Chapter 9
Deformations and Lighting

David Jacobs, Anne Jorstad, and Alain Trouvé

9.1 Introduction

One of the most basic problems in vision is to formulate an effective distance between two images. This should account for two effects. First, pixels can change their position as viewpoint or the shape of objects change. Second, pixels may change their intensity as lighting changes. A good distance should capture both of these transformations. Further, it should allow us to unwind them, so that we can determine correspondences between the two images that we compare.

In many vision problems, intensity changes are primarily due to lighting variation. In this chapter we sketch a general approach to this problem, and then describe some initial results that implement components of this approach. Our work rests on a new approach to modeling lighting effects, which yields a new image distance, with many appealing properties. We will also focus on combining this with both existing and new models of image deformation.

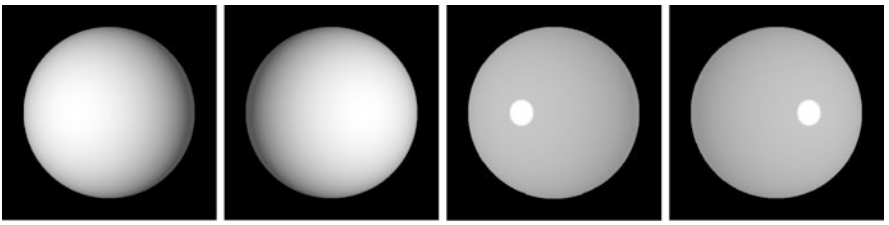
Our image distance aims to capture three ubiquitous properties of real world images. *Geometric* variations have the effect of *deforming* the 2D appearance of an object. For example, a 3D object might deform or have parts that articulate, causing its image to deform. Or, if we view even a rigid 3D object from a new viewpoint, this can have the effect of deforming the resulting image. Also, different instances of objects within the same class can vary considerably in shape; we can treat some of the effects of this shape variation as a deformation. Therefore, we can model a significant number of geometric variations as image deformations.

D. Jacobs (✉) · A. Jorstad
Department of Computer Science, University of Maryland, College Park, MD, USA
e-mail: djacobs@cs.umd.edu

A. Jorstad
e-mail: jorstad@math.umd.edu

A. Trouvé
CMLA, École normale supérieure de Cachan, Cachan, France
e-mail: trouve@cmla.ens-cachan.fr

S.J. Dickinson, Z. Pizlo (eds.), *Shape Perception in Human and Computer Vision*,
Advances in Computer Vision and Pattern Recognition,
DOI 10.1007/978-1-4471-5195-1_9,



(A) (B) (C) (D)

Fig. 9.1 A simple illustration of some problems that arise in image matching. (**A**) and (**B**) are images of a sphere illuminated from different directions. (**C**) is identical to (**A**), with a change in its histogram to create higher contrast. (**D**) is similarly derived from (**B**). We would like to judge the similarity of each image pair, and the extent to which they should be matched by deforming the image or by positing a lighting change. Our proposed algorithm has the following properties: when intensities with smooth shading shift position, as between (**A**) and (**B**), the metric tends to view this as a lighting change, whereas when high contrast intensity patterns shift position, as between (**C**) and (**D**), the metric tends to interpret this as an image deformation. Also, changes in contrast that do not affect image gradient directions but that do introduce high contrast edges, as between (**A**) and (**C**), can be costly

Usually, the effects of image deformations are complicated by simultaneous variations due to changes in *lighting*. As an object articulates, deforms or moves relative to a camera, it also moves relative to the light. When we view two similar objects, we rarely view them under identical lighting. These changes in lighting can have a dramatic effect on the appearance of objects. Moreover, deformations and lighting changes will occur in the presence of *noise*, *occlusions* and other unmodeled image variations that require *robust* matching. Therefore, *our primary goal is to develop robust image matching methods that can simultaneously handle changes in lighting and geometry*.

We model deformations as continuous one-to-one transformations of the image that affect the position of pixels but not their intensity. Lighting, on the other hand, affects the intensity of pixels, but not their position. These effects can be difficult to separate because of the ambiguity that occurs in finding a correspondence between two images. Figure 9.1 illustrates this problem. We can match each pair of images using a pure deformation that warps the bright pixels in one image to the position they occupy in a second image. Or we can explain the image change with a change in lighting position that alters the intensity of all pixels (we show in [5] that any two images can be explained with a single, Lambertian scene and a lighting change). In general, any possible deformation can be combined with alteration in intensities to match two images; our problem is to select the combination of deformation and intensity change that provides the best explanation. For object recognition, it is also important that we assign a cost to this interpretation, so we can judge its validity.

