Correspondence with Cumulative Similiarity Transforms

Trevor Darrell, *Member*, *IEEE*, and
Michele Covell, *Member*, *IEEE*

Abstract—A local image transform based on cumulative similarity measures is defined and is shown to enable efficient correspondence and tracking near occluding boundaries. Unlike traditional methods, this transform allows correspondences to be found when the only contrast present is the occluding boundary itself and when the sign of contrast along the boundary is possibly reversed. The transform is based on the idea of a cumulative similarity measure which characterizes the shape of local image homogeneity; both the value of an image at a particular point and the shape of the region with locally similar and connected values is captured. This representation is insensitive to structure beyond an occluding boundary but is sensitive to the shape of the boundary itself, which is often an important cue. We show results comparing this method to traditional least-squares and robust correspondence matching.

Index Terms—Image correspondence, stereo, motion, contour tracking.

1 Introduction

FINDING corresponding points in image pairs or image sequences is a central problem in computer vision. Most classical methods assume brightness constancy and perform best on high-contrast regions that lie on a single surface. However, many images have visually important features that lie on occluding boundaries between surfaces. If the surfaces themselves are relatively smooth and low contrast, then the only contrast present will be from the occlusion edge. Correspondences at these feature points are readily apparent to human observers, but existing computational techniques fail to find them.

For many detailed image analysis/synthesis tasks, finding precise correspondences at object boundaries is extremely important. Image compositing [25], automatic morphing [15], and video resynthesis [12] all require accurate correspondence and slight flaws can yield perceptually significant errors. To obtain good results, prior approaches for these tasks have relied on either extreme redundancy of measurement, human-assisted tracking, substantial smoothing, or domain-specific feature-appearance models.

In this paper, we describe a new method that can find correspondences when the only contrast is due to an occlusion event (e.g., the foreground is smooth.) Our method uses purely local image analysis, without prior training, and without smoothing or pooling of motion estimates. We define an image transform which characterizes the local structure of an image in a manner insensitive to points in an occluded region, but which *is* sensitive to the shape of the occlusion boundary itself. In essence, our method performs matching on a redundant, local representation of image homogeneity. In this paper, we show examples where color is the attribute analyzed for homogeneity, but our method is applicable to other local image characteristics such as texture, range data, or achromatic image intensity.

- *T. Darrell is with the MIT AI Lab, Room NE43-829, 545 Technology Square, Cambridge MA 02139. E-mail: trevor@ai.mit.edu.*
- *M. Covell is with YesVideo.com, 2192 Fortune Drive, San Jose, CA 95131. E-mail: covell@ieee.org.*

Manuscript received 18 Aug 2000; revised 28 June 2000; accepted 25 Sept. 2000.
Recommended for acceptance by M. Irani.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 110449.

2 Previous Work

In recent years, several researchers have addressed the problem of local image analysis in the presence of occlusion. Methods from the robust estimation literature have been applied to the correspondence problem and have been shown to considerably improve performance near occlusion boundaries.

Black and Anandan pioneered optic flow using robust error norms—these methods substantially discount the effect of outlier contamination due to pixels from a nonforeground surface [5]. Shizawa and Mase derived methods for transparent local flow estimation [30], using an approach which relies on the computation of higher-order image derivatives. Others have explored nonparametric local transforms for correspondence: Bhat and Nayar have advocated the use of rank statistics for robust correspondence [4], while Zabih and Woodfill use ordering statistics combined with spatial structure in the CENSUS transform [37]. Boykov et al. and others have explored correspondence methods which can adapt to use local windows of variable size [26], [24], [11].

However, these methods make a critical assumption: that there will be sufficient contrast in the foreground portion of an analysis window to localize the correspondence match. This is often not true, due either to a uniform foreground surface or low-resolution video sampling. This problem is illustrated in Fig. 1, which shows a foreground region with zero contrast in front of two different background regions. Note that the sign of contrast at the occlusion boundary changes between the two frames. An example in real imagery is shown in Fig. 2: The marked locations pose a considerable challenge for existing robust correspondence methods, since any window large enough to include substantial foreground contrast will include a very large percentage of background pixels.

