Spectral Control of Viscous Alignment for Deformation Invariant Image Matching

by

Christopher Minzer Yang

S.B., Massachusetts Institute of Technology (2008)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author ....
Department of Electrical Engineering and Computer Science
May 22, 2009

Certified by ....
Sai Ravela
Research Scientist
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Spectral Control of Viscous Alignment for Deformation Invariant Image Matching

by

Christopher Minzer Yang

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 2009, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

We present a new approach to deformation invariant image matching. Our approach retains the broad range of linear and nonlinear deformations that viscous alignment methods can model, but introduces a selectivity that is necessary for recognition. Our method models viscous kernels with an over-complete filter basis. The basis is parameterized with a single scalar parameter, the spectral radius r, which selects deformations ranging in complexity from tranlations to "turbulence." The spectral radius is used for cascaded alignment starting from low deformation frequencies and finishing with high deformation frequencies.

Cascaded alignment makes deformation invariant matching for recognition feasible and efficient. Because spectral radii map directly to deformation complexity, their contributions are selectively weighed to calculate the template-target similarity. In this way, our model can distinguish deformations by their relevance to recognition, without losing the flexibility of viscous alignment for handling nonlinear deformations. Our approach is applied to recognize flexible bodies of animals, and results indicate that the method is very promising.

Thesis Supervisor: Sai Ravela
Title: Research Scientist

Acknowledgments

First and foremost, I thank my advisor, Dr. Sai Ravela. Your rigor and your wisdom make you a researcher of the highest caliber, but it is your attention and your dedication that make you an advisor without equal. Thank you for all your help.

I would also like to acknowledge Megan Chesser and Dr. Kevin McGarigal at the Department of Natural Resources Conservation at the University of Massachusetts Amherst for all your work with the marbled salamander CMR data.

I would also like to acknowledge Josh Runge, my fellow MEng student, for all your help on coding the Sloop website.

I would also like to thank my planetary science office mates who (literally) taught me about the world. I wish you the best of luck!

I would like to express my deep gratitude to my closest friends without whom this year, and perhaps all years previous to it, would not have been worth it. My victories are sweetened and my burdens are lessened by your support and affection – thanks for everything.

Finally, I would like to acknowledge my parents. Your unerring and unending love and support are the greatest gifts I have ever received. I would not be where I am or who I am without you. Thank you so much.

This work was funded in part by NSF DBI 6898334.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Deformation invariant matching is the task of recognizing that two different images are of the same object despite some variability in pose, lighting, or viewpoint in the images. There is an inherent trade-off between selectivity (the accuracy of this recognition) and deformation invariance (how much – and what types of – variability we wish to allow). The more variability we accept between images that are "the same" (*i.e.* of the same object), the more likely we are to accidentally call two images of different objects equal, when they really should not be. Conversely, the stricter we are about variability in images, the more likely we are to label two images unequal even if they truly are the same [10].

The work presented here is new approach to deformation invariant matching. In broad terms, our approach measures the similarity between two images by measuring how "difficult" it is to morph one image into the other. Our method makes it simple to parameterize the types of warps allowed or disallowed, which makes the process of evaluating the "difficulty" of a warp more straightforward.

Image processing methods, like the one proposed here, can be applied to ecological problems. For example, algorithms to recognize animals from photographs greatly aid the study of migratory behavior, which can, in turn, aid the development of conservation plans [9]. However, a vast number of animals and plants deform in highly nonlinear ways, which presents a challenging and interesting object recognition problem.

Figure 1-1: The marbled salamander (*Ambystoma opacum*) has a unique marbling pattern on its back. These images demonstrate the variability in patterning and background of images taken from a CMR study.

1.1 Application Motivation

Capture-Mark-Recapture (CMR) studies provide essential information on demography, movement and other ecological characteristics of rare and endangered species [9]. This information is required by conservation managers to focus their strategies on the most relevant threats and life stages, to identify critical habitat areas, and to develop benchmarks for measuring success in recovery plans.

CMR involves capturing the animal of interest, physically marking or tagging it, and releasing it. Once the animal is marked, researchers can track its movement as it is recaptured throughout the study. CMR studies that use physical marking or tagging techniques are intrusive to varying degrees and may even affect the fate or behavior of the creature [9].

Because of this possibility, photographing the animals is sometimes the only realistic technique, but manual photo-identication is not scalable as the image catalog grows; computer-aided recognition has mostly been conned to large-bodied animals [9]. For example, whale shark[1] (*Rhincodon typus*) movements are tracked via semi-automated comparison of amateur photographs. The whale shark system

[1] `http://www.whaleshark.org/`

matches manually identied spots using their spatial geometries [14]. While very useful, such an approach, we argue, is limited in scope, especially when animal surfaces are textured or contain arbitrary patternation that does not lend easily to geometric description.

The marbled salamander (*Ambystoma opacum*) is a creature that is the subject of CMR studies that has just such arbitrary patternation. The salamanders' dorsal patterns can be used as ad hoc fingerprints. Figure 1-1 shows a few marbled salamanders images taken from a CMR study. Ravela applied his pattern recognition algorithm in an effort to automate recognition of the marbled salamander [9]. The recognition algorithm ranks all images in a database against each other using a computational metric to model perceptual visual similarity; images that are the most similar to each other in this way are proposed to the user as the same individual individual. Using this process, then, individuals can be tracked through the study, as they are captured, photographed, and indexed.

Our work is an effort to improve on the performance on this recognition system. The original recognition algorithm is faster but less accurate than our proposed matcher. Eventually, the matcher will be folded into the overall recognition system as an even-finer filter on recognition results.

1.2 Problem Formulation

Two primary difficulties arise in adapting current recognition techniques to animal or plant targets: (1) No succintly parameterized models exist that can handle the broad range of deformations found in nature. (2) As model capacity for nonlinear deformations grows, there is a corresponding loss of selectivity because deforming perceptually unrelated objects into one another also becomes easy [10].

Our work assumes that objects can be recognized by matching their images. We develop an efficiently parameterized deformation invariant matcher. We are particularly motivated by viscous alignment methods [3, 4, 11, 25] that can handle very nonlinear deformations. Prototypically, many such approaches can be viewed as a

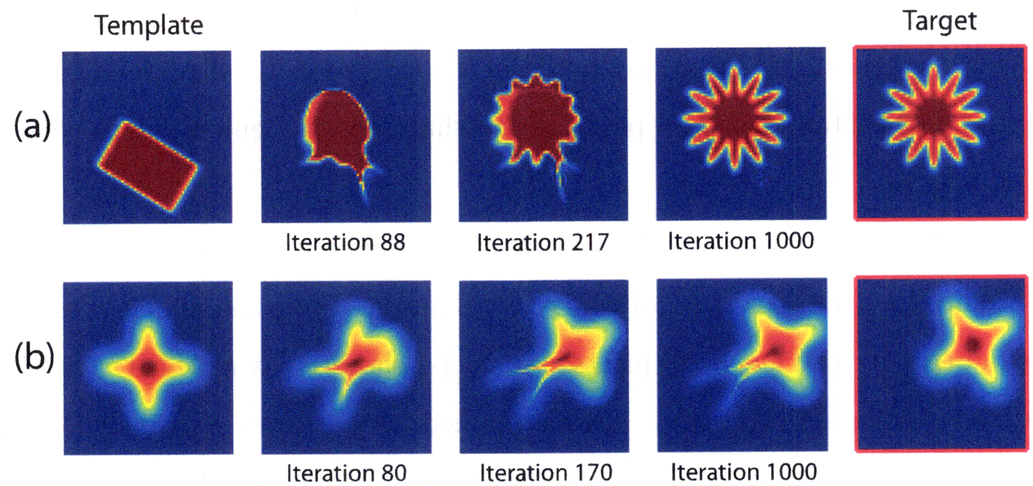

Figure 1-2: (a) Viscous alignment can be used to deform the rectangle completely into the "flower." After the alignment is complete, the template is indistinguishable from the target, which can lead to a loss of perceptual selectivity. The intermediate snapshots during iterations of the alignment are also shown. (b) Viscous alignment can produce highly nonlinear deformations to explain simpler transformation. A translated and rotated "cross" prompts nonlinear transport, as the iteration sequences show. The loss of selectivity and control of alignment prevents this approach from being immediately useful for deformation invariant image matching.

minimization problem of the form:

$$J(\mathbf{q}) \doteq \frac{1}{2} \|Y - X \circ \mathbf{q}\|^2_{\underline{\underline{R}}} + L(\mathbf{q}) \tag{1.1}$$

X and Y are the template and target images respectively, $\underline{\underline{R}}$ is a covariance defining the norm, $\mathbf{q}$ is a deformation vector field, and $X \circ \mathbf{q} \doteq X(\underline{p} - \mathbf{q}(\underline{p}))$ is the deformation of the template by the vector field $\mathbf{q}$. Regularizing constraints are expressed by $L(\mathbf{q})$. If $L(\mathbf{q})$ express smoothness and non-divergence constraints, the Euler-Lagrange equation of this objective leads to a nonlinear PDE that can be solved iteratively [4]. Because smoothness and non-divergence constraints can be viewed as viscous constraints in an evolving fluid, this approach is termed *viscous alignment* or *fluid alignment*. Although only locally convergent, such viscous models can produce very complex deformations without explicit feature point correspondence and even with sparse measurements [25]. Thus whole images or image patches around distinguished locations can be aligned.

