}(\underline{p})) - \mathbf{f_i}(\underline{p}) &= 0 && (7)\\ \underline{\underline{G}}\,\underline{\mathbf{q_i}} &= \underline{\mathbf{f_i}} && (8)\end{aligned}$$

Here $\underline{\underline{G}}$ is the sparse matrix representing the differential operators, and the vectors $\underline{\mathbf{q_i}}$ and $\underline{\mathbf{f_i}}$ are obtained from the corresponding fields. This system represents diffeomorphic, fluid-like flows, which we term "viscous" alignment.

Christensen solved Equation 7 using SOR [4], which was improved using convolution filters [10], and conjugate gradients have been suggested [3]. However, we use spectral methods (the FFT diagonalizes $\underline{\underline{G}}$), which are exact, relatively efficient and pose no issues representing homogeneous dirichlet boundary conditions. The approach may be implemented on a pyramid for even better computational efficiency [10].

Because $X(\underline{p} - \mathbf{q}(\underline{p}))$ is not linearized as in optic flow, complex deformations are produced by advecting the evolving deformation field with the instantaneous displacement field at each iteration (and optionally with restarts, see [4]). The solution, to be sure, is still *local.* Nevertheless, deformations of amazing complexity with no correspondences whatsoever can be generated.

This should not be surprising because Equation 7 represents Navier's equation in equilibrium [18], with an image driven body force. If we drop the Laplacian term, it is a proper (inertia-less) fluid. If we drop the continuity term, we have the Laplace-Beltrami operator. We can thus represent viscoelastic, viscous and fluid-like motions. This flexibility is the basis for success in aligning objects with complex deformations.

3.1. Spectral Interpretation

The spectral interpretation also shows exactly why viscous alignment is a limitation for matching images to recognize objects. In what follows, we drop the explicit notation for iteration i and simply use $\mathbf{q}$ to denote the instantaneous displacement. Let us suppose the vector field $\mathbf{q}(\underline{p}) \doteq (\mathbf{q^x}(\underline{p}) \quad \mathbf{q^y}(\underline{p}))$ and $\underline{p} \doteq (x, y)$ with a Fourier transform pair $\mathcal{Q}(\underline{\omega}) \doteq (\mathcal{Q}^x(\underline{\omega}) \quad \mathcal{Q}^y(\underline{\omega}))$ with $\underline{\omega} \doteq (m, n)$. Similarly, let $\mathbf{f} \leftrightarrow \mathcal{F}$ be a transform pair. We can then write a solution to Equation 8 in wavenumber space, exactly ($m \neq 0$ and $n \neq 0$) of the form:

$$\begin{bmatrix}\mathcal{Q}^x\\ \mathcal{Q}^y\end{bmatrix} = \begin{bmatrix}\mathcal{H}^a & \mathcal{H}^b\\ \mathcal{H}^b & \mathcal{H}^c\end{bmatrix}\begin{bmatrix}\mathcal{F}^x\\ \mathcal{F}^y\end{bmatrix} \quad (9)$$

For the sake of simplicity, let us set $w_2 = 0$, which leads to Laplace-Beltrami (but the problem obviously remains well-regularized to produce complex deformations). This is the resulting filter

$$\mathcal{H}^p = \mathcal{H}^a = \mathcal{H}^b = -\frac{1}{w_1(m^2+n^2)} \quad (10)$$

Equation 10 is simply the Fourier transform of the Laplacian and clearly prescribes a power-law energy spectrum for instantaneous deformations. Thus, this filter is capable of producing complex deformations because $\mathcal{Q}$ can have very high frequencies (see Figure 1 and Figure 3). High frequency deformations may thus arise even when the solutions are apparently "'simpler," as shown in Figure 1. How then do we leverage power-law spectra for recognition?

3.2. Deformation Filters and Cascaded Alignment

We answer this question by approximating viscous alignment with a set of filters that collectively have the power of diffeomorphic alignment, but can be arranged or selected in special ways to factor the total deformation into meaningful parts. This approach is developed in three steps: (i) a tunable and non-singular approximation of the power law, (ii) a Gabor filter basis implementation of the approximation, and (iii) scale-cascaded alignment (SCA). SCA will immediately become useful for image matching, and be discussed in Section 3.3.

i. Laplace Approximation of Power Law: In the viscous formulation, w_1 only controls the convergence rate (provided stability is maintained, see [19]) but not the shape of the spectrum[1]. As a first step, we build a tuner to control the shape of the filter.

