A Deformation and Lighting Insensitive Metric for Face Recognition Based on Dense Correspondences

Anne Jorstad David Jacobs
UMIACS
University of Maryland, College Park

Alain Trouvé
Centre de Mathématiques et de Leurs Applications
Ecole Normale Supérieure de Cachan

Abstract

Face recognition is a challenging problem, complicated by variations in pose, expression, lighting, and the passage of time. Significant work has been done to solve each of these problems separately. We consider the problems of lighting and expression variation together, proposing a method that accounts for both variabilities within a single model. We present a novel deformation and lighting insensitive metric to compare images, and we present a novel framework to optimize over this metric to calculate dense correspondences between images. Typical correspondence cost patterns are learned between face image pairs and a Naïve Bayes classifier is applied to improve recognition accuracy. Very promising results are presented on the AR Face Database, and we note that our method can be extended to a broad set of applications.

1. Introduction

We aim to solve the problems of expression and lighting variation in face recognition within a single framework. We construct a deformation and lighting insensitive metric that assigns a cost to a pair of images based on their similarity. In order to model variations in expression, establishing point correspondences between faces is essential. Our method determines a dense correspondence flow field between pairs of faces, assigning a cost to each pixel pairing based on a novel image metric.

There are two main contributions in this work: 1) we present a new lighting-insensitive metric based on the effect of lighting in 3D scenes, and 2) we present a new framework for optimizing flow fields making use of the Sobolev gradient and a global kernel, leading to increased stability against deformation. The algorithm presented here is able to find reliable correspondences between images that are taken under very different conditions, and the cost function based on these correspondences results in very good recognition accuracy across classes of structured images with variations in deformation and lighting.

Our new deformation and lighting insensitive metric is a function of image gradients and the difference of image gradients, inspired by the known result that image gradients are insensitive to variations in lighting. To find the best pixel correspondences between image pairs, we minimize the sum of the proposed photometric matching costs at each pixel, added to a regularization term that enforces smoothness across adjacent pixel correspondences using a global kernel. Our optimization scheme minimizes over the correspondence flow field making use of a Sobolev gradient, which is smoother and results in superior rates of convergence. The optimization returns correspondence costs for each image pair, which can be compared to make decisions on identity. Based on the photometric and regularization costs calculated at each pixel, we learn a Naïve Bayes Maximum Likelihood model of how same-person and different-person image pairs typically correspond, and we apply this knowledge to improve our results. Experiments are presented on the AR Face Database, and our method is seen to be competitive with the current state-of-the-art.

The standard method for finding dense correspondences is to determine the optical flow between images. Methods of optical flow have traditionally been developed to measure rigid object motion between images in a video sequence. We emphasize that while we construct a method that involves determining a flow field between pairs of images, our goal is to compute a distance between image pairs, and we are not proposing a new method for solving problems in the general optical flow framework. We will sometimes accept incorrect pixel correspondences if this allows the overall image matching cost to be meaningful.

Considerable research has been dedicated to the problem of lighting variation in faces [5]. Solving the expression and lighting problems together has been attempted in several re-

This research was funded by the Office of the Director of National Intelligence(ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.

cent works. Zhao and Gao [15] use only pixels from an edge map to determine the best point pair correspondences between images based on location and Gabor jet information. Xie and Lam [13] also find correspondences between edge pixels, developing a cost function based on Euclidean distances, Gabor maps and gradient directions at each pixel. In a separate work [14] Xie and Lam model a face as a grid of tiles each of which is allowed to translate, rotate and vary intensity linearly to match a second image. Song et al. [12] combine binary edge features with grayscale information using mutual information. In [6], James presents a method in which a simple local descriptor is calculated at each pixel, descriptors at the same coordinates in two images are compared, and the number of sufficiently similar descriptor pairs are tallied, resulting in a surprisingly robust cost function.

We review the use of optical flow for face recognition in Section 2, present our new metric in Section 3, our optimization scheme is described in Section 4, a probabilistic model is introduced in Section 5 to improve our results, and experiments are presented in Section 6.

2. Optical Flow for Face Recognition

Optical flow determines the displacement of every pixel in an image to the most similar pixel in a second image, returning a vector field over the image. Traditional optical flow is based on the intensity constraint equation, which assumes that corresponding object points in two images will have near equal grayscale values,

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t). \tag{1}$$

Using a first order Taylor expansion

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t \tag{2}$$

gives rise to the photometric term to be minimized

$$E_b = \nabla I \cdot w + I_t, \tag{3}$$

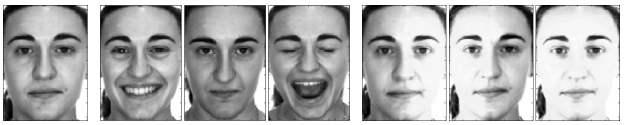
for $w = [\delta x\ \delta y]^T$ and $\delta t = 1$, to which a second term is added to enforce smoothness

$$E_r(w) = |\nabla \delta x|^2 + |\nabla \delta y|^2. \tag{4}$$

Black and Anandan [3] incorporate a robust error function ρ to limit the effect of outliers, allowing them to handle multiple distinct motions in a single image pair by minimizing