Most robust and nonrobust correspondence methods fail when there is no coherent foreground contrast. Transparent-motion analysis [30], [22], [31], [18] can potentially detect motion in these difficult cases, but has not been generally able to provide precise spatial localization of corresponding points (but see [35]). Smoothing methods, such as regularization or parametric motion constraints (affine [3], [6], [20] or learned from examples [7]), can provide approximate localization when good motion estimates are available in nearby image regions, but this is not always the case. If a corpus of training images is available, techniques for feature or appearance modeling can solve these problems, cf. [16], [8].

Several authors have explored methods of finding image layers to pool motion information over arbitrarily shaped regions of support and to iteratively refine parameter estimates [17], [2], [33], but these methods generally assume models of global object motion to define coherence. Token-level edge correspondence are another possible solution, but many images are not amenable to stable, automatic edge extraction. More problematically, edge analysis discards information about image attribute value on both sides of an occlusion edge. This is suboptimal since information from the foreground side can be of value in establishing proper correspondence. Similarly, if a global region segmentation can be performed, correspondence at the surface-shape level can be performed and will solve the problems posed in the previous figures. But, it has proven difficult to obtain global region segmentations that are conserved across corresponding pairs and, even when they are available, the resulting correspondences are extremely sparse since there are far fewer regions than pixels in the image.

As we shall show in this paper, a relatively simple local image transform can easily address the above cases and return pixel-level image correspondences without a global edge or region segmentation, or a global motion model. The essential idea is to consider a transform which captures the local support of the foreground region, and to then perform matching using these support

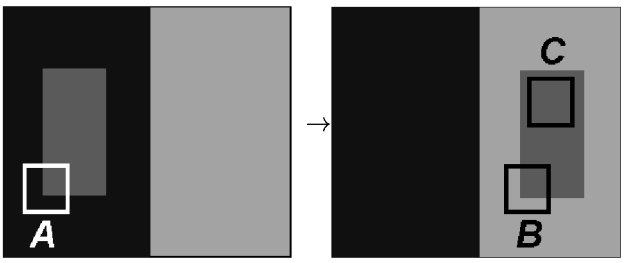

Fig. 1. Correspondence is difficult when a uniform surface moves across different background patterns. Consider the correspondence of window A with windows B or C. Perceptually, window A is closer to window B than window C. But, traditional methods will find the distance between windows A and B to be larger than that between A and C if distance is computed by summing a per-pixel error measure.

functions. Since we do not know the segmentation a priori, we will make conservative guesses at the local support shape by assuming that image regions which are homogeneous in a given attribute value (say, color) belong to the same surface.

3 Radial Cumulative Similarity

Our approach is to define a local image transform with certain desired properties that make features at the occluding boundaries of low-contrast surfaces particularly salient. We apply this transform before using traditional neighborhood distance and contour tracking methods.

Since image contrast determines the ability to find unique correspondences, we design our transform by considering the sources of contrast within a local image window that contains an occlusion boundary. We would like a transform which ignores contrast from the background surfaces but is sensitive to contrast from the occluding boundaries and surface markings of the foreground layer. In general, one does not know a priori whether contrast within a particular window is entirely within the foreground layer, is due to the occlusion boundary between foreground and background, or is entirely within a background layer. We define the "foreground" to be the scene layer on which the central point of the window resides; points on all other layers are considered "background." Ideally, when contrast is in the foreground layer, a transform would capture it fully, both in magnitude and sign. When the contrast is due to an occlusion edge, it is reasonable only to define a template based on the contrast energy, since the sign of contrast is arbitrary with changing background. When contrast is in the background layer, it should be ignored in an ideal template.

We define a robust local image representation that approximates this ideal, without any prior knowledge of the occlusion location. We capture the contrast energy closest to the center of the window, and attenuate all else. Our representation is comprised of a central image-attribute value (typically color) and a local neighborhood of this attribute. The neighborhood computes the local contrast relative to the central attribute attenuated to discount background influence. Many different processes could be used to attenuate background influence, in this paper, we explore radial cumulative probability functions. We compute a measure proportional to the cumulative likelihood that the image attribute is unchanged along the ray from the template center to a particular neighborhood point.