[1]We may choose w_1 to be anisotropic and space-varying, but that is difficult to design.

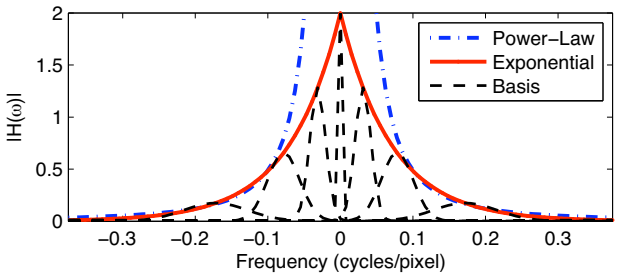

Figure 3. An exponential envelope (red) approximates the power-law envelope (blue) from Equation 10. The basis filters $\mathcal{H}$ are attenuated by the exponential envelope, shown here in 1D.

The tuner is constructed by approximating Equation 10 with the Laplace "distribution": $e(r) = \beta e^{-|r|/2\alpha^2}$, where $r = \sqrt{m^2 + n^2}$. This approximation overcomes the singularity in the power law and it can reasonably approximate any power law using the parameters β (gain) and α (bandwidth), thus producing instantaneous displacements ranging from translations to turbulence. A comparison of the original filter and the tunable approximation is shown in Figure 3 for a power-law of 2.

ii. Gabor filter basis for deformations. The next step in the process is similar to the design and use of Gabor filters for texture. The deformation spectrum is decomposed using a Gabor basis (see Figure 3). The peak power of a filter in the basis is the power of the Laplace approximation at the filter's center frequency. The entire filter bank is an approximate implementation of a power law. Filters in the bank can also be applied in select sequences, which is central deformation invariant matching. Here we discuss the design of the Gabor basis (see Figure 4).

Let us consider frequency rings labelled R_0 to R_N, with N chosen to be logarithmic in the size of the image. On each ring at a radius $r_i = r(R_i)$, we place n_i Gaussians $G_j, j = 1 \ldots n_i$ azimuthally at θ_j and all with scale σ_i. Let us parameterize a Gaussian as $G(r, \theta, \sigma)$ and thus the filter bank at ring R_i is:

$$\mathcal{H}_i = \mathcal{H}(R_i) = \frac{1}{\aleph_i} \sum_{j=1}^{n_i} G(r_i, \frac{2\pi}{n_i} j, \sigma_i) \qquad (11)$$

where $\aleph_i$ is a normalizing constant.

We now construct a scale-cascaded filter bank. The radius $r_0 = r(R_0) = 0$ and $r_1 = 1/2$ cycle and $r_{i+1} = 2r_i$, $\sigma_1 = 1/2$, $\sigma_{i+1} = 2\sigma_i$, and $n_{i+1} = 2n_i$. The amplitudes of filters in the bank are scaled by the magnitudes of the Laplace approximation, as discussed. Thus,

$$\mathcal{H} = \sum_{i=0}^{N} \beta e^{-|r_i|/2\alpha^2} \mathcal{H}(R_i) \qquad (12)$$

This filter bank has two parameters and we choose $\beta = 1$ and $\alpha = \frac{1}{8}$ for our experiments. Figure 4 shows the

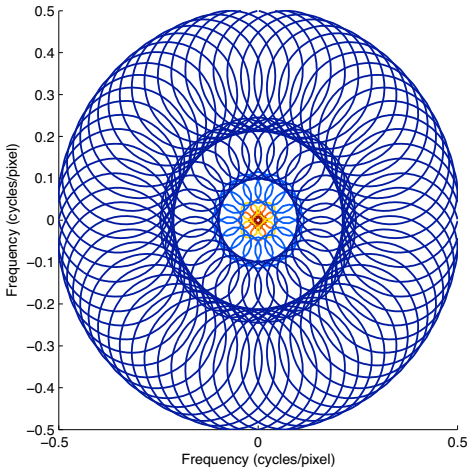

Figure 4. The 1-σ contours of the filters in the Gabor filter bank in frequency domain. Each ring uses twice the number of filters than the previous ring. Similarly, each ring also doubles σ.