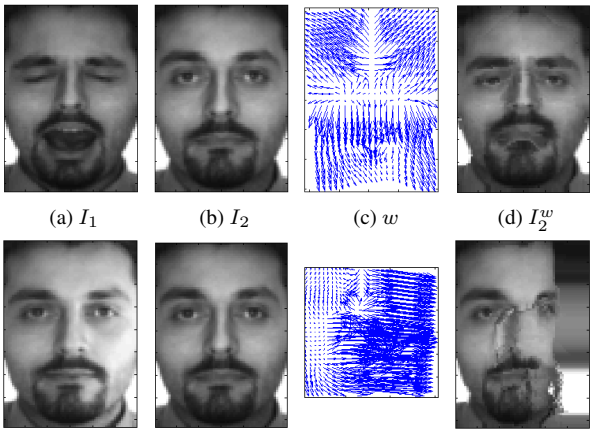
$$E_{B\&A} = \int_\omega \left(\rho_b(E_b^2) + \lambda \rho_r(E_r^2)\right) dxdy. \tag{5}$$

Although optical flow was developed for the rigid object motion tracking problem, it has been successfully applied in face recognition. For example in [2], the flow is calculated between a face and a small variation in pose of that same face. The flow between a new face and the original face is calculated to find correspondences, and then the flow field from the original face is applied to a new face to generate a new pose of the new face. In [8] the length of the flow vectors are used to weight the importance of each pixel before performing image differencing on expression variant image pairs.



(a) (b) (c)

Figure 1. The images of a single individual from the AR Face Database [9]: (a) neutral (b) expression variations (smile, frown, scream) (c) lighting variations (from the right, left, both).



(a) I_1 (b) I_2 (c) w (d) I_2^w

(e) I_1 (f) I_2 (g) w (h) I_2^w

Figure 2. Poor results are achieved when the Black and Anandan flow w is calculated from I_1 to I_2, then the pixels from I_2 are warped backwards along w to generate image I_2^w which corresponds to I_1. The flow here is calculated with a very small regularization weighting. (a)-(d) Change in expression. (e)-(h) Change in lighting.

However, there are limits to using traditional optical flow. The flow between faces is highly nonrigid, often with very large object deformations, and does not involve any intermediate frames between two images separated in time. For example see the expression extremes when comparing Figure 1(a) with the third image in Figure 1(b), or the lighting variations between Figs. 1(a) and 1(c). The challenge of this flow problem is demonstrated using the robust Black and Anandan flow [3], and similar results were observed when using the long range Brox flow [4], which also incorporates a gradient constancy constraint for illumination change robustness. To inspect pixel correspondences, pix-

els from one image can be traced along the flow and pasted into their corresponding positions to create a warped image. When the weight on the regularization term in (5) is very small, it is possible to achieve artificially good-looking results with the Black and Anandan flow, such as in Figure 2(d) generated for $\lambda = 10^{-5}$. Pixels from the tongue in I_1 are matched to lip, skin and beard pixels in I_2, creating false correspondences and a very nonsmooth flow. If the regularization weight is turned up then the resulting flow is almost zero everywhere, and no deformations are captured. If lighting changes are introduced, the method completely breaks down, see Figure 2(h). We want to construct a new metric that can handle large deformations and is insensitive to lighting changes, to be able to find more accurate costs based on dense correspondences between images.

3. A Deformation and Lighting Insensitive Metric

We present a new deformation and lighting insensitive metric, which we will then use in an optical flow-like framework.

3.1. The New Metric

Traditional optical flow relies on the intensity constraint equation (1) to find correspondences between images. Instead of enforcing consistent intensity, we would like to construct a metric where intensities that change as a result of a lighting change in the scene can still be matched. If $w(\vec{x})$ is the flow from image $I_1(\vec{x})$ to image $I_2(\vec{x})$, where $\vec{x}_{ij} = (i, j)$ is the pixel in the $(i, j)^{th}$ position, then $I_2(\vec{x})$ can be warped backwards along this flow to match $I_1(\vec{x})$ by defining

$$I_2^w(\vec{x}) = I_2\left(\vec{x} + w(\vec{x})\right). \tag{6}$$

Any image warped backwards via w will be denoted with a superscript w. Traditional template matching attempts to minimize the warped image difference