We will approach this problem by first developing a Riemannian metric for images that captures deformations and lighting change. This kind of metric is appropriate for image transformations that can be modeled as the result of a continuous sequence of small transformations. That is, it is appropriate when it makes sense to continuously morph from one shape to another, or one lighting condition to another.

Therefore, we can build a distance between images by constructing a metric on any infinitesimal change in the image, and then stitching these together to provide a geodesic distance between any two images.

There is much work already on constructing Riemannian metrics that capture image deformations. Though we have some proposed enhancements to these, our main focus is on developing comparable metrics for intensity variation that capture lighting effects, so that these may be effectively combined. We first propose such a metric, and show that it has many of the properties of existing, successful approaches to lighting change, and also has some desirable properties not possessed by these representations. As an additional benefit, this metric also robustly captures variations in image intensity due to occlusion and clutter.

This metric will provide a theory of computation for deformation and lighting, that encodes our notion of image similarity. However, it is still a considerable challenge to find ways to effectively compute with such an image metric. To address this, we first show that our local metric can be incorporated into an optical flow framework to produce an image distance that performs face recognition effectively. Then, we show that for the intensity component of our metric alone, geodesic distances can be computed extremely efficiently in the wavelet domain. Most of the results in this chapter and further developments are described in greater detail in [13].

9.2 Background

There is a vast amount of work on image matching. First, we note that many approaches have been suggested for comparing images in ways that are insensitive to illumination, typically through a process of locally normalizing the image intensities. Osadchy et al. [19] reviews these methods and proves that several are essentially equivalent. Gopalan and Jacobs [9] performs an experimental comparison of many of these approaches and finds that comparison using simple gradient directions works best. Representing an image in terms of the gradient direction is equivalent to a local normalization that is invariant to additive or multiplicative changes in intensity. We conclude that a good lighting metric can be expected to bear some resemblance to image comparison using gradient direction, as a prototypical example of local normalization.

There is also a great deal of prior work on non-rigid image matching. Much of this has been done in the context of tracking, in which image changes are assumed to be small ([1, 2] provide some entry points to this vast literature). Much work that handles larger deformations has been done, especially in the domain of medical imaging, in which body structures from different individuals are non-rigidly aligned (e.g., [6, 8]), although often these ideas are applied to other domains, such as faces. Riemannian metrics for deformable matching have been developed, (e.g., [3, 17, 22]), as well as novel image descriptors that are invariant to deformations [14]. In general, all these approaches typically find a deformation by combining some penalty on the deformation, to encourage smoothness, with some image term that

encourages the aligned images to have similar intensities. Our proposal fits into this framework; however we propose that the image term should be designed to robustly compare intensities in a way that captures the effects of lighting variation.

Finally, there has been much work on image matching when there is both geometrical and lighting variation, primarily in the context of motion tracking. These approaches attempt to deal with the fact that a moving object deforms in the image and moves relative to the lighting (e.g., [2, 10]). There is also a huge literature on optical flow, some of which explicitly focuses on matching with lighting variation [4, 15, 18].

9.3 Robust Image Metrics for Lighting and Deformation

Our approach will make use of the theory of image metamorphosis as a framework that allows us to embed the notion of simultaneous deformation and lighting change into a proper Riemannian formulation. Prior developments of this framework are summarized in [23]. Intuitively, in this approach we proceed by first defining a cost on infinitesimal image changes. This defines a tangent space for each image. This tangent space spans the image space; that is, any small change to the image is possible. But the cost of these local changes may be quite different from Euclidean distances in the image space. Together, these tangent spaces define a manifold that fills the space of all images. The distance between any two images, then, is a *geodesic* path on this manifold. We can think of this geodesic as the lowest cost metamorphosis from the first image to the second, in which the image is simultaneously deforming and changing intensity.