1-σ contours of the filters in the filter bank. Note that the original kernel is real and so is our filter bank. We thus rewrite Equation 9 as $\mathcal{Q}^x = \sum_{i=0}^{N} \mathcal{H}(R_i)\mathcal{F}^x$ and $\mathcal{Q}^y = \sum_{i=0}^{N} \mathcal{H}(R_i)\mathcal{F}^y$. Such a reparameterization of viscous alignment has not hitherto been proposed.

Note that the off-diagonal term $\mathcal{H}^b$ of Equation 9 is omitted here because $\mathcal{H}^a$ and $\mathcal{H}^c$ sufficiently regularize the problem. The off-diagonal term, which is also real and exhibits similar power-law behavior can, if necessary, be easily incorporated.

Algorithm 1 Scale-cascaded alignment.

INPUTS: Template X, Target Y, Filter bank $\mathcal{H}$
$X_0 \leftarrow X$
for $i = 0$ to N **do**
  $\mathbf{q_{0:0}} \leftarrow 0, j \leftarrow 1$
  while has not converged and $j <$ limit **do**
    Calculate $\mathcal{F}^x_j, \mathcal{F}^y_j$ using Y, $X_i(\underline{p} - \mathbf{q_{0:j-1}}(\underline{p}))$
    Solve: $\mathcal{Q}^x_j = H_i \mathcal{F}^x_j$
    Solve: $\mathcal{Q}^y_j = H_i \mathcal{F}^y_j$
    Update: $\mathbf{q_{0:j}}$ using $\mathbf{q_j}$ and $\mathbf{q_{0:j-1}}$
    $j \leftarrow j + 1$
  end while
  $X_{i+1}(\underline{p}) \leftarrow X_i(\underline{p} - \mathbf{q}(\underline{p}))$
end for

Scale-Cascaded Alignment on the Gabor bank: Algorithm 1 describes the scale-cascaded alignment procedure. It sequentially applies filters in the bank from DC to the highest frequencies. Because there are multiple filters at each radius, we can assert even finer control over the deformation by selecting a subset (or even a single one) from among them. Such fine-grained control is neither necessary in the application nor is it developed here. Further, although

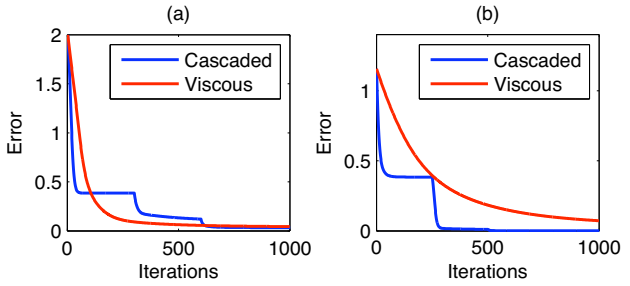

Figure 5. Error sequences for (a) rectangle-to-cross [Figure 1(a,b)] and (b) translating-rotating cross [Figure 1(c,d)]. Red curves show the error sequence using viscous alignment; blue curves show the error sequence using the scale-cascaded approach (SCA). Convergence rates of viscous alignment are not correlated to perceptual similarity.

certain applications may require filters at alternate center frequencies, the overall cascaded procedure will remain the same.

Once filters across scale have sequentially been applied in their entirety, SCA will implement a diffeomorphic alignment and produce a final state (or deformed image) that is close to the result of viscous alignment. However, SCA holds certain advantages. In Figure 5, the red error curve represents viscous alignment, shown in the sequence in Figure 1(a). The blue one (Figure 1(b)), with its characteristic drops, depicts the dissipation of error energy from the lower to higher frequency. For the rectangle-to-cross case, viscous alignment converges relatively quickly, but there is no way to factor the error into perceptually relevant deformation classes. The cascaded approach converges more weakly, but we can see exactly what the contributions of filters at each radius are[2]. For the translating and rotating cross, viscous alignment ironically takes much longer to converge, explaining the rotation and translation with very high frequency deformations. The scale-cascaded alignment approach converges rapidly, identifying the translation and then rotation, leaving negligible numerical residue for higher wavenumber filters to resolve.

Because the dissipation of error energy is tied to a spectral radius, we may "stop" the process at a desired spectral radius or weigh the errors at different radii differently. There is no way to be selective in the original viscous alignment approach, and this difference is key to how scale-cascaded alignment can be adapted for deformation invariant matching.