Formally, given a discrete image function $\mathbf{I}(x, y)$, we compute a local robust representation comprised of two terms, a central value, and a neighborhood function:

$$\mathcal{R}_{\mathbf{I}}(x, y) = \{\mathbf{C}_{\mathbf{I}}(x, y), N_{\mathbf{I}}(x, y, r, \theta)\}.$$

The central value is simply the image value averaged within a small radius of the given image location:

$$\mathbf{C}_{\mathbf{I}}(x, y) = \frac{1}{(2\pi(M_c)^2)} \sum_{r,\theta}^{r \leq M_c} \mathbf{I}(x + r\cos\theta, y + r\sin\theta).$$

We typically keep the central region small, with radius (M_c) of a few pixels.

Our neighborhood function is proportional to the likelihood, for each pixel in the window, that the underlying image attribute is unchanged along a ray from the center of the window to that pixel. To compute the neighborhood function, we first compute the local image-contrast energy, which is simply the MSE between the central point at which the transform is being defined (x, y) and nearby points at a given radial offset (r, θ):

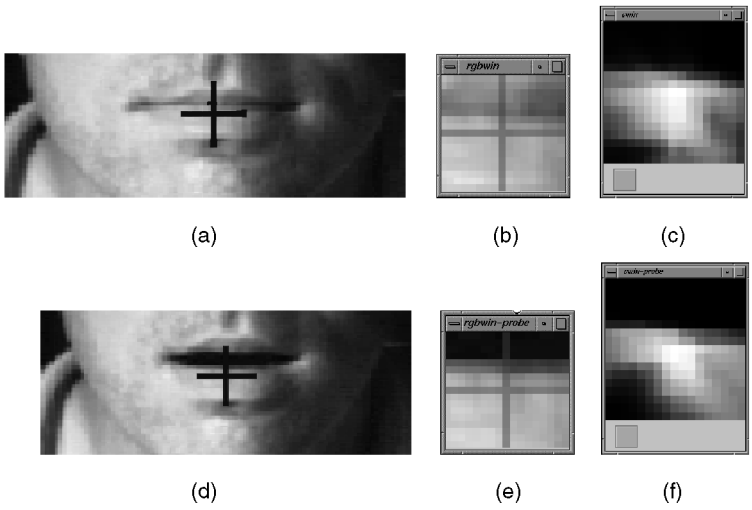

Fig. 2. Finding local correspondences is also difficult in regions with changing scene topology. Images (a) and (d) were taken before and after user's expression changes; (b) and (e) are enlarged views of corresponding points with a cross drawn to indicate the center point of the window. Traditional correspondence methods have difficulty at points such as these, where there is little foreground texture, substantial occlusion, and variable sign of contrast at the occlusion boundary. The RCS transform is stable despite occlusion boundaries of different contrast sign. Images (c) and (f) show the RCS transform of the windows in (b) (e). The RCS transform is shown as two windows, one indicating the neighborhood function, and a smaller one indicating the central attribute value. The computations were performed on color images, which can be viewed at http://www.ai.mit.edu/~trevor/papers/pami110449/.

$$\mathbf{E_I}(x,y,r,\theta) = \alpha||(\mathbf{C_I}(x,y) - \mathbf{I}(x + r\cos\theta, y + r\sin\theta))||^2,$$

where α is a contrast sensitivity coefficient. We then take the integral of $\mathbf{E}$ along a ray from the central point, and take the negative exponential to obtain our neighborhood function:

$$N_\mathbf{I}(x,y,r,\theta) = \exp\{-\int_{\rho<r} \mathbf{E_I}(x,y,\rho,\theta)dp\}.$$

$N_\mathbf{I}$ and $\mathbf{E_I}$ are defined over window coordinates $r \leq M_n$ and $0 \leq \theta \leq 2\pi$. M_n can be set as a parameter of the system, or determined automatically as the largest value needed to include all values of N_I above a small threshold.