However, the issues of selectivity and parameterization prevent us from converting viscous alignment into a deformation invariant matcher for recognition. Illustrated in Figure 1-2(a), we see snapshots of a rectangle deforming (iteratively) into a flower. After convergence, the rectangle is identical to the flower; clearly the result of alignment will not allow us to distinguish between the two. This loss of selectivity cannot be naïvely resolved by stopping the alignment before convergence because intermediate states may bear no meaningful resemblance to the template or the target. Example (b) in Figure 1-2 shows how a complex deformation is generated when only a simple translation and rotation is required. There is no easy way to control viscous alignment to produce this simpler solution. We need an effective way to constrain the viscous model without losing the ability to deform over a broad range. We must then be able to use such a model to match images.

1.3 Proposed Approach

These issues are solved with our proposed approach. Our method, akin to multiscale texture decomposition, uses a spectral representation of deformations [19]. Low wavenumber deformation fields are smooth and global; higher wavenumbers are turbulent and local. This decomposition is produced by using an over-complete filter basis to approximate the regularizing kernel $L(\mathbf{q})$ in the objective (Equation 1.1) in viscous alignment. This basis is parameterized with the spectral radius r, that controls the deformation solution. At $r = 0$, only translations are allowed. At $r = 1/2$ cycles, global affine deformations are admitted. At $r = 1/2$ cycles/pixel, deformation solutions are turbulent.

Images are aligned in a cascaded manner. Higher-frequency deformations are produced subject to the convergence of filters at smaller spectral radii. This approach retain viscous alignment's full deformation power because the entire spectrum can be still be represented. Unlike viscous alignment, however, we can now decompose the reduction in error between template and target by deformation complexity. Thus, perceptually matching images becomes possible, because we can selectively weigh errors that are associated at specific frequency bands of the deformation spectrum. This leads to a perceptually relevant deformation invariant matcher.

To the best of our knowledge, such an approach to deformation invariant matching has not been demonstrated using viscous alignment before. This simply parameterized deformable model is powerful and can be used for matching without detecting corresponding features and possibly with sparse measurements. Results from applying our technique to the marbled salamander CMR data indicate that our approach substantially improves the previous technique's recognition performance.

1.4 Overview

The remainder of this thesis is organized as follows. In Chapter 2 we describe existing work relating to current viscous alignment techniques and deformation invariant matching. In Chapter 3, we will develop our version of viscous alignment. In Chapter 4, we explain how to build a deformation invariant matcher from our formulation of viscous alignment and apply the method to biological image retrieval. Finally, we conclude in Chapter 5 with a summary and discussion of the technique, and possible avenues of future work.

(As an aside, Table 1.1 summarizes the mathematical notation used for the rest of this thesis.)

x	Scalar
$\underline{x}$	N-D column vector
$\underline{x}^T$	N-D row vector
$\mathbf{x}$	Vector field
$\underline{\mathbf{x}}$	Column-rasterized vector field *e.g.* $\begin{pmatrix} \underline{x}^a \\ \underline{x}^b \end{pmatrix}$
X	Scalar field
$\underline{X}$	Column-rasterized scalar field
$\underline{\underline{X}}$	Matrix
$\doteq$	Definition

Table 1.1: Notation summary.

Chapter 2

Related Work

2.1 Deformable Template Matching

Processing (*e.g.* registration, segmentation, matching) using model-based shape matching is a well-studied technique in computer vision. While early research was focused on rigid shape matching (where deformations were simple translations, scaling, rotation, and affine movements), their usefulness is limited because of that rigidity constraint. *Deformable* templates, in contrast, use a flexible template. This flexibility makes these techniques much more versatile and much more capable of dealing with in-class shape deformations and variations [16]. Figure 2-1 summarizes the categories of template-based object matching techniques.

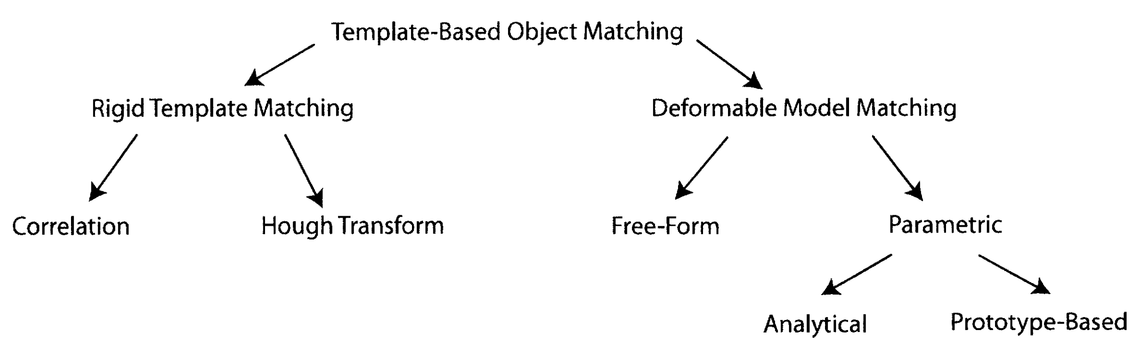

Figure 2-1: Template-based object matching technique summary. The proposed scale-cascaded alignment can be viewed as an example of free-form deformable model matching. This figure follows after [16].

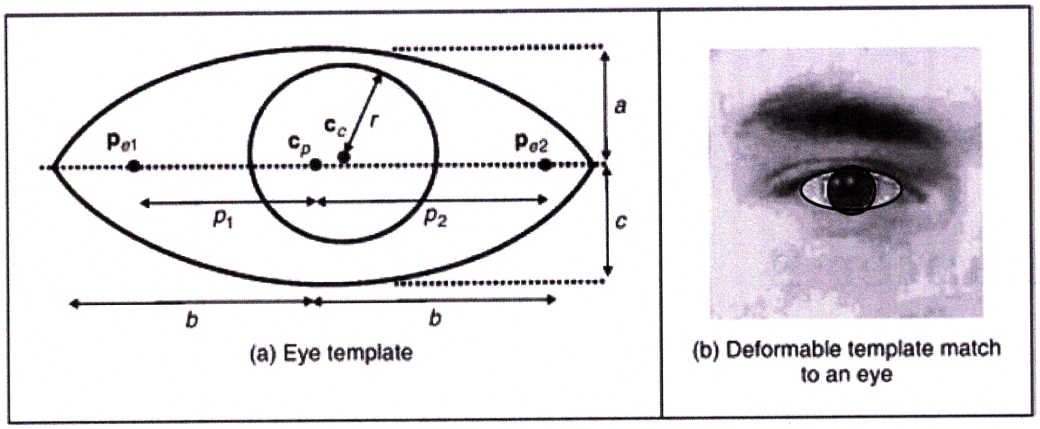

Figure 2-2: An example of analytical parameter template modeling for the task of eye recognition. (a) An eye template model is constructed using two parabolas and a circle. (b) The template parameters are changed until they match an example image. From [24].

2.1.1 Parametric Deformable Templates

Parametric techniques fall into two categories: analytical and prototypical. Analytical techniques rely on parameterizing the template as one or more analytical shapes (*e.g.* splines, circles, ellipses). The template is dynamically changed into a query image by adjusting the parameters of those analytical shapes (*i.e.* increasing the radius of the circle or moving the control points of a spline) in response to query image forces. To match an image, the template parameters are changed until the best match can be found. If the optimal parameters for this image are within a designated "acceptable" range, the image is a match. Figure 2-2 shows how an eye can be modeled using a combination of simple shapes, and adapted to an image of a real eye. Widrow's rubber masks [32] and Fischler and Elschlager's spring-loaded templates [8] were the first examples of such an analytical technique.

Prototypical template matching involves comparison with a so-called "standard," "prototype," or "generic" template image of the relevant class of objects [16]. All images are parameterized in relation to this exemplar template; varying parameters generates new shapes. Techniques based on principal component analysis (*e.g.* [29]) are prototype-based methods. The mean image of some training set is taken to be the prototype image. All other images are parameterized as the weighting of the top

N eigenvectors of that training set's singular value decomposition.

The approach proposed does not assume a shape prior (which would be difficult to use to represent texture) or a prototypical template image. Our alignment method falls into the category of free-form deformable models.

2.1.2 Free-Form Deformable Models

Free-form deformable models do not use a shape prior but rather local constraints on the deformation to regularize the search. Kass *et al.*'s snakes [17] is arguably the first example of such models. A spline (the "snake") is pushed by external constraints and forces from the image (*i.e.* image gradients) toward image features like edges, lines, and contours. The snake dynamically alters its shape and position, trying to seek the minimal energy state. Kass *et al.* defines the snake parametrically as $v(s) = (x(s), y(s))$, and its energy as:

$$E_{\text{total}} = \frac{1}{2} \int_0^1 \left(E_{\text{internal}} + E_{\text{image}} + E_{\text{constraint}} \right) ds \tag{2.1}$$

E_{internal} is the internal energy of the spline:

$$E_{\text{internal}} = \alpha(s) \left| \frac{\partial v}{\partial s} \right|^2 + \beta(s) \left| \frac{\partial^2 v}{\partial s^2} \right|^2 \tag{2.2}$$

The settings α and β control the stretching and flexing (respectively) the spline is allowed – essentially, how nonlinear the spline can be. This is somewhat analogous to our approach's spectral radius parameterization, which similarly controls how global or local a transformation can be, although the formulation is completely different.

E_{image} is the energy of the image along the path of the spline. If we set

$$E_{\text{image}} = w_1 I(x, y) \tag{2.3}$$

the snake will be attracted to light or dark lines, depending on the sign of w_1. If we set

$$E_{\text{image}} = w_2 \left| \nabla I(x,y) \right|^2 \tag{2.4}$$

the snake will be attracted to or repelled by edges, depending on the sign of w_2. $E_{\text{constraint}}$ is used to express high-level, user-defined "springs" (attracting points) and "volcanoes" (repelling points) [17]. Snakes give us a framework in which to solve various low-level problems, including segmentation and edge detection [17].