More precisely, we can express a small change, δI, to an image I, through a deformation, as:

$$\delta I(x) = \partial_t I\big(x - tv(x)\big)\big|_{t=0} = -\big\langle \nabla I(x), v(x)\big\rangle$$

Here t represents time, $\langle , \rangle$ denotes the inner product, and v is a vector field that encodes the instantaneous deformation. This equation is essentially the equation of optical flow. Alternately, we can express the image change as a change in the intensity of pixels, with no deformations, as: $\delta I = h$, where $h(x)$ represents the change in intensity at x. More realistically, we should describe the image change as the sum of these two processes, $\delta I(x) = -\langle \nabla I(x), v(x)\rangle + h(x)$, each of which might explain image changes. Given a metric $|v|_V$ on the deformation, and a metric $|h|_H$ on the intensity change, one can define a combined metric:

$$\|\delta I\|_I^2 = \inf_{v \in V} \big(|v|_V^2 + \lambda \big|\delta I + \langle \nabla I, v\rangle\big|_H^2\big) \tag{9.1}$$

This gives the cost of an image change as the minimum possible cost of a deformation and intensity change, where λ weights the costs. Note that any specific deformation implies a specific intensity change. Most of the work on metamorphosis of images concerns the simplest case in which the metric on intensity change, $|h|_H$, is just the L^2 metric.

9.3.1 Intensity Variation

We now present two new, related metrics for intensity variation, and describe some of their mathematical properties. The analysis in this section concerns intensity variation only, and does not consider deformations. These metrics are closely related to image comparison based on the direction of the image gradient, which has proven to be very robust to lighting changes. At the same time, these metrics also measure contrast variations in a way that fits some intuitive notions about intensity change. Finally, these metrics also produce a robust image comparison method for occlusions, related to L^1 norms.

Intuitively, we want to develop an intensity metric that is related to the likelihood that two images come from the same 3D scene, and that the difference between the two images is due to a lighting variation. Because we are constructing a general metric for images, this distance cannot be based on specific knowledge of the 3D structure of the current scene, but rather should depend on general considerations about the effect of lighting variation on image intensities.

We begin with a simplified, intuitive description of some key issues posed by lighting. The intensity of a pixel mainly depends on the lighting conditions at the corresponding scene point, on the direction of the surface normal at that point, and on the albedo, that is, the fraction of light reflected at that point. When there is a small change in lighting, an image, I, will change by some amount δI. First, we note that as lighting changes, the intensity at a single pixel can change in almost any way. That is, $\delta I(x)$ is not very constrained for any point x. However, δI is often correlated at nearby points, so we will consider the properties of δI as it varies spatially, that is the properties of $\nabla \delta I$. Usually, lighting conditions vary slowly within a scene, so $\nabla \delta I$ is primarily the result of the differential effect that lighting variation has as surface normals and albedos vary throughout a scene. We consider three situations.

First, consider the case of regions of smooth or planar surfaces with nearly uniform albedo. In this case, $\nabla \delta I(x)$ tends to be small, and also the initial image gradient, $\nabla I(x)$ tends to be small. Second, suppose the albedo changes in this region, but the surface normal is constant or changes slowly. In this case, lighting tends to uniformly scale the intensities of pixels in a region, so $\nabla \delta I(x)$ tends to be proportional to $\nabla I(x)$, and $\nabla \delta I(x)$ is large only when $\nabla I(x)$ is large. Finally, consider surfaces with high curvature or curvature discontinuities, in which nearby points have very different surface normals. In this case, nearby points are exposed to different lighting, and affected quite differently even by the same lights. So changing lighting can affect nearby pixels in a way that is almost uncorrelated. Therefore, $\nabla \delta I(x)$ tends to be quite unpredictable, it might be low or high. However, usually (but not always) in this situation $\nabla I(x)$ tends to be high.

As an example to illustrate these situations, consider a v-shaped roof with two sides, facing in two different directions. If a region is on one side of the roof and has uniform albedo, it will tend to have uniform intensity. As the sun moves, the whole region may get brighter or darker, but it will continue to be uniform. So $\nabla I(x)$ and $\nabla \delta I(x)$ will both be low. Suppose, now, the roof is striped, and consider

a region that is still on one side of the roof, but that crosses this stripe. The gradient across this stripe will get stronger when the sun faces the roof more directly, and the whole side of the roof gets brighter. Therefore, the image gradient produced by the stripe changes in proportion to the overall intensity, i.e., $\nabla \delta I(x)$ is proportional to $\nabla I(x)$. Finally, consider a region of the images that crosses the two sides of the roof. As the sun moves, or the light otherwise changes, the intensity on the two sides of the roof also changes. The relationship between these two intensities cannot be predicted without specific knowledge of the geometry and lighting. So, $\nabla \delta I(x)$ is unpredictable across the edge separating the two sides of the roof. However, most of the time this edge will have a strong gradient itself. Therefore, we see that when $\nabla I(x)$ is large, this may signal that $\nabla \delta I(x)$ can also be large, or at least its value is hard to predict from the image alone. The following metric on intensity changes captures these properties:

$$C(\delta I) = \left(\int_{\Omega} \left(\frac{|\nabla \delta I(x)|}{|\nabla I(x)| + \varepsilon} \right)^2 dx \right)^{1/2} \tag{9.2}$$

