Distribution Fields with Adaptive Kernels for Large Displacement Image Alignment

Benjamin Mears
bmears@cs.umass.edu
Laura Sevilla Lara
lsevilla@cs.umass.edu
Erik Learned-Miller
elm@cs.umass.edu

School of Computer Science
UMass Amherst
Amherst, MA

While region-based image alignment algorithms that use gradient descent can achieve sub-pixel accuracy when they converge, their convergence depends on the smoothness of the image intensity values. Image smoothness is often enforced through the use of multi-scale approaches in which images are smoothed and downsampled. Yet, these approaches typically use fixed smoothing parameters which may be appropriate for some images but not for others. Even for a particular image, the optimal smoothing parameters may depend on the magnitude of the transformation. When the transformation is large, the image should be smoothed more than when the transformation is small. Further, with gradient-based approaches, the optimal smoothing parameters may change with each iteration as the algorithm proceeds towards convergence.

We address convergence issues related to the choice of smoothing parameters by deriving a Gauss-Newton gradient descent algorithm based on distribution fields (DFs) and proposing a method to dynamically select smoothing parameters at each iteration. DFs have previously been used in the context of tracking [6]. In this work we incorporate DFs into a full affine model for region-based alignment and simultaneously search over parameterized sets of geometric and photometric transforms. We use a probabilistic interpretation of DFs to select smoothing parameters at each step in the optimization and show that this results in improved convergence rates.

Following the notation of Baker and Matthews [1], let $T(\mathbf{x})$ be an image containing a fixed region for which we want to find the corresponding region in $I(\mathbf{x})$, where $\mathbf{x} = (x,y)^{\mathrm{T}}$ is a column vector. We refer to $T(\mathbf{x})$ as the template image and $I(\mathbf{x})$ as the input image. Let $W(\mathbf{x};\mathbf{p})$ be the parameterized set of warps, where $\mathbf{p}$ are the parameters. The goal of region-based alignment is to find the $\hat{\mathbf{p}}$ that minimizes some distance measure between $T(\mathbf{x})$ and $I(\mathbf{W}(\mathbf{x};\hat{\mathbf{p}}))$.

One of the early image alignment algorithms was the gradient-based Lucas-Kanade (LK) algorithm [3]. In the LK method, the feature space consists of intensity values and the similarity measure is the L2 distance. Our method is inspired by the LK algorithm, but we replace the image and template with their DF representations. The basic idea of a DF is to represent a region in an image as a normalized histogram, i.e., a probability distribution, over feature values at each pixel. In this work we use grayscale intensity values as the feature values although other features could be used instead (e.g. edge intensities, RGB values for color images, etc.). The simplest DF consists of probability distributions over binned intensity values where each probability distribution is degenerate and is given by

$$D(I,\mathbf{x},f) = \begin{cases} 1 & \text{if } I(\mathbf{x}) \in \text{bin } f \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $D(I,\mathbf{x},f)$ is the value of the DF for image I at position $\mathbf{x}$ and bin f.

Using Eq. (1) alone to represent an image provides few additional benefits over using the image itself. Indeed, when the number of bins equals the number of intensity values, the representation contains the same information as the image. Additional benefits can be gained though if the DF is "smoothed" to "spread" the information in the image. In particular, the DF can be convolved with a three-dimensional Gaussian filter with a standard deviation of σ_{xy} in the spatial directions and σ_f in the feature space dimension. By convolving with a Gaussian filter, some degree of uncertainty is allowed in both the location and value of an image pixel. Like image blurring, convolving a DF with a Gaussian filter spreads information about intensity values to neighboring pixels, but does so with a smaller loss of information.

There are various trade-offs that need to be considered when choosing the σ_{xy} and σ_f values used for smoothing the DFs for alignment. For instance, a larger σ_{xy} value may allow for a larger basin of attraction but may result in a less precise final alignment. And if the σ_{xy} value is chosen to be too large or too small, it can cause the algorithm to diverge.

Rather than choose fixed σ values, the values can be chosen automatically based on the current location in the search space. In this work, the σ values are chosen using the probabilistic view of DFs by maximizing the log likelihood of the current warped input image under the DF of the template image. Treating the DF of one image as an independent pixel model, this is defined to be the sum of the log probabilities of each pixel of the current warped input image under the corresponding probability distribution of the template image's DF. Our method is similar to the approach used by Narayana *et al.* for choosing the pixelwise kernel variances in their background subtraction algorithm [4, 5].

In our approach, the log likelihood of the current warped input image, $I' = W(\mathbf{p})$, under the DF of the template, smoothed using the parameters σ_{xy} and σ_f, is given by

$$l(\sigma = \{\sigma_{xy}, \sigma_f\}|T,I',R) = \sum_{\mathbf{x}\in R} \log\left(D_\sigma(T,\mathbf{x},\text{bin}(I'(\mathbf{x}))\right), \tag{2}$$

where bin is the binning function that takes an intensity value and maps it to the appropriate histogram bin.

Since truncated Gaussian kernels are used for efficiency to smooth the DFs, it is possible that some entries of a DF are zero. To deal with the problem of zero probabilities (in which a single outlier can cause the likelihood to be zero), we replace $\log\left(D_\sigma(T,\mathbf{x},\text{bin}(I'(\mathbf{x}))\right)$ in Eq. (2) with $\log\left(\max\left(.0001, D_\sigma(T,\mathbf{x},\text{bin}(I'(\mathbf{x}))\right)\right)$. At each iteration in our method, an exhaustive search is performed over a finite set of σ_{xy} and σ_f values and the σ_{xy} and σ_f values that maximize Eq. (2) are used to convolve the two DFs.

At each iteration of a forward-inverse compositional Gauss Newton optimization, we use the above method to dynamically select the σ values used to blur the DFs of the current warped input image and template. Further, similar to the simultaneous inverse compositional (SIC) algorithm described by Baker and Matthews [2], we extend the algorithm to also search over bias and gain parameters. We achieve impressive convergence results compared to other existing algorithms and also show that the adaptive kernel parameters produce convergence rates better than or equal to the best convergence rates produced by any of a large set of fixed parameter values.

[1] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004.

[2] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, CMU Robotics Institute, December 2003.

[3] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, 1981.

[4] M. Narayana, A. Hanson, and E. Learned-Miller. Background modeling using adaptive pixelwise kernel variances in a hybrid feature space. In *CVPR*, 2012.

[5] M. Narayana, A. Hanson, and E. Learned-Miller. Improvements in joint domain-range modeling for background subtraction. In *BMVC*, 2012.

[6] L. Sevilla-Lara and E. Learned-Miller. Distribution fields for tracking. In *CVPR*, 2012.