Snakes are just one example of such dynamically deforming methods (so-called "active models"); others include [6, 13, 22, 27]. These models all tend to be low-dimensional in the types of deformations they can produce or describe. Our cascaded viscous alignment, in contrast, is easily parameterized and can encompass deformations between simple translation to turbulence.

2.2 Correspondence-Based Deformation

Correspondence-based deformation warps images by corresponding a small set of features and constructing a global warp based on that correspondence. In [30] and [5] dense correspondence fields on two images are used to compare their similarity. Belongie *et al.* obtained excellent results on digit recognition using the *shape contexts* of feature points [2]. The method proposed here, however, uses no feature correspondence and thus is applicable for whole image matching. Alternatively, this method can be adapted to use correspondences by applying the matcher to the neighborhood of potential correspondences.

2.3 Viscous Image Alignment

Christensen has done much work on using a viscous fluid model as a regularizer for nonlinear image deformation for registration [4]. Using viscous constraints allow large distance, nonlinear kinematics that would be impossible using regularization methods based on linear elasticity or thin plates. Christensen uses a Navier-Poisson Newtonian fluid model as his viscous constraint because it allows for large, nonlinear deformations while maintaining a continuous homeomorphic map with smooth deformations of the template [4]. Templates are aligned to query images, for each image point $\underline{x}$ in the unit cube, by iteratively solving:

$$\mu\nabla^2\mathbf{q}(\underline{x}) + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{q}(\underline{x})) + \mathbf{f} = \underline{0} \tag{2.5}$$

where $\mathbf{q}$ is the deformation vector field (the instantaneous velocity field), μ and λ are viscosity constants, and $\mathbf{f}$ is the "body force" generated by the difference between evolving template and target. Christensen solved the objective given in Equation 2.5 using successive-over-relaxation (SOR) [4]. His computational techniques were improved by Gramkow and Bro-Nielsen's work using convolution filters [11].

We can relate Christensen's formulation to our prototypical alignment objective (Equation 1.1) by setting the following:

$$\begin{aligned} \frac{\partial L(\mathbf{q})}{\partial \mathbf{q}} &\doteq \mu\nabla^2\mathbf{q} + (\lambda+\mu)\nabla(\nabla\cdot\mathbf{q}) & (2.6)\\ \frac{\partial}{\partial \mathbf{q}}\left(\|Y - X\circ\mathbf{q}\|^2_{\underline{\underline{R}}}\right) &= \mathbf{f} & (2.7)\end{aligned}$$

(This formulation is developed more fully in Chapter 3.) Our work shares his viscous model but differs from previous work in this area in several key ways.

First, our spectral solution to the objective PDE improves both [4] and [11]. The spectral interpretation of the fluid kernel is related with Heeger's spatio-temporal filters [12], but the deformable model formulation is different. Second, our Gabor basis approximation of the viscous kernel is new. Thirion [28] uses a Gaussian to regularize, but there is no connection to image matching, and the filters proposed here use higher order (Gabor) filters. Finally, our cascaded approach jumps the gap between alignment and matching. Previous viscous approaches were developed for registration, not recognition [1].

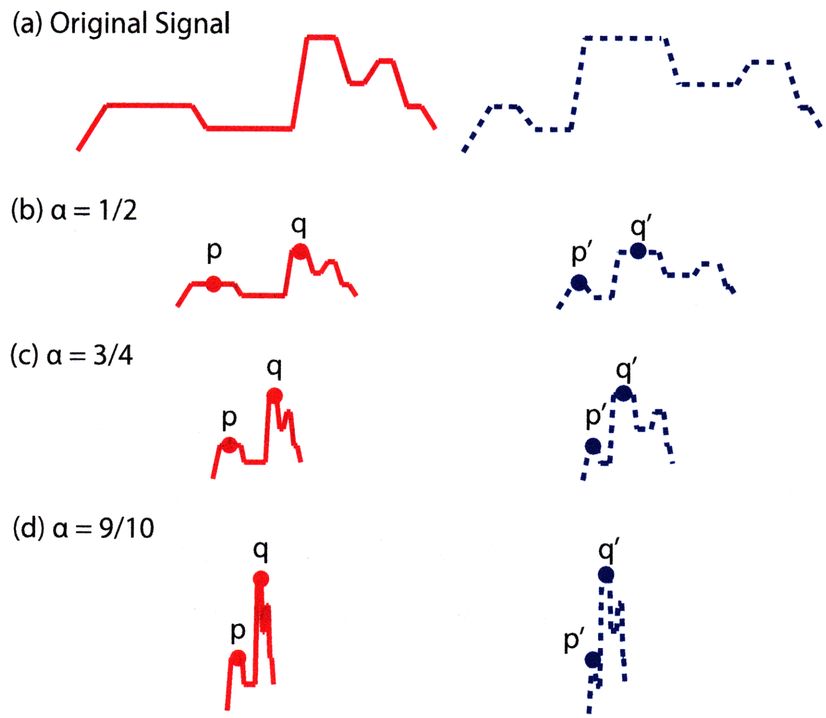

Figure 2-3: Geodesic distance is invariant to deformation. (a) shows the original 1D signals (intensity is height). The signal on the right is a deformed version of the signal on the left. The curve (p', q') is the deformed version of the curve (p, q). (b)–(d) show the effect of the embedding for increasing values of α. As α approaches 1, the length of the curve (p, q) approaches the length of the curve (p', q'). This is the geodesic distance, and it is deformation invariant. Figure follows after [21].

2.4 Deformation Invariant Matching

The most closely related work that we draw motivation from is that of Ling and Jacobs, who use the *geodesic-intensity histogram* (GIH) as a local descriptor that is deformation invariant [21]. The (2D) images of comparison are embedded in a 3D space where the intensity of the image is the z dimension. Let $I_1(x, y)$ be a 2D image and let $I_2(u, v)$ be a deformed version of I_1. Because deformation is homeomorphic (and thus invertible), we can write each image's coordinates in terms of each other. Thus, we say $I_2(u, v) = I_1(x(u, v), y(u, v))$.

We then embed both images in a 3D space. Let σ_1, σ_2 be the embeddings of I_1 and I_2 respectively. We write:

$$\sigma_1 = (x' = (1-\alpha)x, y' = (1-\alpha)y, z' = \alpha I_1(x,y)) \tag{2.8}$$

$$\sigma_2 = (u' = (1-\alpha)u, v' = (1-\alpha)v, w' = \alpha I_2(x,y)) \tag{2.9}$$

α is the *aspect weight*, which controls the weight on the intensity compared to the weight on the image coordinate – as α approaches 1, the less significant the spatial differences are and the more significant intensity differences become [21]. Figure 2-3 illustrates this for a simple 1D example. As α approaches 1, the differences in intensity dominate the arc length between the two points on both curves.

Let γ_1 be a curve on σ_1, and let γ_2 be the deformed version of γ_1 on σ_2. Ling and Jacobs show that as $\alpha \to 1$, the arc lengths of σ_1 and σ_2 become the same, *regardless of the deformation relating* I_1 *and* I_2. Therefore the geodesic distance, the distance of the shortest path between two points on the embedded surfaces, is deformation invariant [21]. They use this to build a geodesic-intensity histogram descriptor that is deformation invariant.

There are similarities to the effect of their invariance parameter α and our spectral radius parameter r, but both the representations and problem formulations are entirely different.

Chapter 3

Scale-Cascaded Alignment

Our objective is to exploit the power of viscous alignment for deformation invariant matching. The main obstacle to achieving this goal is that there is no way to control the complexity viscous alignment's solution. As we saw earlier, viscous alignment can generate very complex deformations when simpler ones would work just as well. Intuitively, we would like simpler deformation possibilities to be exhausted before we search for more complex ones.

We solve this problem with *scale-cascaded alignment*. We interpret the viscous constraints of alignment as a filter in the frequency domain. This is useful because we will show that spectral radius is directly related to deformation complexity. The lower the frequency, the "simpler" the solution, and the higher the frequency the more "complicated" the solution. If we only allow a certain frequency of the viscous filter through, then the deformations generated will be constrained to be of a certain "complexity" – this is the key to controlling viscous alignment! Alignment proceeds as a continuation over filter frequency: we allow alignment to "use" the high frequencies of the filter only after alignment using lower frequencies has converged. Thus, "simpler" deformations are produced before "complex" ones.

We develop SCA in four stages. First, we review a prototypical method for viscous alignment, for context and motivation. Second, we spectrally reinterpret the underlying equations of motion. Third, we introduce the spectral radius as a simple control for deformation and analyze precisely how spectral radius relates to deformation field complexity. Finally, we present and discuss our cascaded approach to alignment in detail.