Here $C(\delta I)$ denotes the cost of a small change to the intensities, and the integral is over the image. This cost has the desired properties: (1) When $\nabla I(x)$ is low, large values of $\nabla \delta I(x)$ are very expensive; (2) When $\nabla \delta I(x)$ is proportional to $\nabla I(x)$ the cost is fairly constant; (3) When $\nabla I(x)$ is high, the value of $\nabla \delta I(x)$ is not important in determining the overall cost. Note that ε is a small constant. This prevents division by zero, and adds other useful properties, as we will see later.

We can also consider a somewhat more complex metric that bases the cost of a change not only on the magnitude of its gradient but also on its direction, with a lower cost for changes in the direction of the image gradient. This cost has interesting theoretical properties, though we have not yet evaluated it experimentally.

$$C(\delta I) = \left(\int_{\Omega} \frac{\left((1 + \rho(|\nabla I|)) \left\langle \frac{\nabla \delta I}{|\nabla I| + \varepsilon}, \frac{\nabla I^{\perp}}{|\nabla I^{\perp}|} \right\rangle^2 + \left\langle \frac{\nabla \delta I}{|\nabla I| + \varepsilon}, \frac{\nabla I}{|\nabla I|} \right\rangle^2 \right)}{(2 + \rho(|\nabla I|))} \right)^{1/2} \tag{9.3}$$

ρ is a function that serves as a weight between the two terms. $\rho(|\nabla I|)$ goes to 0 as $|\nabla I|$ goes to 0, and becomes large as $|\nabla I|$ goes to infinity. One natural choice for ρ would be $\rho(|\nabla I|) = \log(1 + |\nabla I|)$. $\nabla I^{\perp}/|\nabla I^{\perp}|$ is a unit vector orthogonal to the image gradient, while $\nabla I/|\nabla I|$ is a unit vector in the direction of the gradient. Equation (9.3) becomes identical to (9.2) as $|\nabla I|$ goes to 0.

Intuitively, this metric divides the change in the image into two parts, one orthogonal to the image gradient, and one in the direction of the image gradient. The first type of change is more costly than the second. This allows us to also express the fact that typically the direction of $\nabla \delta I(x)$ is the same as the direction of $\nabla I(x)$, because the direction of $\nabla I(x)$ is more likely to be the direction in which scene properties are changing rapidly. Note that this is true in our roof example.

We have been able to analytically determine the distance between pairs of images for some simple cases that show that our metric has a number of favorable properties. We will summarize these properties here, while omitting derivations:

1. *Relation to direction of gradient comparisons:* Let us consider the above metrics for the case in which $\nabla \delta I(x)$ is always in the direction perpendicular to $\nabla I(x)$. In this case, we can show that the cost of a local change in image intensities is approximately proportional to the change in the direction of the image gradient. That is, if we denote the change in the direction of the image gradient by: $\delta\theta(x)$, the cost in (9.2) reduces to:

$$C(\delta I) = \left(\int_{\Omega} \left(\frac{|\nabla I(x)|}{|\nabla I(x)| + \varepsilon} \right)^2 \left(\delta\theta(x)\right)^2 dx \right)^{1/2}$$

   For (9.3), we can select a parameter ρ so that $\rho(|\nabla I|)$ is always high. This assigns minimal cost to all intensity changes in which $\nabla \delta I(x)$ is parallel to $\nabla I(x)$. So, by proper parameter selection we can make our metric reduce to one that measures the change in the direction of the image gradients. As noted above, many current approaches to lighting insensitive image comparison use representations like this, so this assures us that we can achieve good performance with this metric.
2. *Contrast change:* When $\nabla \delta I(x)$ is in the same direction as $\nabla I(x)$, this alters the contrast in an image, without changing the direction of the image gradient. For this case, we can derive a closed form solution for the total cost of the geodesic path between two images, which we denote $d(I_0, I_1)$. We can show that when the magnitude of the gradient in an image changes from $|\nabla I_0|$ to $|\nabla I_1|$ then our metric assigns this a cost proportional to:

$$d(I_0, I_1) = \left(\int_{\Omega} \left(\log \frac{|\nabla I_1(x)| + \varepsilon}{|\nabla I_0(x)| + \varepsilon} \right)^2 dx \right)^{1/2}$$