We call the representation $\mathcal{R}$ the *Radial Cumulative Similarity* (RCS) transform, since it reflects the radial homogeneity of a given attribute value at a particular point. The main benefit of the RCS transform is invariance to sign of contrast at an occluding boundary as well as invariance to background contrast, while remaining sensitive to foreground and occluding edge contrast energy. This can be seen by the similarity of Figs. 2c and 2f; despite dissimilar background structure and occlusion contrast sign reversal, the RCS transformed pairs are very similar.

We choose the radial integration step for its mathematical and conceptual simplicity, but several other neighborhood functions are possible. In general, one could use any pixel-fill or diffusion operators which finds a set of locally connected and similar pixels. These latter methods would have the additional advantage that they could capture nonconvex local support structure.

4 Finding Point Correspondences

To find a correspondence result for a point (x, y) in an image $\mathbf{I}$, we can simply search for the displacement in a second image $\mathbf{I}'$ which has minimum RCS distance:

$$(u_\mathbf{x}, v_\mathbf{x}) = \arg \min_{u,v=-M_w}^{u,v\leq M_w} D_\lambda(\mathcal{R}_\mathbf{I}(x,y), \mathcal{R}'_\mathbf{I}(x+u, y+v)),$$

where $\mathbf{x} = (x, y)$.

We define RCS distance by computing the weighted L_2 error of the transformed data using a combination of neighborhood difference and central value difference terms.

$$D_\lambda(\mathcal{R}_\mathbf{I}(x,y), \mathcal{R}_{\mathbf{I}'}(x',y')) = (1-\lambda)\Delta N + \lambda \Delta C.$$

The neighborhood difference ΔN is defined as the MSE between $N_\mathbf{I}(x,y,r,\theta)$ and $N_{\mathbf{I}'}(x',y',r,\theta)$ computed over r, θ; ΔC is the MSE between $\mathbf{C}$ and $\mathbf{C}'$.

The bias term λ expresses a trade-off between the contribution of the central attribute error and the neighborhood function error. Generally, the neighborhood error is the most important, since it captures the spatial structure at the given point. However, in certain cases of spatial ambiguity, the central attribute value is critical for making the correct match unambiguous. For example, in the image shown in Fig. 2d, the neighborhood component of the RCS transform would be roughly equal for the marked point and a point located just below the top lip (centered in the dark region of the open mouth). A modest value of λ disambiguates this case.

However, central attribute information alone is clearly inadequate to establish reliable correspondences—in the same figure, many spatially disjoint points may have equal attribute value (color or hue). Using $\lambda = 1$ would lead to random correspondences being indicated between all such equal color points. We have empirically compared the performance under varying values of the λ parameter, and found values between $0.05 \leq \lambda \leq 0.25$ seem to give the best results. In all of the examples presented in this paper, we set $\lambda = 0.1$.

4.1 Anisotropic Regularization

For many image features, the simple matching procedure described above—without any regularization step—can yield good correspondences. However, to overcome errors due to the aperture problem and sensor noise, it is often necessary to regularize or smooth correspondence results. Unfortunately, isotropic smoothing or regularization have negative effects at object boundaries, as information from multiple object surfaces may be pooled together. Fortunately, the RCS representation offers a natural way to overcome this limitation since the neighborhood similarity function is a natural mask to use in anisotropic smoothing.

The simplest approach is to smooth estimated displacements, using the neighborhood function as a weight:

$$(u^+_\mathbf{x}, v^+_\mathbf{x}) = \frac{1}{\sum W_\mathbf{I}} \sum_{i,j=-M_s}^{i,j\leq M_s} W_\mathbf{I}(x,y,i,j)(u_{\mathbf{x}'}, v_{\mathbf{x}'}),$$

where $\mathbf{x}' = \mathbf{x} + (i, j)$ and $W_\mathbf{I}$ is the Cartesian sampling of $N_\mathbf{I}$, $W_\mathbf{I}(x,y,i,j) = N_\mathbf{I}(x,y,\sqrt{i^2+j^2}, atan2(j,i))$.