3.1 Viscous Alignment

A 2D template image X is aligned to the target image Y, and both images are discretized on domain Ω. Let $\underline{p} = (x, y)^T \in \Omega$ be the discrete position in a gridded field, and let $\mathbf{q}$ be a continuous displacement vector field. $X(\underline{p} - \mathbf{q}(\underline{p}))$ is used to denote the image X displaced by field $\mathbf{q}$.[1]

We seek the deformation field $\mathbf{q}$ that maximizes the a posteriori probability $P(\mathbf{q}|X, Y)$. Using Bayes' rule we write:

$$P(\mathbf{q}|X, Y) \propto P(Y|X, \mathbf{q})P(X)P(\mathbf{q}) \tag{3.1}$$

The right-hand side of Equation 3.1 consists of (1) the data likelihood $P(Y|X, \mathbf{q})$, (2) an amplitude prior $P(X)$ which is independent of displacements, and (3) a displacement prior $P(\mathbf{q})$. We will suppose that these component densities are all Gaussian and thus produce a quadratic objective:

$$\begin{aligned} J(\mathbf{q}) = {} & \frac{1}{2}\sum_{\underline{r}\in\Omega}\sum_{\underline{s}\in\Omega}\{[Y(\underline{r}) - X(\underline{r} - \mathbf{q}(\underline{r}))]\, C(\underline{r}, \underline{s})\, [Y(\underline{s}) - X(\underline{s} - \mathbf{q}(\underline{s}))]\} \\ & - \ln P(X) + L(\mathbf{q}) \end{aligned} \tag{3.2}$$

Here C is the field associated with the matrix $\underline{\underline{C}}$, the inverse of covariance of the likelihood. We assume that $\underline{\underline{C}}$ is static with respect to the deformation field for the optimization. The displacement prior is based on an energy function $L(\mathbf{q})$ modeled with divergence and non-smoothness penalties [25]:

$$L(\mathbf{q}) = \frac{w_1}{2}\sum_{\underline{z}\in\Omega}[\nabla\mathbf{q}(\underline{z})^T\nabla\mathbf{q}(\underline{z})] + \frac{w_2}{2}\sum_{\underline{o}\in\Omega}[\nabla\cdot\mathbf{q}(\underline{o})]^2 \tag{3.3}$$

[1] Since the displacement field $\mathbf{q}$ is real-valued, $X(\underline{p} - \mathbf{q}(\underline{p}))$ may be evaluated using interpolation.

The Euler-Lagrange equation of the objective (Equation 3.2) is a highly nonlinear PDE, so we solve iteratively. At iteration i, we solve the following:

$$\begin{aligned}
X_i(\underline{p}) &\leftarrow X(\underline{p} - \mathbf{q_{0:i-1}}(\underline{p})) && (3.4)\\
\delta \underline{X}_i &\doteq \underline{\underline{C}}\left[\underline{Y} - \underline{X}_i\right] && (3.5)\\
\frac{\partial L(\mathbf{q_i}(\underline{r}))}{\partial \mathbf{q_i}(\underline{r})} &= \nabla X_i(\underline{r}) \delta \underline{X}_i(\underline{r}) && (3.6)\\
&\doteq \mathbf{f_i}(\underline{r}) && (3.7)
\end{aligned}$$

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 column 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 and at point $\underline{r}$, by fixing $\mathbf{f_i}$, we have a linear system:

$$\begin{aligned}
w_1 \nabla^2 \mathbf{q_i}(\underline{r}) + w_2 \nabla(\nabla \cdot \mathbf{q_i}(\underline{r})) - \mathbf{f_i}(\underline{r}) &= 0 && (3.8)\\
\underline{\underline{G}}\underline{\mathbf{q_i}} &= \underline{\mathbf{f_i}} && (3.9)
\end{aligned}$$

Here $\underline{\underline{G}}$ is the sparse matrix representing the differential operators, and the column vectors $\underline{\mathbf{q_i}}$ and $\underline{\mathbf{f_i}}$ are obtained from rasterizing their corresponding fields.

Christensen solved Equation 3.8 using SOR [4], which was improved using convolution filters [11], 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 results are excellent and may be combined with pyramid approaches for even better computational efficiency [11].

The elegance and utility of this method lies in the fact that $X(\underline{p} - \mathbf{q}(\underline{p}))$ is not linearized as in optic flow.[2] Complex deformations are thus produced by advecting the evolving deformation field with the instantaneous displacement field at each it-

[2]See Appendix A for an in-depth discussion of how optical flow relates to viscous alignment.

eration (and optionally with restarts, see [4]). The solution, to be sure, is still local. Nevertheless, we, as have many others, seen deformations of amazing complexity with no correspondences whatsoever.

This should not be surprising really because Equation 3.8 represents Navier's equation in equilibrium [23], with an image-driven body force. If we drop the Laplacian term $\nabla^2 \mathbf{q_i}(\underline{p})$, it is a proper (inertia-less) fluid. If we drop the continuity term $\nabla(\nabla \cdot \mathbf{q_i}(\underline{p}))$, 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.

3.2 Spectral Interpretation

The spectral interpretation 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, so we can rewrite the viscous alignment objective (Equation 3.8) as:

$$w_1 \nabla^2 \mathbf{q}(\underline{p}) + w_2 \nabla(\nabla \cdot \mathbf{q}(\underline{p})) = \mathbf{f}(\underline{p}) \tag{3.10}$$

Each grid position has two components $\underline{p} \doteq (x, y)^T$ and so too do the fields $\mathbf{q}(\underline{p})$ and $\mathbf{f}(\underline{p})$. Let us define them as $\mathbf{q}(\underline{p}) \doteq (Q^x(\underline{p}), Q^y(\underline{p}))^T$ and $\mathbf{f}(\underline{p}) \doteq (F^x(\underline{p}), F^y(\underline{p}))^T$.

We rewrite the instantaneous objective in Equation 3.10 in terms of its components (we drop the dependence on $\underline{p}$ for clarity):

$$w_1 \left(\frac{\partial^2 Q^x}{\partial x^2} + \frac{\partial^2 Q^x}{\partial y^2} \right) + w_2 \left(\frac{\partial^2 Q^x}{\partial x^2} + \frac{\partial^2 Q^y}{\partial yx} \right) = F^x \tag{3.11}$$

$$w_1 \left(\frac{\partial^2 Q^y}{\partial x^2} + \frac{\partial^2 Q^y}{\partial y^2} \right) + w_2 \left(\frac{\partial^2 Q^x}{\partial xy} + \frac{\partial^2 Q^y}{\partial y^2} \right) = F^y \tag{3.12}$$

We define a wavenumber space $\underline{\omega} \doteq (m, n)^T$, and the following Fourier pairs:

$$\mathbf{q}(\underline{p}) \doteq (Q^x(\underline{p}), Q^y(\underline{p}))^T \quad \leftrightarrow \quad \mathcal{Q}(\underline{\omega}) \doteq (\mathcal{Q}^x(\underline{\omega}), \mathcal{Q}^y(\underline{\omega}))^T \tag{3.13}$$

$$\mathbf{f}(\underline{p}) \doteq (F^x(\underline{p}), F^y(\underline{p}))^T \quad \leftrightarrow \quad \mathcal{F}(\underline{\omega}) \doteq (\mathcal{F}^x(\underline{\omega}), \mathcal{F}^y(\underline{\omega}))^T \tag{3.14}$$

We can rewrite the componentized objective equations (Equations 3.11 and 3.12) in Fourier space (omitting the dependence on $\underline{\omega}$ for clarity):

$$w_1\left(-m^2\mathcal{Q}^x - n^2\mathcal{Q}^x\right) + w_2\left(-m^2\mathcal{Q}^x - mn\mathcal{Q}^y\right) = \mathcal{F}^x \tag{3.15}$$

$$w_1\left(-m^2\mathcal{Q}^y - n^2\mathcal{Q}^y\right) + w_2\left(-mn\mathcal{Q}^x - n^2\mathcal{Q}^y\right) = \mathcal{F}^y \tag{3.16}$$

We can rewrite this system of equations in matrix form:

$$\underline{\underline{\mathcal{G}}} \begin{bmatrix} \mathcal{Q}^x \\ \mathcal{Q}^y \end{bmatrix} = \begin{bmatrix} \mathcal{F}^x \\ \mathcal{F}^y \end{bmatrix} \tag{3.17}$$

where

$$\underline{\underline{\mathcal{G}}} = \begin{bmatrix} -w_1(m^2+n^2) - w_2m^2 & -w_2mn \\ -w_2mn & -w_1(m^2+n^2) - w_2n^2 \end{bmatrix} \tag{3.18}$$

This is easy (and exact) to invert (for $m \neq 0$ and $n \neq 0$). The determinant of $\mathcal{G}$ is:

$$\begin{aligned} \det \underline{\underline{\mathcal{G}}} &= \left[-w_1(m^2+n^2) - w_2n^2\right]\left[-w_1(m^2+n^2) - w_2m^2\right] - (-w_2mn)^2 \\ &= w_1^2(m^2+n^2)^2 + w_1w_2m^2(m^2+n^2) + w_1w_2n^2(m^2+n^2) \\ &\quad + w_2^2m^2n^2 - w_2^2m^2n^2 \\ \det \underline{\underline{\mathcal{G}}} &= (w_1^2 + w_1w_2)(m^2+n^2)^2 \end{aligned} \tag{3.19}$$

Finally, we can write:

$$\begin{bmatrix} \mathcal{Q}^x \\ \mathcal{Q}^y \end{bmatrix} = \begin{bmatrix} \mathcal{H}^a & \mathcal{H}^b \\ \mathcal{H}^b & \mathcal{H}^c \end{bmatrix} \begin{bmatrix} \mathcal{F}^x \\ \mathcal{F}^y \end{bmatrix} \tag{3.20}$$

where

$$\mathcal{H}^a = \frac{-w_1(m^2+n^2) - w_2 n^2}{(w_1^2 + w_1 w_2)(m^2+n^2)^2} \tag{3.21}$$

$$\mathcal{H}^b = \frac{w_2 mn}{(w_1^2 + w_1 w_2)(m^2+n^2)^2} \tag{3.22}$$

$$\mathcal{H}^c = \frac{-w_1(m^2+n^2) - w_2 m^2}{(w_1^2 + w_1 w_2)(m^2+n^2)^2} \tag{3.23}$$

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). Thus, the filter $\mathcal{H}^b = 0$, and:

$$\mathcal{H}^p = \mathcal{H}^a = \mathcal{H}^c = -\frac{1}{w_1(m^2+n^2)} \tag{3.24}$$

Equation 3.24 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 nonlinear (as a function of the grid) deformations because the instantaneous deformation field $\mathcal{Q}$ can have all frequencies (see Figure 1-2 and Figure 3-1).