   That is, the cost at each point is generally proportional to the log of the ratio of the change, though it becomes linear when the gradient magnitude is near zero. This means that when there is a strong image gradient, such as an edge, changing the contrast has little cost. However, changing a low contrast region of the image into a strong edge can have a high cost, even if the direction of the image gradient does not change. This cost is not captured by most current lighting insensitive representations, and seems useful.
3. *Robustness:* Our proposed metrics accounts for occlusions robustly. We can analytically determine the cost of transforming a region of uniform intensity into an arbitrary image pattern caused by occlusion. This cost is similar to, and bounded by, a constant times the bounded variation norm (BV-norm) of the image gradient of the new image pattern. That is, when I_0 is a constant image, and I_1 is not, for the geodesic distance we have:

$$d(I_0, I_1) \leq \frac{1}{\varepsilon} \int_{\Omega} \left| \nabla I_1(x) \right| dx$$

This implies that the cost of transforming one image pattern into a totally different pattern is bounded by the sum of the BV-norms of the two patterns. This is a type of L^1-norm on the image gradients, and is known to be much more robust than L^2-norms.

9.3.2 Interaction Between Deformations and Intensity Change

We can combine our metric for intensity variation with standard metrics for deformation using (9.1). With this formulation, deformations and intensity changes compete to explain image changes. We have analyzed some simple cases of this, in order to gain intuitions. We will consider here the case of an image pattern undergoing a small translation. The cost of explaining such a change through a deformation cost is always low, and is independent of the image content. However, this change can also be explained through intensity variation. We have shown that for a smoothed edge, this cost will vary linearly with the sharpness of the edge. So a small translation of a sharp, high contrast edge is very expensive to explain with an intensity variation. Translation of a smooth, gradual edge is more easily modeled as an intensity variation with our metric (Fig. 9.1).

When there is a more complicated deformation of image intensities, the cost of explaining this with deformations rises, while the cost of explaining this through intensity variations will be the aggregate of a collection of local translations, and depend on the number and sharpness of edges that are shifting.

Consequently, explaining the motion of a number of sharp edges in an image as an intensity variation is very expensive. If a set of sharp edges are deformed, even quite a lot, we will tend to interpret this as a deformation. However, when there are fewer, more gradual edges in the image, we will be more inclined to interpret a complex deformation of these patterns as an intensity variation.

9.4 Using These Metrics for Image Comparison

The metrics we have described form the basis for several algorithms that we have created. First, we have used the local image metric described in Eqs. (9.1) and (9.2) to build an optical flow algorithm. Our goal is not motion understanding, the traditional focus of optical flow research, but to apply this algorithm to compare images with changes in lighting and shape. Second, we show that if we only consider the intensity component of our local metric, given in Eq. (9.2), we can compute an approximation to geodesic distances on a Riemannian manifold constructed with this local metric using a very efficient algorithm that works in the wavelet domain. Finally, an initial foray into the computation of geodesic costs that take account of intensity changes and deformations can be found in [13].

9.4.1 A Deformation and Lighting Insensitive Metric for Face Recognition

Our primary goal is to create a Riemannian manifold of images, in which geodesic distances represent image similarity. However, our local image metric can be more

Table 9.1 Face recognition experiments, comparing our new methods with top previous methods. Run time is given on common hardware, when available. Results shown on recognition with changes in expression, lighting, or both

Method	Time (sec)	Expression	Lighting	Overall
Image Differencing	3.1×10^{-5}	83.0 %	9.0 %	46.0 %
Normalized Cross-Correlation	7.2×10^{-3}	84.0 %	59.3 %	71.7 %
Significant Jet Point [24]	–	80.8 %	91.7 %	86.3 %
Binary Edge and MI [21]	–	78.5 %	97.0 %	87.8 %
Gradient Direction	3.8×10^{-4}	85.0 %	95.3 %	90.2 %
Our Optical Flow Approach	1.0	89.6 %	98.9 %	94.3 %
Our Wavelet Approach	1.3×10^{-3}	93.7 %	96.7 %	95.2 %
Pixel Level Decisions [11]	5.6×10^{-4}	98.0 %	94.0 %	96.0 %
Our Wavelets Thresholded	1.3×10^{-3}	97.3 %	97.0 %	97.2 %

easily evaluated by using it as the cost function in an optical flow algorithm. We then use the correspondences from optical flow to judge similarity, evaluating our results quantitatively on a Face Recognition task. For reasons of space, we only briefly summarize our algorithm here; a full description may be found in [12].