$$E_b^{L^2}(w) = \frac{1}{2}\sum_{i,j} \|I_2^w - I_1\|_{L^2}^2. \tag{7}$$

In the image manifold where each point on the manifold is an $M \times N$ image, the usual Euclidean metric defines a structure in a local neighborhood around point I. Letting δI denote an infinitesimal image variation, this infinitesimal metric is $\|\delta I\|_{L^2}$. In the discrete case we take

$$\delta I = I_2^w - I_1, \tag{8}$$

so $\|\delta I\|_{L^2}$ is just (7). Our new metric instead defines a different Riemannian structure on images using the new infinitesimal metric

$$\|\delta I\|_I^2 = \frac{1}{2}\int \frac{\|\nabla \delta I\|^2(x,y)}{\|\nabla I\|^2(x,y) + \epsilon^2} dxdy, \tag{9}$$

where ϵ is a small positive constant of the order of the image noise. As a simple approximation, we then take our new photometric energy term to be

$$E_b(w) = \frac{1}{2}\sum_{i,j} \frac{\|\nabla(I_2^w - I_1)\|^2}{\|\nabla I_1\|^2 + \epsilon^2}, \tag{10}$$

where for the moment the norms and gradients are all taken to follow their standard Euclidean definitions in L^2.

The idea that lighting change on a surface can be represented as multiplication by a scalar and addition by a constant [10] is integrated into the robust optical flow calculation in [7] to develop a lighting-insensitive optical flow algorithm. Our metric goes further, and is designed to be insensitive to intensity changes caused by the effects of lighting variation in 3D scenes. We normalize by the gradient of the image because a high image gradient often signals a rapid change in scene properties, such as a change in albedo or a point with high curvature. At these locations, a change in lighting conditions can have a significant effect on the image gradient. For example, a brighter light can scale the image gradient. Changing the location of a light can magnify or weaken the gradient at the edge of a polyhedron, as the two sides forming the edge are exposed differently to the light. Therefore, at locations with large image gradients, a significant change in the gradient is often due to lighting effects. At the same time, regions with small image gradients often signal scene regions with uniform albedo and surface normals. For Lambertian objects with uniform albedo and surface normals, variations in lighting cannot induce large gradients. Therefore, while it is not impossible for a lighting change to turn a small gradient into a large one, it is less likely, and so is more heavily penalized by our metric.

The derivation of our new metric removes the restriction that movement between images be less than one pixel, a limitation [1] that comes from applying first order finite differencing to a first order Taylor Expansion (2). Many long-range optical flow methods have been developed to get around this restriction, often using hierarchical coarse-to-fine strategies [4]. Our method is able to capture larger movements by optimizing over a dual space related through a global kernel, see Section 3.3, and the new method is seen to handle typical face deformations better than traditional optical flow.

In addition to minimizing E_b, a metric based on similarities between the gradients of the intensities, we also want to take into account the total deformation required to arrive at this similarity, so we include a regularization term E_r that depends on the smoothness of the flow w. Traditional optical flow minimizes the sum of the L_2-norm squared gradients of the flow (4). Instead, we introduce a more general Sobolev-type quadratic cost penalizing irregular w,

$$E_r(w) = \frac{1}{2}\langle K^{-1}w,\ w\rangle_G, \tag{11}$$

where K is a symmetric positive definite matrix as will be discussed below, and the definition of the G-inner product is defined in (13).

Equations (10) and (11) are combined into the proposed Deformation and Lighting Insensitive (DLI) energy function:

$$E_{\text{DLI}}(w) = (1-\lambda)E_b(w) + \lambda E_r(w). \tag{12}$$

In our experiments we take the weighting constant $\lambda = .01$.

3.2. The Sobolev Gradient

Since E_b in (10) involves derivatives, the usual Euclidean gradient $\nabla E_{\text{DLI}}(w)$ will not be smooth enough to be used in an efficient gradient descent method. Instead we use a Sobolev gradient $\nabla_{_K} E_{\text{DLI}}(w)$, which is smoother and results in superior rates of convergence [11], so the optimization scheme gets caught in fewer local minima, and our algorithm is able to arrive efficiently at more accurate solutions. We first define a general inner product

$$\langle u, v\rangle_{_G} = \sum_{i=1}^{M}\sum_{j=1}^{N} \langle u_{ij}, v_{ij}\rangle_{_{\mathbb{R}^2}}. \tag{13}$$

where $G := \mathbb{R}^{M\times N\times 2}$, the dimension of the flow w. Then taking the Sobolev inner product

$$\langle u,\ v\rangle_{_K} = \langle K^{-1}u,\ v\rangle_{_G} \tag{14}$$

used in the regularization term (11), the relation between the regular gradient and the Sobolev gradient is given by

$$\nabla_{_K} f = K\nabla f, \tag{15}$$

where K is a smoothing operator regularizing the Euclidean gradient. To derive (15), it is sufficient to consider the variation δf of any smooth function f and follow the framework of differential forms. The definition of the gradient of a function f for any inner product defined by some K is the unique vector written $\nabla_{_K} f$ satisfying the following equality for any vector w:

$$\delta f = \langle \nabla_{_K} f,\ \delta w\rangle_{_K}. \tag{16}$$

From this, $\delta f = \langle \nabla f(w),\ \delta w\rangle_{_G} = \langle \nabla_{_K} f(w),\ \delta w\rangle_{_K} = \langle K^{-1}\nabla_{_K} f(w),\ \delta w\rangle_{_G}$, and equating the first terms of the $\langle .,.\rangle_{_G}$ expressions we get $\nabla f(w) = K^{-1}\nabla_{_K} f(w)$, which is equivalent to (15). Since $\nabla E_r(w) = K^{-1}w$ directly from (11), we get that $\nabla_{_K} E_r(w) = w$, where K^{-1} no longer appears, and only K is needed for the computation of $\nabla_{_K} E_b = K\nabla E_b$. Here w can be considered as an element of a Reproducing Kernel Hilbert Space (RKHS).

We choose K to be the matrix form of a 2D convolution with a symmetric positive definite kernel k,

$$Ku \equiv k * u, \tag{17}$$

where we abuse notation slightly to consider u as an $MN\times 1$ column vector on the left and as an $M\times N$ image on the right. Here k is an $M\times N$ kernel, and K is the $MN\times MN$ matrix representation of this kernel. Multiplying K by the vector representation of u, (17) holds for corresponding elements. With this choice of K, any matrix-vector product involving K can be computed very efficiently with the Fast Fourier Transform (FFT). We therefore accept periodic boundary conditions, as will be discussed further at the end of Section 4.2.

3.3. Choice of Kernel

The convolution kernel k associated with the matrix K used in (11) must be positive definite in order to define an inner product. We select a Gaussian-like kernel for its smoothing properties. The most obvious choice of such a kernel is defined for all (x, y) as

$$k(x,y) = \exp\left(\frac{-1}{s^2}\left(x^2+y^2\right)\right). \tag{18}$$

We will use derivatives of this kernel to define the derivative filters discussed in Section (4.2). The scale parameter used is $s = 0.0075p$ where p is the perimeter of the image, this value having been empirically determined to be robust.

When defining (11) we instead use a Cauchy kernel which was observed to provide better results experimentally,

$$k(x,y) = \frac{1}{1+\frac{1}{s^2}\left(x^2+y^2\right)}, \tag{19}$$

where the scale parameter $s = \frac{1}{32}p$.

A second kernel is defined for each s with $s_2 = \frac{s}{4}$, and the final kernel is the weighted average of these two kernels ($\frac{1}{4}$ the kernel with smaller scale, $\frac{3}{4}$ the larger). All parameters and kernel choices were tuned on simple synthetic datasets consisting of polygons on a white background, to be as general as possible. At the start of the iterations, the kernel of larger scale dominates, aligning large regions in the image. As the iterations progress, smaller features become more significant and the effect of the smaller kernel predominates.

The kernel has the same dimensions as the image. Convolving with such a global kernel allows our algorithm to capture large-scale image deformations, including long-range translations and large rescalings, that other flow algorithms require multiscale methods to achieve.

4. The Optimization Scheme

The optimization is performed using a modified gradient descent algorithm. To find a point where the energy function $E(w)$ is minimized, we start with $w = 0$, and at every iteration calculate $\nabla_{_K} E$, then update w using a standard

gradient descent update

$$w_{n+1} = w_n - \Delta t \cdot \nabla_K E(w_n). \tag{20}$$

In fact, the actual implementation uses a dual variable α_n such that $w_n = K\alpha_n$ initialized at $\alpha_0 = 0$. Using the fact that $\nabla_K E = K\nabla E$, the update becomes

$$w_n = K\alpha_n \tag{21}$$

$$\alpha_{n+1} = \alpha_n - \Delta t \cdot \nabla E(w_n), \tag{22}$$

which involves only the usual Euclidean gradient. The step size Δt is is initially defined to be 0.01. If an iteration results in a cost smaller than the previous cost, we accept the new α_{n+1} and update $\Delta t = 1.1 \cdot \Delta t$. If an iteration results in a larger cost, then we do not accept the new α, and instead update $\Delta t = 0.5 \cdot \Delta t$ and try again. For the next calculation, we use the α_{n+1} which had resulted in too high a cost, as it was found that this helps move away from local minima as in a rudimentary deterministic annealing algorithm, but no α_{n+1} is accepted as a solution if the cost it produces is not smaller than that at the previous accepted step.

The optimization scheme is terminated when either the gradient at the current α is within a small threshold of zero, or when the size of Δt has been decreased to within a small threshold of zero and no nearby α has resulted in a smaller overall cost. Like all implementations of the Gradient Descent algorithm, our algorithm will usually stop at a local minimum, but it was observed that optimizing over α using Sobolev gradients allows the optimization scheme to proceed much further before terminating.