While this is useful for alignment, high frequency deformations may undesirably arise even where the solutions are apparently "simpler," as shown in Figure 1-2. Such a broadband response is not selective enough for recognizing objects by matching their images. Although no perceptual basis for modeling deformations as Navier's equation (or Laplace-Beltrami) has been shown, can we leverage power-law spectra to gain selectivity for recognition without losing its kinematic range?

3.3 Deformation Filters

To answer the question, let us look at the viscous formulation under discussion. As a scalar variable, w_1 only controls the convergence rate (provided stability is maintained, see [25]) but not the shape of the spectrum. We may choose w_1 to be anisotropic and space-varying, but that is difficult to design. So as a first step,

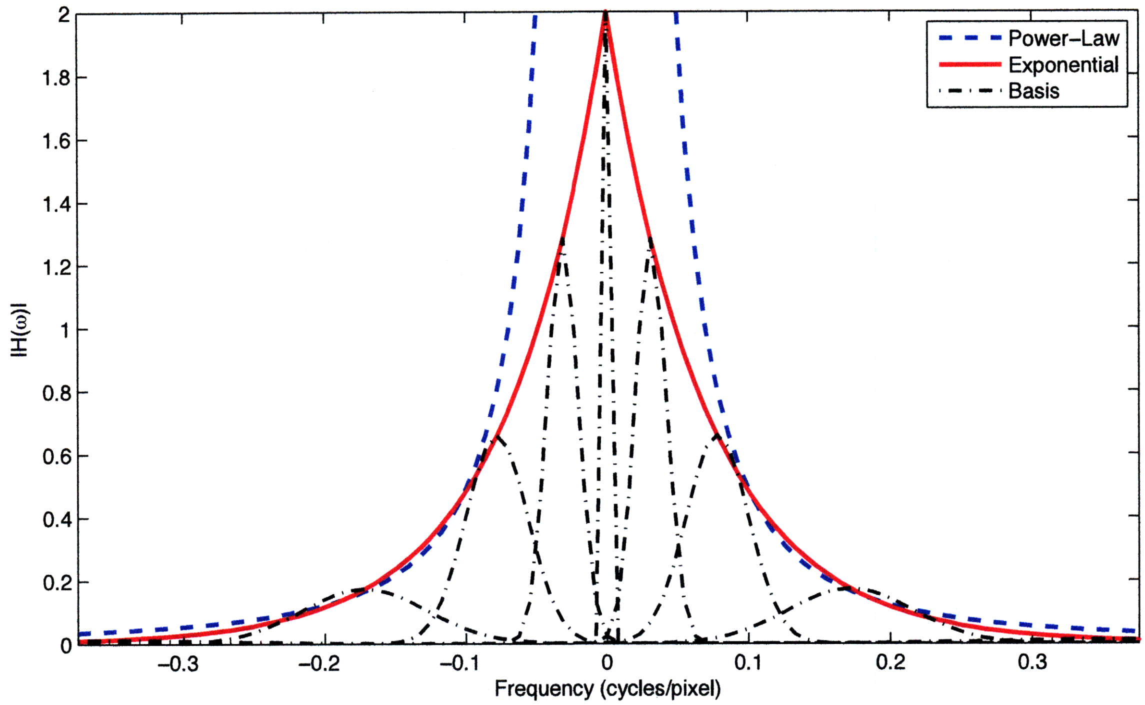

Figure 3-1: An exponential envelope (red) approximates the power-law envelope (blue) from Equation 3.24. The basis filters $\mathcal{H}$ are attenuated by the exponential envelope, shown here in 1D.

we build a tuner to control the complexity of deformations – from translations to turbulence.

3.3.1 Laplacian Envelope

Let us approximate the filter $\mathcal{H}^p$ (Equation 3.24) with the Laplace "distribution":

$$e(r) = -\beta e^{-|r|/2\alpha^2} \tag{3.25}$$

where $r = \sqrt{m^2 + n^2}$, which overcomes the singularity in the power law and can be adapted to many different spectral profiles. The parameter β can be interpreted as controlling the gain of the filter, and α controls the bandwidth. A comparison of the original filter and the tunable approximation is shown in Figure 3-1.

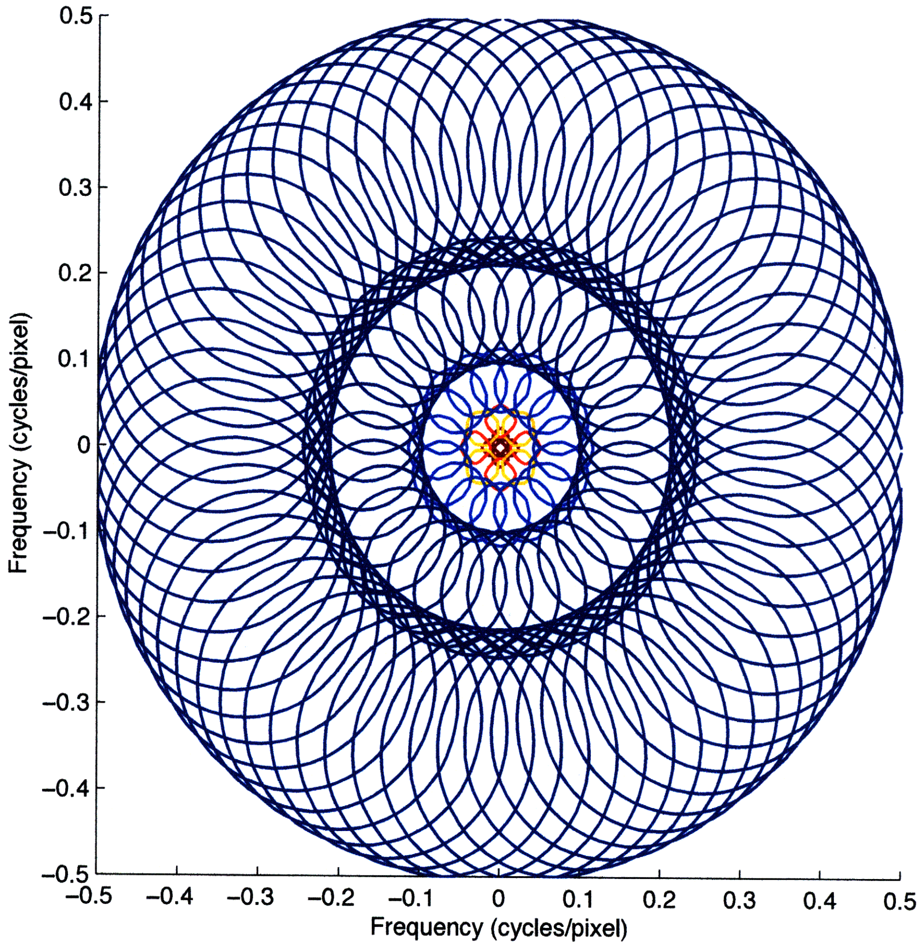

Figure 3-2: The 1-σ contours of the filters in the Gabor filter bank in frequency domain. Each ring uses twice the number of filters than the previous ring. Similarly, each ring also doubles σ.

3.3.2 Building the Basis

The value of representing texture spectra using multi-scale filters is well understood [19]. Similarly, the deformation spectrum is decomposed into multiple scales, here using a Gabor basis. Figure 3-1 shows the basis. The power (peak) of any filter in this basis is constrained by the corresponding power of the Laplace approximation (Equation 3.25) at its location in the frequency spectrum. This intuition is further developed in the rest of this section.

Let us consider rings labelled R_0 to R_N in $\underline{\omega}$. N is designed to be logarithmic in the size of the image. In each ring R_i at a radius $r_i = r(R_i)$ from the origin, we place n_i Gaussians G_j, $j = 1 \ldots n_i$ azimuthally at θ_j and all with scale σ_i. Each Gaussian is parameterized as $G(r, \theta, \sigma)$. The filter bank at ring R_i is thus:

$$\mathcal{H}_i = \mathcal{H}(R_i) = \frac{1}{\aleph_i} \sum_{j=1}^{n_i} G(r_i, \frac{2\pi}{n_i} j, \sigma_i) \tag{3.26}$$

where $\aleph_i$ is a normalizing constant. The filter $\mathcal{H}_0$ is simply a unit impulse at the origin. For all other filters, $r_1 = 1/2$ cycle and $r_{i+1} = 2r_i$, $\sigma_1 = 1/2$ and $\sigma_{i+1} = 2\sigma_i$, and $n_1 = 4$ and $n_{i+1} = 2n_i$. Thus we obtain the cascade. The amplitudes are scaled by the magnitudes of the Laplace approximation. Thus,

$$\mathcal{H} = \beta \mathcal{H}_0 + \sum_{i=1}^{N} \beta e^{-|r_i|/2\alpha^2} \mathcal{H}(R_i) \tag{3.27}$$

This filter bank has two constants α and β – which set how much we want high frequency deformations to be used and the step size of the update. Figure 3-2 shows the 1-σ contours of the filters in the filter bank. Note that the original kernel is real and so is our filter bank.

We have not modeled the off-diagonal term $\mathcal{H}^b$ in Equation 3.20. It is also real and exhibits similar power-law behavior, and it can easily be incorporated. $\mathcal{H}^a$ and $\mathcal{H}^c$ sufficiently regularize the problem, however. We thus rewrite Equation 3.20 as

$$\mathcal{Q}^x = -\mathcal{H}\mathcal{F}^x = -\sum_{i=0}^{N} \mathcal{H}_i \mathcal{F}^x \tag{3.28}$$

$$\mathcal{Q}^y = -\mathcal{H}\mathcal{F}^y = -\sum_{i=0}^{N} \mathcal{H}_i \mathcal{F}^y \tag{3.29}$$

Such a reparameterization of viscous alignment has not been proposed before.