The weighting function removes the influence of outlier points in the smoothing process. However, taking the spatial average of estimated displacement is nonrobust when there are multiple peaks in the displacement distance function. Even with a perfect weighting mask which discounts all nonforeground points, it can easily be the case that a particular point in a regular pattern can match equally well against several locations. It is, of course, subject to debate how one should represent this and how to choose which peak(s) to return. But, it is clear that one should not return an answer that is intermediate between the two peaks, which would be the result of naively averaging the estimated displacements.

Our preferred method is to instead find the mode of the smoothed distribution. We integrate the RCS distance over a spatial smoothing window $-M_s \leq i, j \leq M_s$, weighted by the neighborhood map evaluated at the chosen point:

$$(u^*_\mathbf{x}, v^*_\mathbf{x}) = \arg \min_{u,v=-M_w}^{u,v\leq M_w} \sum_{i,j=-M_s}^{i,j\leq M_s} W_\mathbf{I}(x,y,i,j) D_\lambda(\mathcal{R}_\mathbf{I}(x+i, y+j), \mathcal{R}'_\mathbf{I}(x+i+u, y+j+v)).$$

4.2 Validity Test

Correspondence search computed with the RCS transform generally outperforms traditional methods near the occluding boundaries of otherwise smooth (low-texture) objects. However, RCS-based correspondence may perform worse at interior points, especially those which are highly textured. This lower performance of RCS away from occlusion boundaries is not surprising: when analyzing an image window of a single surface where brightness constancy holds and there is no occlusion, suboptimal performance will result since the RCS neighborhood function does not capture all the possible contrast in the window. In addition, if there is high contrast in both spatial directions the neighborhood function may well be degenerate, reducing to approximately a single point.

Fortunately, occlusion-free regions of high contrast are cases where traditional methods perform exceedingly well and it is straightforward to devise a test for the validity of the RCS-based result.

We use the results of a search using an L2-norm to determine the appropriateness of the RCS transform. Failure of L2 to find a match below a certain residual threshold indicates a point at which RCS should be used. This hybrid approach is more accurate since, when an L2 match is valid, it is generally more precise than an RCS match at the same point. However, it is also more computationally expensive, as it can require the computation of two full

Fig. 3. Correspondence search for two example features. Top row shows an image pair. Left column shows two features selected from left image. Images labeled (R, Rn, Rc, L2, Rs, Ra) show the distance surface of the feature compared to each possible offset in the right image. Images (R, Rn, Rc) are computed using the RCS transform with no smoothing and varying λ to illustrate the trade off between neighborhood and color terms (R, Rc, and Rn use $\lambda = 0.1, 1.0, 0.0$, respectively). (L2) shows the result using mean-squared-error distance; result with robust norms were qualitatively similar. (Rs) shows the result using an RCS transform ($\lambda = 0.1$) and isotropic smoothing. (Ra) shows the result using the anisotropic smoothing method described in the text. The corner feature is not near an occlusion boundary, so L2 and RCS methods all have distinct minima at the correct correspondence, and both isotropic and anisotropic smoothing preserve that minima. However, the finger feature is dominated by contrast due to an occlusion boundary and, thus, L2-based methods and isotropic smoothing approaches fail to find the correct correspondence.

correspondence searches (L2 followed by RCS) per pixel.[1] The threshold is computed empirically by evaluating the minimum L2 error over regions of the image known to be from a single object; the threshold is set to be approximately two standard deviations above the mean of this error.

Fig. 3 shows the results of correspondence search using features in an image pair. (See caption text.) From the results for the corner feature, we can see that a distinct minima is visible when using unsmoothed RCS, L2 distance, isotropic smoothed RCS, and anisotropic smoothed RCS. This is unsurprising; since this feature point is not near the boundary of a moving object, the L2 metric or isotropic smoothing will yield the correct result. However, the finger-tip feature is near an occluding boundary. We can see the inferior result provided by the L2 norm in this case, as well as by an isotropic smoothing. Only the anisotropic smoothed RCS result offers a distance image which is relatively smooth and whose minima corresponds to the correct result.