3.3. Deformation Invariant Matching

Let us say that the error sequence due to scale-cascaded alignment is $\Delta(X,Y) = [\Delta_0\ \Delta_1 \dots \Delta_N]$, corresponding to the dissipation of the optimization objective energy by each filter $\mathcal{H}_0 \dots \mathcal{H}_N$. The entire sequence can be thought of as a vector, as shown for example in the plot in Figure 5.

Deformations are scored as a weighted sum with a scoring curve, that is $e(X,Y) = \Delta(X,Y)W$. There are several ways to choose W. As a primer, consider three image patches in Figure 6(a,b,c). Though it is obvious to the reader that template (b) is the true match to the target (a), the template (c) is a better match with multiscale PCA (see Section 4). Note that PCA is a special case of the objective (Equation 1) sans the regularization constraint. Thus, as shown in Figure 6(d), the starting point of the error sequence shows the wrong template to be a better match. But as we run the scale-cascaded alignment, the correct template aligns itself rapidly while the incorrect one struggles. As the alignment proceeds to higher wavenumbers, the error reduction is dominated by work done "undoing" the effects of nuisance variables (specularity and noise, for example). This must be discounted. It appears that there is a band within the error sequence that accurately depicts the relative "perceptual closeness" of the two targets to the template. Its active region corresponds roughly to $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$.

Thus, we choose a parameterization of W with $j \geq 0$ as the iteration sequence (concatenated over the cascade)

$$W(j;\sigma_c,O) = \left(\frac{j}{\sigma_c}\right)^{\left(\frac{2}{O}\right)} e^{-\left(\frac{j}{\sigma_c}\right)^O} \qquad (13)$$

Here, σ_c is the cut-off point and O is the order of the weight curve. For the three-patch example, the scoring curve is also shown in Figure 6, obtained by setting $\sigma_c = 600$ and $O = 10$.

Scoring functions can be designed (or learned) to reflect any set of preferences for deformation invariance. For example, we might prefer to be invariant to global transformations and turbulent motion (*i.e.* noise) but be sensitive to all other deformations. Or we may continually discard higher-frequency deformations in favor of lower ones. This is difficult to model in other approaches, such as Ling and Jacobs [16], but presents no difficulty in the scale-cascaded approach. Further, note that we do *not* simply terminate alignment at a preset spectral radius, though that is useful if alignment were the main objective. For recognition, we argue for fully applying SCA (i.e. completely "morph" template to target) and then post process the error curves with W.

4. Application

We apply our method to a problem in conservation biology that requires the recognition of individual marbled salamanders [8]. A database of over 6000 images was collected over six years in the field by trapping individual salamanders and photographing them (see Figure 7(b)). The scientific objective is to track their movement individually

[2]These filters were run to their iteration limit.

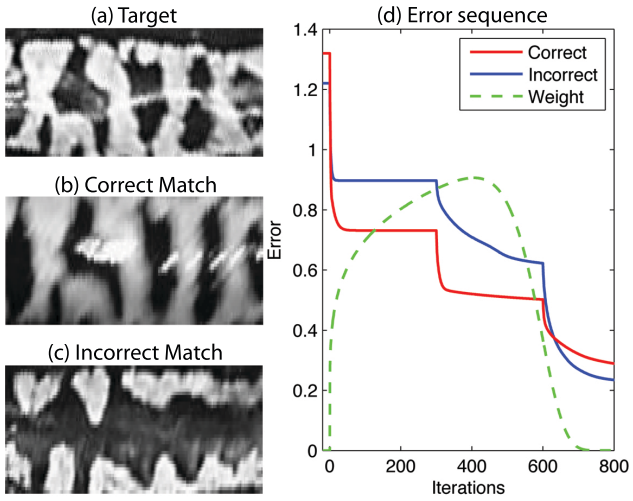

Figure 6. Error sequences generated by aligning (b) to (a) and (c) to (a) are shown in (d) in red and blue respectively. The scoring curve is shown in dashed green. Although (b) is the true match, it has higher initial and terminal error than (c).