3.3.3 Understanding the Filter Bank

Each filter ring $\mathcal{H}_i$ in the bank corresponds to a certain type of deformation. We will discuss some of the filters in greater detail to facilitate understanding of the power of our approach.

$\mathcal{H}_0$: Pure Translation

Let us consider a template X. The target Y is X, but shifted by the vector $\underline{s}_0 = (s_x, s_y)^T$. The ideal deformation field we can find is clearly the constant vector field $\mathbf{q}(\underline{p}) = \underline{s}$. Therefore:

$$Q^x = s_x \quad \leftrightarrow \quad \mathcal{Q}^x = s_x \delta(m, n) \tag{3.30}$$

$$Q^y = s_y \quad \leftrightarrow \quad \mathcal{Q}^y = s_y \delta(m, n) \tag{3.31}$$

where $\delta(m, n)$ is a unit impulse as the origin. In our formulation, $\mathcal{Q}^x = -\mathcal{H}\mathcal{F}^x$ and $\mathcal{Q}^y = -\mathcal{H}\mathcal{F}^y$. To generate the desired deformation field, either $\mathcal{H}$ must be a unit impulse or the forcing terms $\mathcal{F}^x, \mathcal{F}^y$ must be. Even in this constrained case where Y is a translated version of X (and in general), $\mathcal{F}^x$ and $\mathcal{F}^y$ will have high-frequency components *i.e.* is not an impulse at the origin. To generate the preferred deformation field, $\mathcal{H}$ must be the impulse. We have defined $\mathcal{H}_0$ to be precisely that!

Using $\mathcal{H} = -\beta\mathcal{H}_0 = -\beta\delta(m, n)$, we can use the DC Value Theorem [20] to write:

$$\mathbf{q} = -\beta \sum_{\underline{p}} \nabla X(\underline{p}) \left(Y(\underline{p}) - X(\underline{p}) \right) \tag{3.32}$$

$\mathbf{q}$ is indeed a constant vector field. It moves in the direction of increasing the overlap between Y and X. As the template converges to the target, $X = Y$ so $\mathbf{q}$ approaches 0. Here β is *literally* the step size, and we must ensure that β is small enough so that the template does not oscillate around the true convergence point.

$\mathcal{H}_1$: Affine Transformations

Let us consider a target Y that is an affine transformation of the template X. Table 3.1 summarizes the ideal deformation field for each type of affine transformation.

Notice that all the fields are linear combinations of x and y. The Fourier transforms of x and y are complicated to write out explicitly, but they are 0 at the origin and everywhere else except on the m and n axes. On the axes, nearly all their power is located $(m, n) = (\pm 1/2 \text{ cycles}, 0)$ and $(m, n) = (0, \pm 1/2 \text{ cycles})$. That means if we

Affine Type	Y	Ideal Field
Scale	$X(a_1x, a_2y)$	$Q^x = (1-a_1)x$ $Q^y = (1-a_2)y$
Shear	$X(a_1y, a_2x)$	$Q^x = (1-a_1)y$ $Q^y = (1-a_2)x$
Rotate	$X(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$	$Q^x = x(1-\cos\theta) - y\cos\theta$ $Q^y = x\sin\theta + y(1+\sin\theta)$

Table 3.1: Summary of affine transformations of the template X to form the target Y and the corresponding deformation field to recover that transformation. Notice that all the fields are linear combinations of x and y, which means that a filter with the power spectrum characteristics of $(x+y)$ can generate these affine deformations.

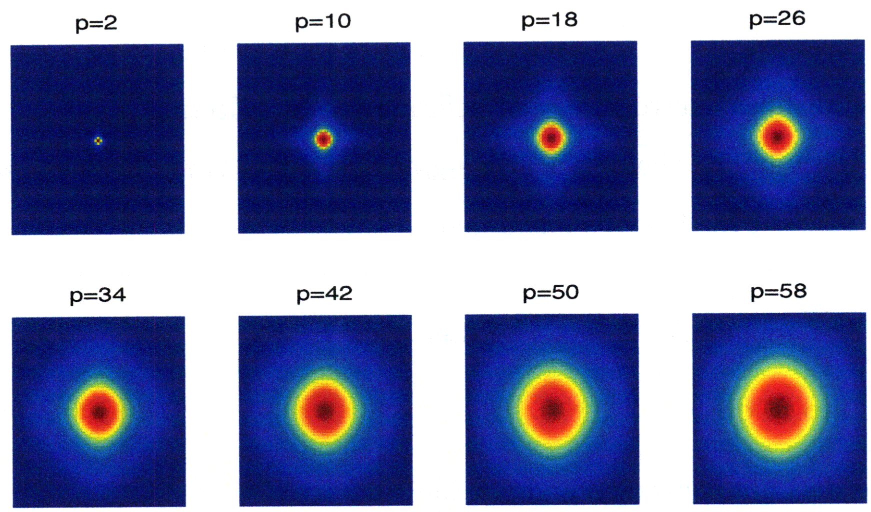

Figure 3-3: Power spectra of Fourier transform of polynomial deformation fields of various powers p (blue is low, red is high). As p increases, the power spectrum broadens. Local, more "complicated" deformation fields occupy higher-frequency bands than global, "simpler" ones.

construct a filter with these characteristics, then it can generate affine deformations. The filter bank $\mathcal{H}_1$ is precisely this filter.

Higher-Order Transformations

Less global (more local) transformations require deformation fields that are higher-order functions of grid position. For the sake of explanation, let the deformation field be a polynomial function of grid position of power p, *i.e.* $Q^x = O(x^a y^b)$ and $Q^y = O(x^a y^b)$. We can plot $||\mathcal{Q}^x||$ and $||\mathcal{Q}^y||$. Figure 3-3 show the power spectra for fields $a = b = p$ for various even values of p. As the exponent increases, the more complex the deformation field and the broader the energy. Of course, we are not constrained only to polynomial deformation fields for solutions, but it illustrates the intuition neatly. Our filter bank cascade on spectral radius captures the idea that the more complicated the deformation, the higher frequency filter band needed to generate it.

3.4 Scale-Cascaded Alignment

Algorithm 1 describes our alignment procedure, obtained as a continuation of the spectral radius. We note that 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 further developed here.

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 $\mathcal{F}_j^x, \mathcal{F}_j^y$ using Y and $X_i(\underline{p} - \mathbf{q_{0:j-1}}(\underline{p}))$
    Solve: $\mathcal{Q}_j^x = -H_i \mathcal{F}_j^x$
    Solve: $\mathcal{Q}_j^y = -H_i \mathcal{F}_j^y$
    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

In Figure 3-5, we see the primary benefit of the cascaded approach. The red error curve is the result of using Equation 3.8 for aligning the images in Figure 1-2. The blue one, with its characteristic drops, depicts the dissipation of error energy from the lower to higher frequency in sequences shown in Figure 3-4. For the rectangle to "flower" 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

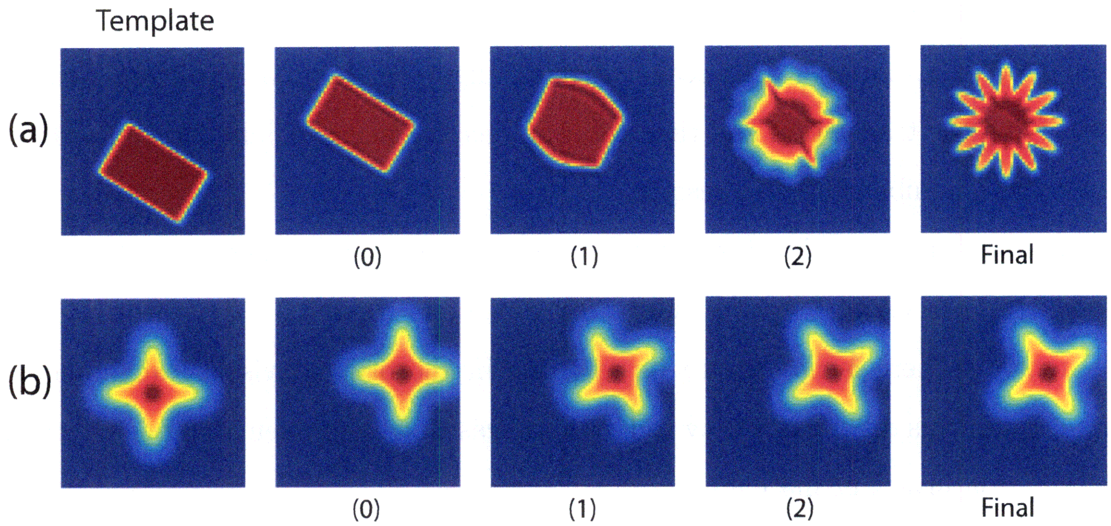

Figure 3-4: Scale-cascaded alignment of template/target pairs shown in Figure 1-2. Each column shows the converged image using the sub-band $\mathcal{H}_i$. Notice how each deformation corresponds well to a perceptual notion of complexity.