4.3 Contour Tracking Application

The RCS transform can be also be used to track an entire occluding contour in an image, using dynamic contour models. Tracking visual features in a series of images is an important task both for vision-based control and for segmentation applications. Dynamic contours [27] and related active tracking techniques are well-suited for both applications because they combine simple, light-weight object models with rapid updates. Dynamic contours track boundaries by minimizing the sum of an external force based on

1. If precomputing L2 correspondences is too computationally expensive, another approach would be to test explicitly for degeneracy of the neighborhood function. Regions that are prone to aliasing can be detected by checking whether the magnitude of the radial cumulative similarity function, $||N||$, is below a certain threshold; if so the RCS transform should not be used.

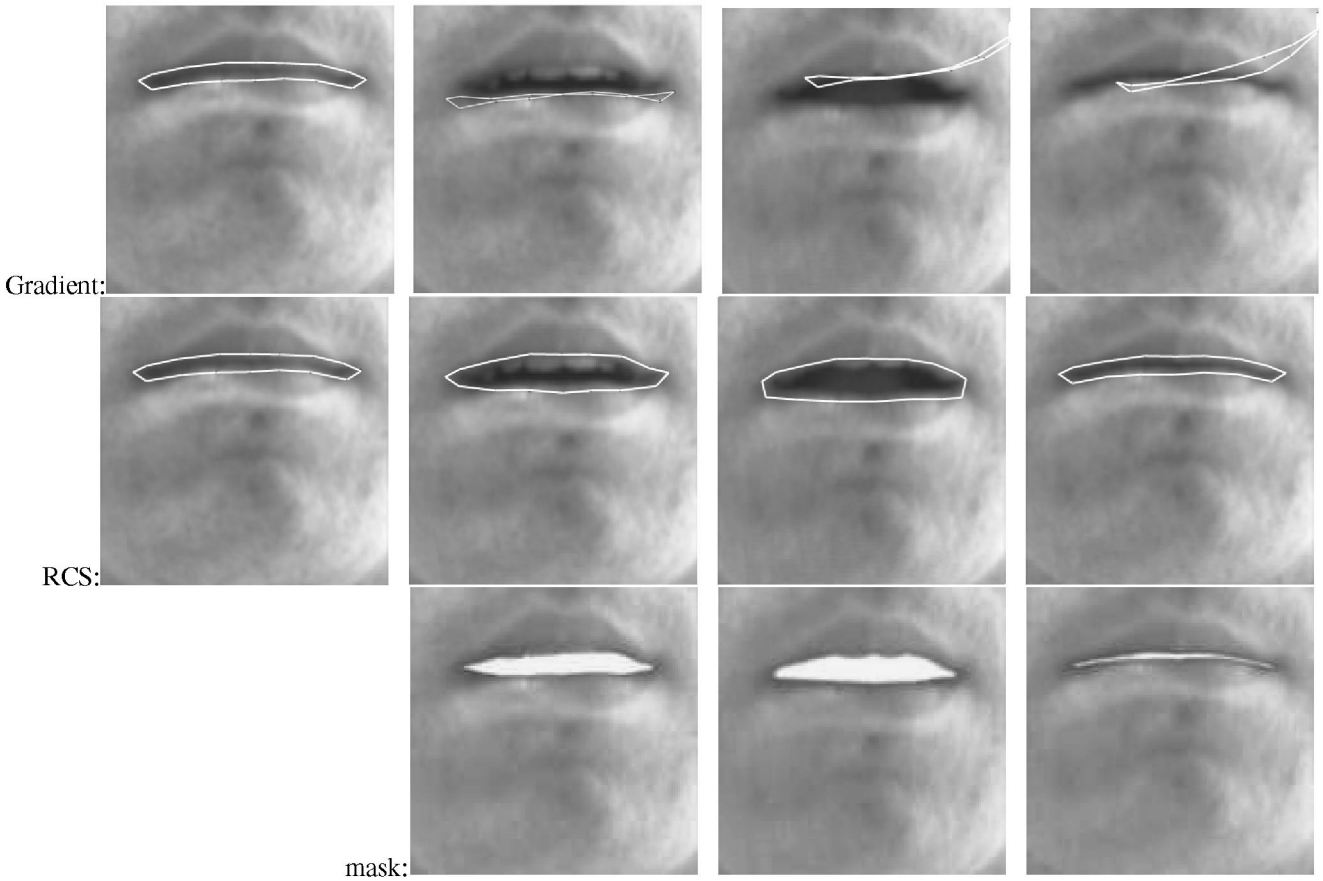

Fig. 4. The RCS transform can also aid tracking when using a dynamic contour model, by defining the external energy term using RCS distance measures. The top image sequence shows the tracking result obtained with a gradient-based external energy term; the second sequence shows the tracking result where RCS-based external energy was used. The contour was initialized by hand in the first frame, four frames from sequence of 110 are shown here. This sequence is difficult for the gradient-based method since the lower edge of the upper teeth is similar to the lower edge of the upper lip in the gradient image, using the RCS these are easily distinguished. The RCS neighborhood function can also be used to define a segmentation mask, as shown in bottom sequence.

a local image measure, and an internal force, based on a shape-dynamics model. A dynamic contour tracks the indicated boundary by finding the shape that minimizes the combined external and internal forces. The external force drives the dynamic contour according to the current image appearance, while the internal force increases the spatial and temporal continuity of the tracked boundary.

Dynamic contours usually employ a simple image contrast measure to define the external forces on the model. This approach works well as long as the boundary being tracked is not an occluding boundary, such as that of a silhouette. When the boundary to be tracked is an occluding boundary, the dynamic contour often confuses background texture for the desired boundary.[2] To overcome this problem, we use an RCS-based external energy term for the dynamic contour model. As with point correspondences, our image model describes the local contrast energy pattern but is largely insensitive to changes in background contrast.

Given a dynamic contour model, we can simply associate each nodal point with an RCS transform value taken directly from the image pattern at the initial nodal points of the user-drawn contour. The external force in subsequent frames is then computed according to the distance between the desired RCS transform value and the value of the proposed node location. If rotational invariance is desired, we can formulate a 1D version of the RCS transform, using a ray from each RCS transform evaluated along the full length of the dynamic contour. We use rays that are perpendicular to the local contour orientation. This also avoids the ambiguities introduced by matching sparse nodal points since the external force term is determined by the RCS appearance along the full length of the contour.