4.1. The Gradient of the DLI Metric

In order to use a gradient descent method, we must calculate the gradient of the DLI energy function (12),

$$\nabla E_{\text{DLI}}(w) = (1-\lambda)\nabla E_b(w) + \lambda \nabla E_r(w). \tag{23}$$

Since $\nabla E_r(w) = K^{-1}w = \alpha$ we get

$$\nabla E_r(w) = \alpha, \tag{24}$$

and all that remains is to solve for $\nabla E_b(w)$.

4.2. The Gradient of the Photometric Norm

For any given definition of the photometric norm E_b, the regular Euclidean gradient can be calculated directly through applications of the chain rule and finite differencing. However, since this cost involves the computation of derivatives of warped images, we will consider a slightly more general situation using low-pass filtered directional derivatives.

Before describing this more general framework, we consider the simple example of the template matching definition of E_b defined in (7). For this, the gradient would be calculated as

$$\nabla E_b(w) = (I_2^w - I_1)(\nabla I_2)^w, \tag{25}$$

with the warped image gradient term $(\nabla I_2)^w$ resulting from an application of the chain rule. Using the more complex metric for E_b from (10), the gradient could be derived similarly.

Instead, to increase robustness, we make use of more general gradient-like filters with larger regions of support than those used by traditional finite difference methods. Instead of calculating a true gradient ∇I we will instead calculate HI for $H = [H_x\ H_y]^T$, where H_x and H_y represent convolutions with more general x- and y-directional derivative filters h_x and h_y of the low-pass kernel k from (18).

We introduce a diagonal weighting matrix C on $\mathbb{R}^{MN\times MN}$ with dimensions as in (17) to serve as the denominator, with diagonal coefficient

$$C_{ij,ij} = \left(|(H_x I_1)_{ij}|^2 + |(H_y I_1)_{ij}|^2 + \epsilon^2\right)^{-1}. \tag{26}$$

The metric (10) can now be expressed as

$$E_b(w) = \frac{1}{2}\langle CH_x(I_2^w - I_1),\ H_x(I_2^w - I_1)\rangle_{\mathbb{R}^{M\times N}} \tag{27}$$

$$+ \frac{1}{2}\langle CH_y(I_2^w - I_1),\ H_y(I_2^w - I_1)\rangle_{\mathbb{R}^{M\times N}} \tag{28}$$

$$= \frac{1}{2}\langle \Delta_C(I_2^w - I_1),\ (I_2^w - I_1)\rangle_{\mathbb{R}^{M\times N}} \tag{29}$$

where $\Delta_C = H_x^T C H_x + H_y^T C H_y$ is a discrete Laplacian operator combining the directional derivatives and the weighting factors. Note that the multiplication by C has a linear cost with respect to the number of pixels, and the multiplication by H_x^T (respectively H_y^T) is a convolution with the adjoint filter of h_x (respectively h_y).

To calculate the gradient of E_b we will make use of the symmetry of the matrix Δ_C to get

$$\frac{\partial E_b}{\partial w_{ij}}(w) = [\Delta_C(I_2^w - I_1)]_{ij} \nabla I_2(\vec{x}_{ij} + w_{ij}) \tag{30}$$

or equivalently

$$\nabla E_b(w) = [\Delta_C(I_2^w - I_1)](\nabla I_2)^w. \tag{31}$$

To perform the computations efficiently, FFTs are used to compute the convolutions. This means that we accept periodic boundary conditions, despite not having periodic images. In order to avoid driving the optimization by pixels near the boundaries, which are the least important points for our purposes, we multiply the cost function by a weighting function that diminishes the weights of the pixels closest to each boundary smoothly down to zero, thereby approximating periodic boundary conditions. This is implemented by premultiplying (26) with this weighting at each point.

4.3. The Algorithm

The optimal pixel correspondences between images I_1 and I_2 are determined by the flow w from I_1 to I_2 that minimizes the cost $E_{\text{DLI}}(w)$ from (12). The optimization algorithm is summarized in Algorithm 1.