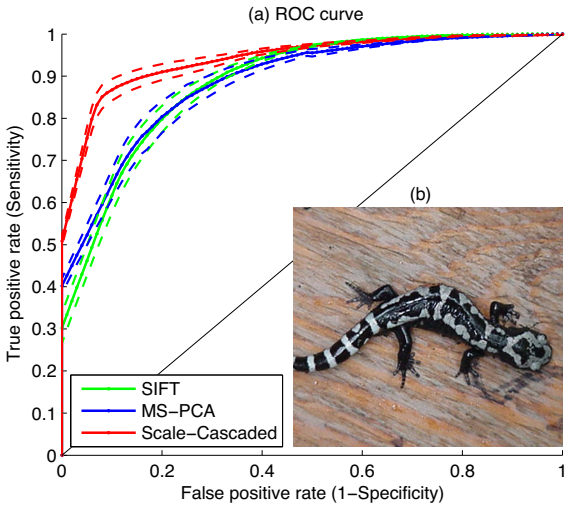

Figure 7. (a) ROC curve for deformation invariant matcher (red), MS-PCA (blue), and SIFT (green). The dashed lines show the variance for each method over 150 queries. (b) Salamanders have highly deformable bodies; images are varied in background and specular effects.

so that we may establish migratory patterns and thus develop appropriate conservation strategies. For our purpose, we need to index the database by matching individuals.

As shown in Figure 7(b), marbled salamanders have highly deformable bodies. The images themselves are blurry, noisy, and specular (see Figure 6). To solve the animal pose problem, Gamble *et al*. [8] artificially straighten the salamanders along the medial axis, removing large lower-order deformations. Then the authors extract patches between key features (the feet) and match them using multiscale PCA [8]. The ROC curve of this method is shown in blue in Figure 7(a), determined using relevance judgements provided by users who examined the top $N = 20$ retrievals over 150 queries.

The primary difficulties in improving the performance of their system are still the existence of specularity and nonlinear deformations (see Figure 6 and Figure 8). Specularity is difficult to explicitly marginalize in this application because the color of light and the color of the salamander are the same (see Figure 7(b)). Fortunately, specularity is localized and very high frequency deformations are necessary to "paint" out differences between images due to specularity. This suggests that we may be able to compensate further by down-weighting the contributions of large wavenumber deformations in W. The nonlinearity in pose remains as a difficulty but, due to "straightening", it does not occupy the high frequencies required to marginalize noise and specularity. Indeed, the application of SIFT features to the recognition problem on the straightened images produces a slightly worse result than MS-PCA (see Figure 7(a)), which further highlights the existence of local variability. We supposed that a mid frequency portion of the deformation spectrum is worth exploring and apply the scale-cascaded alignment with the scoring curve shown in Figure 6 (dotted green curve).

We used the score to re-rank the top 20 ranks for which relevance judgements were available using the same 150 queries. The ROC shows marked improvement. Figure 8 shows examples of two retrievals. In each case, we show the target, the top few MS-PCA retrievals (right) and the corresponding ranks of the deformation invariant matcher (left). The Label "Y01C1S2P932" indicates that the salamander is from year 2001. Thus, matches are found across many years. Although both SIFT and MS-PCA match well (as can be seen in Figure 7(a)), our scale-cascaded approach ranks the retrievals better by ignoring the errors corresponding to low and high frequency deformations.

5. Conclusion

We demonstrate a new approach to deformation invariant matching. Our approach retains the benefits of viscous alignment and introduces the selectivity necessary for recognition. It is a simply parameterized model, and can be efficiently implemented using spectral methods. It does not require explicit correspondence and can be easily tuned to prefer simpler or specific explanations of the template-target misfit. This allows us to mitigate the effects of nonlinear deformations in a recognition task, producing a clear and substantial improvement.

There are several areas to explore: (a) Explicitly incorporating a changing uncertainty C (Equation 1), as alignment progresses [3, 19]. (b) Examining ways to learn W,