We can also define a segmentation mask for the selected object out of the sequence automatically, using a continuous-valued alpha-channel sequence created from the contour position and from the RCS profiles along the contour. (For details of this method, see [14].) We tested our tracking and segmentation performance on an image sequence containing a mouth opening and closing. The mouth sequence shows a low-texture, deformable occlusion (the inside boundary of the lips), with a similarly colored background (the teeth, gums, and the opposite lip) moving along a related but distinct path.

For each of these sequences, the user manually initialized the contour in the first frame and deformable contour tracking was performed with both traditional and RCS-based external energy models. As expected, the edge-based dynamic contour gave poor matches when confronted with a cluttered background with changing contrast. The RCS dynamic contours were more successful, and remained attached to the lips throughout the sequence. Fig. 4 shows these sequences, and the computed segmentation mask from the RCS result.

2. To help us disambiguate the contrast edges, of course, we could use better models of object shape and/or object motion. Much recent work has improved the shape-dynamics models for dynamic contours and related techniques [13], [9], [32]. We could also track multiple hypotheses [23] and use future shape distortions to select the correct tracking sequences. These approaches require detailed shape and motion models for each object that we hope to track; they can be applied independently with our method when appropriate.

5 Conclusion

Radial Cumulative Similarity (RCS) is a new image transform that describes local image homogeneity, comprised of a central attribute value and a function of the surrounding radial structure. We compute radial similarity with the cumulative error of an attribute value relative to the value at the center of an image window. This representation can be insensitive to structure outside an occluding boundary, yet model the boundary itself. When used for correspondence search, it can track foreground surfaces near occlusions where there is no foreground contrast other than from the occlusion boundary. Using the local cumulative similarity function as a weighting measure, anisotropic smoothing reduces noise in the displacement estimate but does not introduce outlier contamination. The RCS transform can also serve as the basis for image measurements in a dynamic contour tracking system. An external energy term can be defined using the RCS transform evaluated at nodal points of the contour, as well as along rays perpendicular to the contour spine. By compositing the neighborhood function in the RCS representation, an accurate continuous alpha channel corresponding to the foreground/background mask function can be computed. We believe the RCS transform can be integrated into many different window-based correspondence algorithms, and yield improved performance when tracking low-contrast occluding contours or features.