Algorithm 1 Find Optimal Correspondences

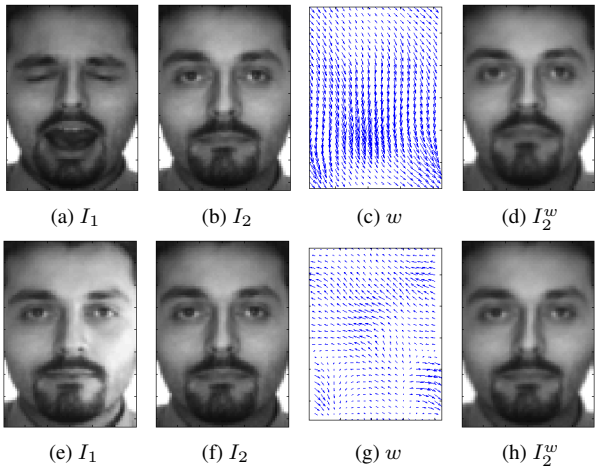
Input images I_1 and I_2, initialize $\alpha_0 = 0$
repeat
  $w_n = k * \alpha_n$
  Calculate $\nabla E(w_n)$ from (23) using (24) and (31)
  $\alpha_{n+1} = \alpha_n - \Delta t \cdot \nabla E(w_n)$, update Δt
until $\|\alpha_{n+1} - \alpha_n\| <$ threshold
return final matching cost from (12)

The optimization takes approximately 1 second to converge for a pair of images of dimension 83×59, running Matlab on a 3.16 GHz processor.

Inspecting representative image pairs reveals that our algorithm is robust to changes in expression and lighting. In Figure 3, the flow w is calculated from I_1 to I_2, then the pixels from I_2 are warped backwards along w to generate I_2^w which corresponds to I_1. We see that in 3(d), the top lip and nose from I_2 has been matched very accurately to the location of the top lip in I_1, and the top of the face has been deformed slightly to align with I_1. Below the top lip, the regularization became more important than pixel intensity matching so the rest of the mouth remained smooth, rather than having the discontinuous flow that would be required to match both closed sets of lips in I_2 to the open lips in I_1. We note that generating flows and warped images is not the goal of our algorithm. We are searching for distance values between image pairs, and we accept some imperfect correspondences when this preserve smoothness. It will be seen in Section 6 that the smooth correspondences we achieve from our calculations are sufficient to serve as the basis for an accurate identification algorithm. In Figure 3(h), the algorithm has accurately detected that although there has been a change in lighting in the scene, there is no deformation of the face, and the calculated flow is small, mostly accounting for imperfect alignment between images.

5. Learning Typical Correspondence Patterns

Because all images are known to be of faces, typical correspondences between faces can be learned via Naïve Bayes classification to improve the recognition results. Based on the cost values obtained from the DLI metric, we learn a Gaussian model at each pixel between faces of the same person across variations in expression and lighting, and we learn a separate model for correspondences between faces of different people, also allowing for variations in expression and lighting. The found correspondence costs between an unknown probe face and a known gallery face can then be compared to each model.



(a) I_1 (b) I_2 (c) w (d) I_2^w

(e) I_1 (f) I_2 (g) w (h) I_2^w

Figure 3. Results from our proposed flow calculation. (a)-(d) The algorithm is robust to large deformations, where the top lip has been correctly matched between images while keeping the overall flow smooth. (e)-(h) The algorithm correctly identifies that in spite of significant change in lighting there has been no deformation, and the flow is small.

After the correspondences between images have been calculated, at each pixel we have a photometric cost in the x- and y-directions, and a regularization cost in the x- and y-directions (recall that the gradient and the flow w both have x- and y- components),

$$E_b(w) = \frac{1}{2}\sum_{i,j} \frac{\|\nabla(I_2^w - I_1)\|^2_{\mathbb{R}^2}}{\|\nabla I_1\|^2_{\mathbb{R}^2} + \epsilon^2} = \frac{1}{2}\sum_{i,j} \left(E^x_{b_{ij}}\right)^2 + \left(E^y_{b_{ij}}\right)^2 \quad (32)$$

$$E_r(w) = \frac{1}{2}\langle K^{-1}w,\ w\rangle_G = \frac{1}{2}(K^{-1}w)^T w = \frac{1}{2}\left(K^{-\frac{1}{2}}w\right)^2 = \frac{1}{2}\sum_{i,j} \left(E^x_{r_{ij}}\right)^2 + \left(E^y_{r_{ij}}\right)^2 \quad (33)$$

The cost vector for an image pair correspondence at each pixel (i,j) is $\vec{E_{ij}} = [E^x_{b_{ij}}\ E^y_{b_{ij}}\ E^x_{r_{ij}}\ E^y_{r_{ij}}]$, and the total cost (12) at each pixel can be rewritten as

$$E_{\text{DLI}}(w)_{ij} = (E^x_{b_{ij}})^2 + (E^y_{b_{ij}})^2 + (E^x_{r_{ij}})^2 + (E^y_{r_{ij}})^2. \quad (34)$$

We can use Maximum Likelihood estimation to learn the typical Gaussian distribution for the flow costs between same person image pairs at each pixel. Given training data of many same person image pairs, we calculate the optimal pixel correspondences between each pair using Algorithm 1. For each pixel, a Gaussian is fit through the 4D cost vectors found for that location. The probability that two new images both come from the same person can then be calculated at each pixel.