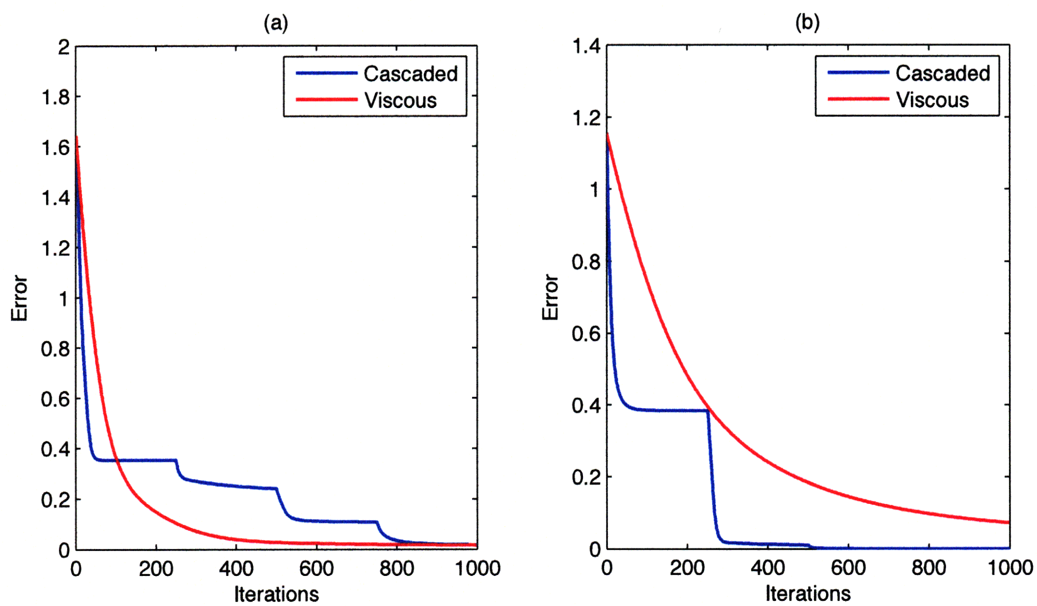

Figure 3-5: Error sequences for rectangle to flower (a) and translating-rotating cross (b). Red curves show the error sequence using viscous alignment [Figure 1-2]; blue curves show the error sequence using the scale-cascaded approach [Figure 3-4]. Convergence rates of viscous alignment are not correlated to perceptual similarity.

each radius[3] are. 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 we can assign the dissipation of error energy to spectral radii, 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] These filters were run to their iteration limit.

Chapter 4

Deformation Invariant Matching

With scale-cascaded alignment, we are ready to build the deformation invariant matcher. We introduce *scoring curves*, which selectively weigh the error curve generated by the SCA, and discuss their design. Then we apply our deformation invariant matcher to a problem in conservation biology. We review the setup of the problem, existing results, and finally present the results of our technique.

4.1 Deformation Invariant Matching

Scale-cascaded alignment results in an error sequence $\Delta(X, Y) = [\Delta_0 \Delta_1 \ldots \Delta_N]$ corresponding precisely to the dissipation of the optimization objective energy by each filter $\mathcal{H}_0 \ldots \mathcal{H}_N$. The entire sequence can be thought of as a vector, as shown in the plot in Figure 3-5(d).

Deformations are scored as a weighted sum with a scoring curve, that is:

$$e(X, Y) = \Delta(X, Y)W \tag{4.1}$$

There are several ways to choose W. If we suppose that lower frequency deformations are perceptually irrelevant, we can weigh them lower. If the image is prone to high-frequency noise or localized specularities, we may wish to discount higher frequency deformations. Although estimating W is best viewed as a learning problem, here we

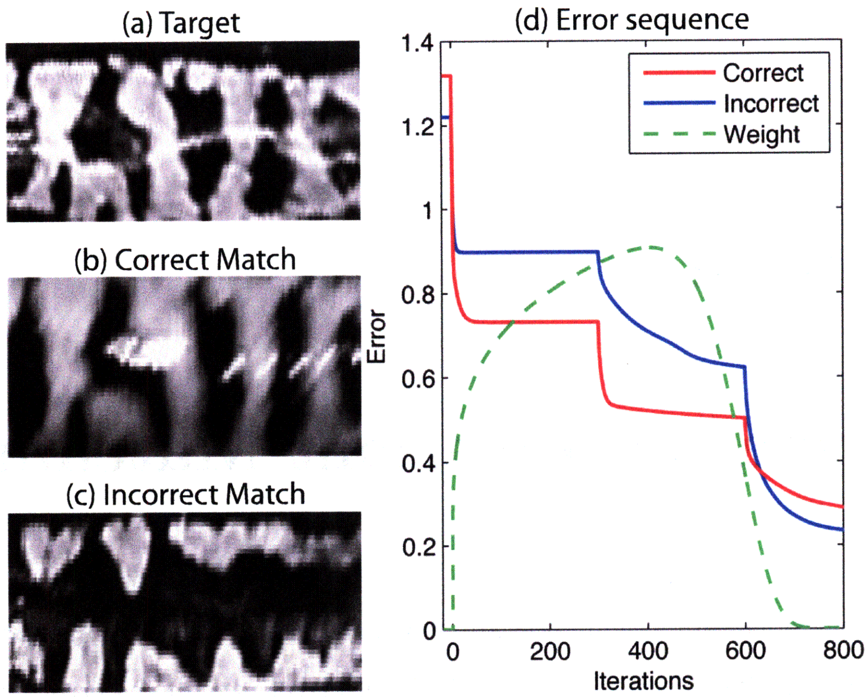

Figure 4-1: Error sequences generated by aligning (b) to (a) and (c) to (a) are shown in (d) in red and blue respectively. The scoring curve is shown in dashed green. Although (b) is the true match, it has higher initial and terminal error than (c).

design scoring curves manually.

As a primer, we present an example of three image patches in Figure 4-1. Though it is obvious to the reader that template (b) is the true match to the target (a), multiscale PCA (with sample statistics computed over 6000 images) identified the template (c) as a better match. Note that PCA is a special case of the objective (Equation 3.2) when the displacement and associated regularization constraint are dropped. Thus, as shown in Figure 4-1(d), the starting point of the error sequence also 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. Thus, there is a band within the alignment sequence that accurately depicts the relative closeness of the two targets to the template. Its active region corresponds roughly to $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$.

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

$$W(j; \sigma_c, O) = (\frac{j}{\sigma_c})^{(\frac{2}{O})} e^{-(\frac{j}{\sigma_c})^O} \tag{4.2}$$

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 4-1, 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 affine transformations and turbulent motion (*i.e.* noise) but be sensitive to all other deformations. This is difficult to model in other approaches, such as Ling and Jacobs [21], but presents no difficulty in the scale-cascaded approach.

4.2 Experimental Application

We apply our method to capture-mark-recapture study of marbled salamanders [9]. A database of 6021 images was collected over six years in the field by trapping individual salamanders and photographing them (see Figure 4-2). The scientific objective is to track each salamander's movement individually 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.

4.2.1 Baseline Work

As shown in Figure 4-2, marbled salamanders have extremely deformable bodies. The images themselves are blurry, noisy, and specular (see Figure 4-1). To solve the animal pose problem, Gamble *et al.* [9] artificially straighten the salamanders by detecting their medial axis manually and warping the image so that the medial axis becomes straight in the rectified image, removing large lower-order deformations. Then the authors extract patches between key features (the feet) and match them

(a)

(b)

(c)

Figure 4-2: Salamanders show extreme variability in pose, background, and specularity.

using multiscale PCA [9]. Illumination was compensated by contrast-normalization, and specularity was not explicitly dealt with. The ROC curve of this method is shown in blue in Figure 4-3, determined using relevance judgements provided by users who examined the top $N = 20$ retrievals over 150 queries.

4.2.2 Challenges

The primary difficulties in improving the performance of our system are still the existence of specularity and local deformations (see Figure 4-1 and Figure 4-4). 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 4-2). As pointed out in Chapter 4, specularity causes large variance in the patch covariance, which may partially explain MS-PCA's relative success. Because specularity is also localized, very high frequency deformations are necessary to "paint" it out. We mitigate this by down-weighting the contributions of large wavenumber deformations in W. (The population covariance technique described Chapter **??** is promising but was not implemented.)

The nonlinearity in pose remains as a difficulty, but it does not occupy the high frequencies required to marginalize noise and specularity. Thus, a mid frequency portion of the deformation spectrum is worth exploring. Therefore, we apply the

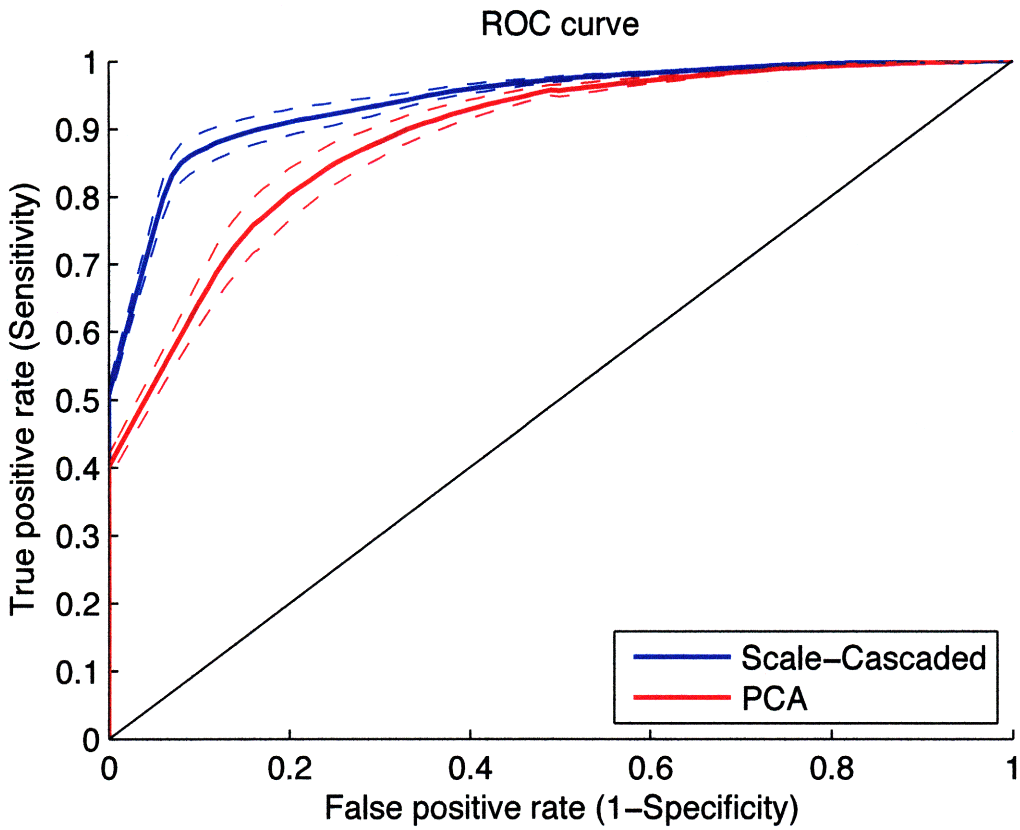

Figure 4-3: ROC curve for deformation invariant matcher (blue) and MS-PCA (red). The dashed lines show the variance for each method over 150 queries.

scale-cascaded viscous alignment with the scoring curve shown in Figure 4-1 (dotted green curve).