- The intensity component of our local metric is given in Eq. (9.2). To measure deformations, we introduce a new regularization term: $E_r(\alpha) = \frac{1}{2}\langle \alpha, k * \alpha \rangle_G$. Here G denotes a generalized inner product on vector fields, and the flow field is $k * \alpha$, where k is a smooth kernel, such as a Gaussian, and $*$ denotes convolution. α can be considered to represent the dual elements of the flow field. This results in smoother gradients and superior rates of convergence.
- The kernel, k, is a combination of fine and coarse scale Gaussians. This allows the method to converge accurately to a solution that is effective at a fine scale, while avoiding convergence to local minima.
- Optimization over correspondence fields is performed using a modified gradient descent. Convolutions are calculated using the fast Fourier Transform, for efficiency.

We experiment by applying this algorithm to face recognition. To compare two face images, we compute the optical flow between them. At each pixel, this provides us with four values: a flow vector and an intensity gradient change. We use a naïve Bayes classifier to determine whether the resulting values are more consistent with two images of the same person, or of two different people. Results of experiments with the Martinez data set [16] are shown in Table 9.1. This data set contains images of 100 individuals taken with variations in lighting and facial expression. Recognition based on our local metric performs competitively.

9.4.2 Geodesics for Image Comparison with a Lighting-Insensitive Metric

Using Eq. (9.2) as a local metric, we can define an image manifold in which distances will tend to be low when intensity changes are due to lighting variation. It turns out that we can compute approximate geodesic distances on this manifold very efficiently by working in the wavelet domain. While this metric does not explicitly account for deformations, there is a close relationship with an approximate algorithm for computing the Earth Mover's Distance [20], which suggests that this metric will also be insensitive to small deformations. Using this local metric, we must solve the following:

$$I_{\text{geod}}(t) = \arg\min_{I(t)} \frac{1}{2} \int_0^1 \sum_{x,y} \frac{\|\nabla \delta I(x, y, t)\|^2}{\|\nabla I(x, y, t)\|^2 + \varepsilon^2} dt$$

Here, t parameterizes the images along a geodesic path traveled starting at $t = 0$, and ending at $t = 1$. We then rewrite the image in the wavelet domain, using orthonormal wavelets whose horizontal and vertical components, H, and V, are approximations to horizontal and vertical first derivative operators. This gives us a local cost:

$$E_{\text{wav}}(I) = \frac{1}{2} \sum_{m,n} \frac{\delta H^2 + \delta V^2}{H^2 + V^2 + \varepsilon^2}$$

In the wavelet domain, each wavelet basis location is now independent of its neighbors, as the local descriptions of the gradients are handled during the wavelet filtering, a result of the orthogonality of the wavelets. This allows us to rewrite the geodesic computation as:

$$I_{\text{geod}}(t) = \frac{1}{2} \sum_{m,n} \arg\min_{H(t), V(t)} \int_0^1 \frac{H'^2 + V'^2}{H^2 + V^2 + \varepsilon^2} dt$$

where H' and V' denote derivatives taken with respect to t. Note that $H(t)$ and $V(t)$, the horizontal and vertical components of the images that lie on the geodesic path, can be computed independently for each location. Each of these minimization problems can be converted to a differential equation using the Euler-Lagrange equations, and solved numerically. The boundary conditions of these equations are $H(0)$, $H(1)$, $V(0)$ and $V(1)$, two corresponding wavelet coefficients in each image. Because the geodesic cost is invariant when both images are rotated together, there are three degrees of freedom in these conditions, allowing us to construct a look-up table for the geodesic cost at a single image location. Computing the geodesic distance between two images, then, can be done by adding together the results of a table look-up for each image location, which can be done in about a millisecond.

When taken at a single scale, this metric is closely linked to our original metric on intensity changes. We then combine information over many scales (see [13, 20]),

building a metric on the $(H, V)_{(m,n)}$ space which is a direct product of the metric on each (H, V) fiber. Mapping to the wavelet domain creates greater stability to deformations.