Assuming pixel independence, we multiply the probabilities over all pixels in an image for the final probability value. We compute the probability P_{same} that two images are from the same person, and probability P_{diff} that two images are not from the same person, repeating the above process using training data from different person image pairs. The ratio $P_{\text{same}}/P_{\text{diff}}$ is used as the final similarity metric between pairs of face images, as this is a more discriminatory metric than P_{same} alone. In practice we calculate the log likelihood ratio. For a new image pair I_1 and I_2, a new set of cost values $\vec{E}^{new}$ is calculated where each pixel location is as in (34). The final similarity value for this image pairing is then

$$S(I_1, I_2) = \frac{P_{\text{same}}(\vec{E}^{new}(w))}{P_{\text{diff}}(\vec{E}^{new}(w))}. \tag{35}$$

We write this similarity function in terms of the image pair, while in the original DLI energy function (12) the cost was written in terms of the flow between the two images.

6. Experiments

Experiments are performed on the subset of the AR Face Database [9] dealing with expression and lighting, see Figure 1. There are seven images of each individual: a neutral face, three variations in expression (smile, frown, scream), and three variations in lighting (from the left, from the right, from both sides). The standard 100 person aligned and cropped faces are used. We resize each image to be 83×59 pixels, as images of this size return the most accurate results with our algorithm. Similarly resized images have been used successfully in many other algorithms [6, 13, 14]. Our algorithm is fully automatic, so no other input is required.

The neutral faces of all individual are taken to be the gallery, and the other six images of each person are compared to each gallery image separately. We found that warping the neutral images to the non-neutral images is more stable, and so the gallery images take the place of I_2 in our algorithm, and the neutral faces are warped backwards along the calculated flows to generate the I_2^w. Nearest neighbor matching is applied, so that the neutral image that results in the lowest correspondence cost for an unknown non-neutral image defines the identity of the unknown image. Results are presented from the direct output of the optimization scheme minimizing (12) in the first row of Table 1. To use the probabilistic model from Section 5 to maximize (35), half the dataset is used as training data, where the same number of different person image pairs are used as available same person image pairs ($6 \times 50 = 300$), with different person image pairs chosen randomly, given that each type of variation is equally represented. The other half of the data is used for testing. The dataset is divided in half randomly five times, and the average accuracy of the five trials is presented in the second row of Table 1. The same testing galleries are used for both the direct and learned methods. The results of our algorithm are broken down for each expression and lighting variation in Table 2. The lowest observed accuracy is on the challenging "scream" case, where our results are 30% higher than recently reported results [12, 15].

Cost Function	*Expression*	*Lighting*	*Overall*
Direct	82.0%	96.0%	89.0%
After Learning	89.6%	98.9%	94.3%
Smile gallery After Learning	86.8%	91.2%	89.0%
Borders removed After Learning	85.1%	96.4%	90.7%

Table 1. Identification Accuracy found when directly minimizing equation (12), and after applying the probabilistic model from equation (35). Rows 1-2: for a gallery of neutral faces. Row 3: for a gallery of smile faces. Row 4: when 10% of the border pixels have been removed from each edge for a gallery of neutral faces.

Variation	*Accuracy*	*Variation*	*Accuracy*
Smile	97.6%	Left light	98.8%
Frown	91.6%	Right light	99.6%
Scream	79.6%	Both lights	98.4%

Table 2. Identification Accuracy broken down by variation for a gallery of neutral faces.

To test the gains in robustness coming from our new lighting-insensitive photometric energy norm (10), we ran our optimization scheme replacing E_b in (12) with the L^2 warped image difference metric from (7). Results are presented in the first row of Table 3. It is seen that this direct image differencing breaks down when lighting variation is considered, and the new metric presented in this paper is more accurate in all cases.

To test that our algorithm is robust when both lighting and expression are varied at once, we use the smile faces as our gallery, and repeat the above experiment, so that all the lighting variation images are being warped from a neutral face with harsh lighting to a smiling face with ambient lighting. See Table 1. The recognition accuracy of many algorithms is directly related to the alignment of the outline of the head and neck. To test that we are capturing true face information and not simply capturing the head and neck outlines, we remove 10% of the pixels on each edge of the image after the flow has been calculated, and determine the matching cost only from the remaining pixels. From Table 1 we see that very little accuracy is lost. As a comparison, we consider the simple Gradient Direction method, which has been found to be one of the most robust methods against changes in lighting [5]. This method determines the direction of the image gradient at each pixel, and measures the distance between images as the sum of the angles between their gradient directions at each pixel coordinate. The Gradient Direction accuracy decreases by 7% in this case.