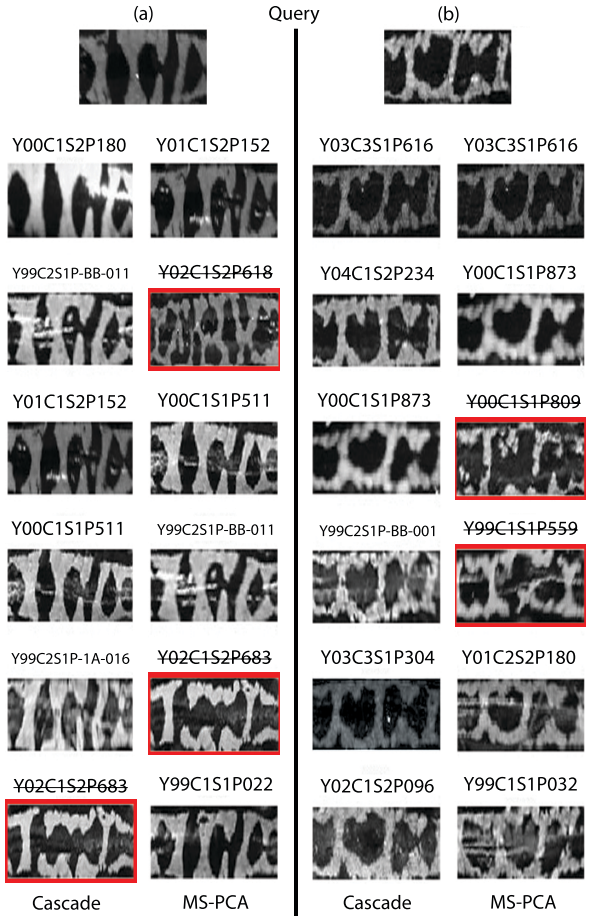

Figure 8. Two sample retrievals comparing the deformation invariant matcher and MS-PCA. The queries are marked (a) and (b) and the retrievals proceed in rank order down a column per method (self-matches are omitted and appear at rank 1). Mismatches are highlighted in red and their labels are crossed out.

and learn sparse-prior models for learning activations of $\mathcal{H}$ or the filter bank itself. (c) Using this approach to calculate deformation statistics better [3]. (d) Extending PCA-SIFT [15], and (e) reformulating our approach with mutual information measures [6, 23].

References

[1] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86(414):376–387, 1991.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *TPAMI*, 24(4):509–522, April 2002.

[3] G. Charpiat, O. Faugeras, and R. Keriven. Image statistics based on diffeomorphic matching. In *ICCV*, pages 852–857, 2005.

[4] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, Oct 1996.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *TPAMI*, 23(6):484–498, June 2001.

[6] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. In *MICCAI, LNCS 2489*, pages 541–548, 2002.

[7] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, C-22(1):67–92, Jan 1973.

[8] L. Gamble, S. Ravela, and K. McGarigal. Multi-scale features for identifying individuals in large biological databases: an application of pattern recognition technology in amphibian research. *Journal of Applied Ecology*, 45(1):170–180, Feb 2008.

[9] S. Geman. Invariance and selectivity in the ventral visual pathway. *Journal of Physiology (Paris)*, 100(4):212–224, Oct 2006.

[10] C. Gramkow and M. Bro-Nielsen. Comparison of three filters in the solution of the navier-stokes equation in registration. In *Proc. Scandinavian Conference on Image Analysis*, pages 785–802, 1997.

[11] D. J. Heeger. Optical flow using spatiotemporal filters. *IJCV*, 1(4):279–302, Jan 1988.

[12] A. Hill, C. Taylor, and A. Brett. A framework for automatic landmark identification using a new method of nonrigid correspondence. *TPAMI*, 22(3):241–251, Mar 2000.

[13] A. K. Jain, Y. Zhong, and M.-P. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71(2):109 – 129, 1998.

[14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, Jan 1988.

[15] Y. Ke and R. Suthankar. Pca-sift: a more distinctive representation for local image descriptors. In *CVPR*, volume 2, pages 506–513, 2004.

[16] H. Ling and D. W. Jacobs. Deformation invariant image matching. In *ICCV*, pages 1466–1473, 2005.

[17] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *TPAMI*, 15(6):580–591, 1993.

[18] G. V. Middleton and P. R. Wilcock. *Mechanics in the Earth and Environmental Sciences*. Cambridge University Press, 1996.

[19] S. Ravela, K. Emanuel, and D. McLaughlin. Data assimilation by field alignment. *Physica D: Nonlinear Phenomena*, 230(1-2):127 – 145, 2007.

[20] R. Szeliski and J. Coughlan. Hierarchical spline-based image registration. In *CVPR*, pages 194–201, Jun 1994.

[21] J.-P. Thirion. Fast non-rigid matching of 3D medical images. In *MRCAS*, Nov 1995.

[22] T. Vetter, M. J. Jones, and T. Poggio. A bootstrapping algorithm for learning linear models of object classes. In *CVPR*, pages 40–46, 1997.

[23] P. A. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[24] B. Widrow. The "rubber-mask" technique-II. pattern storage and recognition. *Pattern Recognition*, 5(3):199–211, 1973.