We show the results of using this distance in Table 9.1. Although it doesn't explicitly account for deformations, this cost, based on geodesic distances, slightly outperforms our method based on optical flow. When combined with a thresholding technique developed in [11], this method produces the best current results on this data set.

9.5 Conclusions

In this chapter, we describe a new, local image distance. Our overall goal is to use this local metric to define an image manifold, in which geodesic distances measure the similarity of images in which shape and lighting may have changed. We have argued analytically that our local metric has some intuitive properties, and begun to evaluate it experimentally. First, we have shown that by integrating our local metric into an optical flow framework we can achieve strong performance on a face recognition task. Second, we have shown that using our new, lighting insensitive metric on intensity changes, we can compute geodesics very efficiently in the wavelet domain, and also achieve excellent face recognition results. Further work remains to develop a complete approach to image comparison in the presence of lighting and shape changes, but these initial results encourage us to believe that this problem can be profitably addressed using Riemannian manifolds.

References

1. Aggarwal JK, Cai Q (1999) Human motion analysis: a review. Comput Vis Image Underst 73:90–102
2. Baker S, Matthews I (2004) Lucas-Kanade 20 years on: a unifying framework. Int J Comput Vis 56(3):221–255
3. Beg MF, Miller MI, Trouvé A, Younes L (2005) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vis 61(2):139–157
4. Brox T, Bruhn A, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for time warping. In: ECCV, vol 4, pp 25–36
5. Chen H, Belhumeur P, Jacobs D (2000) In search of illumination invariants. In: IEEE proc comp vis and pattern recognition, vol I, pp 254–261
6. Cootes TF, Taylor CJ (2001) Statistical models of appearance for medical image analysis and computer vision. In: Proc. SPIE medical imaging, pp 236–248
7. Criminisi A, Blake A, Rother C, Shotton J, Torr PHS (2007) Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. Int J Comput Vis 71(1):89–110
8. Durrleman S, Pennec X, Trouvé A, Thompson P, Ayache N (2008, in press) Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. Med Image Anal

9. Gopalan R, Jacobs D (2010) Comparing and combining lighting insensitive approaches for face recognition. Comput Vis Image Underst 114:135–145
10. Hager G, Belhumeur P (1998) Efficient region tracking with parametric models of geometry and illumination. IEEE Trans Pattern Anal Mach Intell 20(10):1125–1139
11. James AP (2010) Pixel-level decisions based robust face image recognition. In: Oravec M (ed) Face Recognition, chap 5. INTECH, pp 65–86
12. Jorstad A, Jacobs D, Trouvé A (2011) A deformation and lighting insensitive metric for face recognition based on dense correspondence. In: IEEE conference on computer vision and pattern recognition (CVPR)
13. Jorstad A (2012) Measuring deformations and illumination changes in images with applications to face recognition. PhD thesis, University of Maryland
14. Ling H, Jacobs D (2005) Deformation invariant image matching. In: IEEE international conference on computer vision, vol II, pp 1466–1473
15. Martinez A (2003) Recognizing expression variant faces from a single sample image per class. In: CVPR, vol 1, pp 353–358
16. Martinez A, Benavente R (1998) The AR face database. CVC technical report #24
17. Miller MI, Trouvé A, Younes L (2006) Geodesic shooting for computational anatomy. J Math Imaging Vis 24(2):209–228
18. Negahdaripour S (1998) Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. IEEE Trans Pattern Anal Mach Intell 20:961–979
19. Osadchy M, Jacobs D, Lindenbaum M (2007) Surface dependent representations for illumination insensitive image comparison. IEEE Trans Pattern Anal Mach Intell 29(1):98–111
20. Shirdhonkar S, Jacobs D (2008) Approximate earth movers distance in linear time. In: CVPR
21. Song J, Chen B, Wang W, Ren X (2008) Face recognition by fusing binary edge feature and second-order mutual information. In: IEEE conf on cybernetics and intelligent systems, pp 1046–1050
22. Trouvé A, Younes L (2005) Local geometry of deformable templates. SIAM J Math Anal 37(1):17–59
23. Trouvé A, Younes L (2005) Metamorphoses through Lie group action. Found Comput Math 5(2):173–198
24. Zhao S, Gao Y (2008) Significant jet point for facial image representation and recognition. In: ICIP, pp 1664–1667