4.2.3 Results

We re-ranked the top 20 ranks for which relevance judgements were available. We use the same 150 queries as in [9]. To set the filter bank parameters, we manually ran alignments on sample template/target pairs for various settings. The settings $\alpha = 1/8$ and $\beta = 1$ were numerically stable and converged reasonably well. SCA was run with a 300 iteration maximum per filter $\mathcal{H}_i$. Patches were resized to be 64×64 pixels. A small Gaussian blur (width of 3 pixels) was applied to patches before alignment. With these settings, each alignment took approximately 5 minutes.

The ROC shows marked improvement (Figure 4-3). Figure 4-4 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.

MS-PCA matches well (as can be seen in Figure 4-3). Even if the MS-PCA mismatches look plausible, our scale-cascaded approach ranks the retrievals better. Our method succeeds because it can relate image alignment into deformation classes that are relevant for perceptual image similarity.

Figure 4-4: 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. Mismatches are highlighted in red and their labels are crossed out.

Chapter 5

Conclusion

We demonstrate a new approach to deformation invariant image matching for recognition. Our approach retains the broad range of deformations that viscous alignment methods can model, which is possible because the template is never linearized (in contrast to optic flow) and nonlinear deformations can be iteratively produced. However, our approach remodels the viscous constraint with a filter bank that 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. It is this utility that allows us to mitigate the effects of nonlinear deformations in a recognition task, where the results show a clear and substantial improvement. Our choice of parameterization has biological motivation [12], and more concretely relates representation of texture and deformation via a common bank of filters [19].

There are several areas that we wish to explore. First, we wish to explicitly incorporate the change in amplitude uncertainty C (Equation 3.2) as alignment progresses. Although we assume C was constant with respect to the deformation for our experiments, it is not because the original covariance contains position errors in addition to amplitude errors that must be factored out [3, 25]. Second, we want to examine ways to learn W, and learn sparse-prior models for learning activations of $\mathcal{H}$ or the filter bank itself [26]. Third, we want to examine if this approach can calculate deformation statistics better [3]. Fourth, although correspondence was not sought, we can extend PCA-SIFT [18] with this approach easily. Finally, it may be useful to reformulate our approach with mutual information measures [7, 31].

Appendix A

Deriving Optical Flow from Viscous Alignment

We can easily show that optical flow [15] is a special linearized case of the generalized viscous alignment described in Chapter 3.

As before, X and Y are the template and target images respectively and $\mathbf{q}$ is the deformation field. Let each discrete grid point $\underline{p} \doteq (x, y)^T$. We wish to solve the nonlinear objective function from Chapter 3.1 (Equation 3.2):

$$J(\mathbf{q}) = \frac{1}{2} \sum_{\underline{r}} \sum_{\underline{s}} \left[Y(\underline{r}) - X(\underline{r} - \mathbf{q}(\underline{r}))\right] C(\underline{r}, \underline{s}) \left[Y(\underline{s}) - X(\underline{s} - \mathbf{q}(\underline{s}))\right] - \ln P(X) + L(\mathbf{q}) \tag{A.1}$$

Optic flow assumes no observation noise, which means that $\underline{\underline{C}}$ is the identity so that $C(\underline{r}, \underline{s})$ is 1 when $\underline{r} = \underline{s}$ and 0 otherwise. Therefore we can rewrite Equation A.1 as:

$$J(\mathbf{q}) = \frac{1}{2} \sum_{\underline{r}} \left[Y(\underline{r}) - X(\underline{r} - \mathbf{q}(\underline{r}))\right]^2 - \ln P(X) + L(\mathbf{q}) \tag{A.2}$$

In optic flow, we linearize the image difference. In our notation, we make the following approximation:

$$X(\underline{p} - \mathbf{q}(\underline{p})) \approx X(\underline{p}) - \frac{\partial X}{\partial \underline{p}}^T \mathbf{q}(\underline{p}) \tag{A.3}$$

where $\frac{\partial X}{\partial \underline{p}}^T = (\frac{\partial X}{\partial x} \; \frac{\partial X}{\partial y})$. We define the quantity $\Delta X(\underline{p})$ as:

$$\Delta X(\underline{p}) \doteq Y(\underline{p}) - X(\underline{p}) \tag{A.4}$$

We can thus rewrite the objective given in Equation A.2 as:

$$J(\mathbf{q}) = \frac{1}{2} \sum_r \left[\Delta X(\underline{r}) + \frac{\partial X}{\partial \underline{r}}^T \mathbf{q}(\underline{r}) \right]^2 - \ln P(X) + L(\mathbf{q}) \tag{A.5}$$

To minimize this function, we differentiate with respect to $\mathbf{q}$ at every grid position $\underline{r}$:

$$\frac{\partial J(\mathbf{q}(\underline{r}))}{\partial \mathbf{q}(\underline{r})} = \frac{\partial X}{\partial \underline{r}} \left[\Delta X(\underline{r}) + \frac{\partial X}{\partial \underline{r}}^T \mathbf{q}(\underline{r}) \right] + \frac{\partial L(\mathbf{q}(\underline{r}))}{\partial \mathbf{q}(\underline{r})} = 0 \tag{A.6}$$

In an optic flow formulation, we assume that an image undergoes constant motion, described by the velocity field $\mathbf{q}$, over some small time step δt [15]. We can think of our template and target images X and Y as the start and end images resulting from that motion. The velocity field $\mathbf{q}$ is constant over time, and is regularized using smoothness *i.e.* $\frac{\partial L}{\partial \mathbf{q}} = \nabla^2 \mathbf{q}$ [15]. Let $E(x, y, t)$ be the irradiance function [15] that describes the apparent motion; we can then rewrite our variables in the optic flow notation given in [15]:

$$E(x, y, t) = X(\underline{p}) \tag{A.7}$$

$$E(x, y, t + \delta t) = Y(\underline{p}) \tag{A.8}$$

$$E_t(x, y) = \Delta X(\underline{p}) \tag{A.9}$$

$$(E_x, E_y)^T = \frac{\partial X}{\partial \underline{p}} \tag{A.10}$$

$$(u, v)^T = \mathbf{q_i}(\underline{p}) \tag{A.11}$$

We can thus write:

$$\begin{bmatrix} \nabla^2 u \\ \nabla^2 v \end{bmatrix} = - \begin{bmatrix} E_x \\ E_y \end{bmatrix} (E_t + E_x u + E_y v) \tag{A.12}$$

This is precisely the optic flow formulation as given by [15]. Optic flow is thus a special case of viscous alignment where (1) we use only smoothness to regularize the velocity field and (2) the image body force is linearized.

Bibliography

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

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, Apr 2002.

[3] G. Charpiat, O. Faugeras, and R. Keriven. Image statistics based on diffeomorphic matching. In *ICCV*, volume 1, pages 852–857, Washington, DC, USA, 2005. IEEE Computer Society.

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

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995.

[6] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 484–498. Springer, 1998.

[7] 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 '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, pages 541–548, London, UK, 2002. Springer-Verlag.

[8] 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.

[9] 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.

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

[11] 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 (SCIA'97)*, pages 785–802, Lappeenranta, Finland, 1997.

[12] D. J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1(4):279–302, 1988.

[13] A. Hill, C. J. Taylor, and A. D. Brett. A framework for automatic landmark identification using a new method of nonrigid correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):241–251, Mar. 2000.

[14] J. Holmberg, B. Norman, and Z. Arzoumanian. Estimating population size, structure, and residency time for whale sharks *Rhincodon typus* through collaborative photo-identification. *Endangered Species Research*, 7:39–53, May 2009.

[15] B. K. P. Horn. *Robot Vision.* MIT Press, 1986.

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

[17] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

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

[19] P. Kruizinga, N. Petkov, and S. E. Grigorescu. Comparison of texture features based on gabor filters. In *ICIAP '99: Proceedings of the 10th International Conference on Image Analysis and Processing*, pages 142–147, 1999.

[20] J. S. Lim. *Two-Dimensional Signal and Image Processing.* Prentice Hall, 1990.

[21] H. Ling and D. W. Jacobs. Deformation invariant image matching. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, volume 2, pages 1466–1473, Washington, DC, USA, 2005. IEEE Computer Society.

[22] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(6):580–591, 1993.

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

[24] M. S. Nixon and A. S. Aguado. *Feature extraction and image processing.* Academic Press, 2008.

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

[26] S. Roth and M. Black. Field of experts: A framework for learning image priors. In *Proc. CVPR*, pages 860–867, 2005.

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

[28] J. P. Thirion. Fast non-rigid matching of 3D medical images. In *Proceedings of the Conference on Medical Robotics and Computer Assisted Surgery (MRCAS'95)*, Baltimore, November 1995.

[29] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

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

[31] P. A. Viola and W. M. Wells III. Alignment by maximization of mutual information. In *International Journal of Computer Vision*, pages 16–23, 1995.

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