When compared to other methods in the literature, the

Method	*Expression*	*Lighting*	*Overall*
Proposed Framework with image differencing	84.0%	8.7%	46.3%
Significant Jet Point [15]	80.8%	91.7%	86.3%
Binary Edge Feature and MI [12]	78.5%	97.0%	87.8%
Gradient Direction [5]	86.0%	96.0%	91.0%
Elastic Shape-Texture Matching [13]	98.3%*	97.2%	97.8%*
Elastic Local Reconstruction [14]	99.2%*	98.6%	98.9%*
Proposed Method	**89.6%**	**98.9%**	**94.3%**
Pixel Level Decisions [6]	99.0%	97.0%	98.0%

Table 3. Comparison with other methods that address both lighting and expression variation on the AR Face Database using a gallery of neutral expression and lighting. *The challenging "scream" case is not included in these expression tests, so these results are not directly comparable.

method proposed here is found to be very competitive, see Table 3. The AR Face Database is a tightly controlled and therefore relatively simple dataset. With a robust error function incorporated into our algorithm to limit the effect of outliers, we expect that our algorithm will be able to handle much less controlled datasets. Unlike other algorithms [6], our method does not rely heavily on input image alignment, as we calculate dense correspondences based on global considerations. We foresee many ways to extend the unified framework presented in this paper to incorporate more robustness, to be able to handle greater variations that cause other algorithms to fail. Nothing in our algorithm is specific to faces, the method can be applied to any class of images with deformations and lighting variation that exhibit a standard structure.

7. Conclusion

Finding reliable image metrics is a fundamental problem in Computer Vision. We have presented an algorithm to perform recognition tasks in the presence of deformation and lighting variations in well-structured images. Our primary contributions are the introduction of a metric that handles lighting variation in a new way, and a method to optimize over this metric. The new lighting-insensitive metric is based on the effect of lighting in 3D scenes. The optimization scheme makes use of smooth Sobolev gradients to efficiently optimize over a flow field that determines dense correspondences between potentially deformed images taken under very different conditions. The mathematics inspiring this work is rigorously motivated. We have validated the efficacy of our metric and optimization scheme by applying them to the problem of expression and lighting variant face recognition. Typical correspondence cost patterns from our metric were learned between face image pairs and a Naïve Bayes classifier was applied to improve recognition accuracy. Our very general algorithm is seen to be competitive with the current state-of-the-art on the AR Face Database, and it lays the groundwork for many possible extensions to handle significantly more challenging datasets.

8. Acknowledgments

We would like to thank D. Sun and S. Roth for making their implementation of the Black and Anandan Optical Flow [3] available. We would also like to thank A. James for making the Pixel Level Decisions [6] code available. This material is based upon work supported by the National Science Foundation under Grant No. 0915977.

References

[1] J. Barron, D. Fleet, and S. Beauchemin. Performance of Optical Flow Techniques. *IJCV*, 12:43–77, 1994.

[2] D. Beymer and T. Poggio. Face Recognition From One Example View. *IEEE Conf. on Computer Vision*, p. 500, 1995.

[3] M. J. Black and P. Anandan. The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields. *Computer Vision and Image Understanding*, 63:75–104, 1996.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Time Warping. *ECCV*, 4:25–36, 2004.

[5] R. Gopalan and D. Jacobs. Comparing and Combining Lighting Insensitive Approaches for Face Recognition. *Computer Vision and Image Understanding*, 114:135–145, 2010.

[6] A. P. James. Pixel-Level Decisions Based Robust Face Image Recognition. In M. Oravec, editor, *Face Recognition*, chapter 5, pp. 65–86. INTECH, 2010.

[7] Y.-H. Kim, A. M. Martinez, and A. C. Kak. Robust Motion Estimation Under Varying Illumination. *Image and Vision Computing*, 23, 2004.

[8] A. Martinez. Recognizing Expression Variant Faces From a Single Sample Image per Class. *CVPR*, 1:353–358, 2003.

[9] A. Martinez and R. Benavente. The AR Face Database. *CVC Technical Report #24*, 1998.

[10] S. Negahdaripour. Revised Definition of Optical Flow: Integration of Radiometric and Geometric Cues for Dynamic Scene Analysis. *PAMI*, 20:961–979, 1998.

[11] J. W. Neuberger. *Sobolev Gradients and Differential Equations, 2nd Edition*. Springer, 2010.

[12] J. Song, B. Chen, W. Wang, and X. Ren. Face Recognition by Fusing Binary Edge Feature and Second-Order Mutual Information. In *IEEE Conf. on Cybernetics and Intelligent Systems*, pp. 1046–1050, 2008.

[13] X. Xie and K.-M. Lam. Elastic Shape-Texture Matching for Human Face Recognition. *Pattern Recogn.*, 41:396–405, 2008.

[14] X. Xie and K.-M. Lam. Face Recognition Using Elastic Local Reconstruction Based on a Single Face Image. *Pattern Recogn.*, 41:406–417, 2008.

[15] S. Zhao and Y. Gao. Significant Jet Point For Facial Image Representation and Recognition. In *ICIP*, pp. 1664–1667, 2008.