References

[1] A. Amini, T. Weymouth, and R. Jain, "Using Dynamic Programming for Solving Variational Problems in Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 9, pp. 855-867, Sept. 1990.

[2] S. Ayer and H. Sawhney, "Layered Representation of Motion Video Using Robust Maximum Likelihood Estimation of Mixture Models and MDL Encoding," *Proc. Int'l Conf. Computer Vision,* 1995.

[3] J.R. Bergen, P.J. Burt, K. Hanna, R. Hingorani, P. Jeanne, and S. Peleg, "Dynamic Multiple-Motion Computation," *Artificial Intelligence and Computer Vision,* Y.A. Feldman and A. Bruckstein, eds., Elsevier Science Publishers B.V., 1991.

[4] D. Bhar and S. Nayar, "Ordinal Measures for Visual Correspondence," *Proc. Conf. Computer Vision and Pattern Recognition,* 1994.

[5] M. Black and P. Anandan, "A Framework for Robust Estimation of Optical Flow," *Proc. Int'l Conf. Computer Vision,* pp. 263-274, 1993.

[6] M. Black and Y. Yacoob, "Tracking and Recognizing Rigid and Non-Rigid Facial Motions Using Local Parametric Models of Image Motions," *Proc. Int'l Conf. Computer Vision,* 1995.

[7] M. Black, Y. Yacoob, A. Jepson, and D. Fleet, "Learning Parameterized Models of Image Motion," *Proc. Conf. Computer Vision and Pattern Recognition,* 1997.

[8] M. Black, Y. Yacoob, and D. Fleet, "Modeling Appearance Change in Image Sequences," *Proc. Third Int'l Workshop Visual Form,* 1997.

[9] A. Blake and M. Isard, "Three-Dimensional Position and Shape Input Using Video Tracking of Hands and Lips," *Proc. SIGGRAPH '94,* pp. 185-192, 1994.

[10] C. Bregler and M. Slaney, "Snakes," IRC-TR 1995-017, Interval Research Corp., Palo Alto, Calif., 1995. see http://web.interval.com/papers/1995-017/.

[11] Y. Boykov, O. Veksler, and R. Zabih, "A Variable Window Approach to Early Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 12, pp. 1283-1294, Dec. 1998.

[12] C. Bregler, M. Covell, and M. Slaney, "Video Rewrite," *Proc. SIGGRAPH,* 1997.

[13] T. Cootes, C. Taylor, A. Lanitis, D. Cooper, and D. Graham, "Building and Using Flexible Models Incorporating Grey-Level Information," *Proc. Int'l Conf. Computer Vision,* pp. 242-246, 1993.

[14] M. Covell and T. Darrell, "Dynamic Occluding Contours: A New External Energy Term for Snakes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1999.

[15] M. Covell and M. Withgott, "Automatic Morphing: Spanning the Gap between Motion Estimation and Morphing," *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing,* 1994.

[16] M. Covell and C. Bregler, "Eigenpoints," *Proc. Int'l Conf. Image Processing,* 1996.

[17] T. Darrell and A. Pentland, "Robust Estimation of a Multi-Layer Motion Representation," *Proc. IEEE Workshop Visual Motion,* 1991.

[18] T. Darrell and E. Simoncelli, "Nulling Filters and the Separation of Transparent Motions," *Proc. Computer Vision and Pattern Recognition,* 1993.

[19] B. Galvin, B. McCane, and K. Novins, "Virtual Snakes for Occlusion Analysis," *Proc Computer Vision and Pattern Recognition,* pp. 294-299, 1999.

[20] G. Hager and P. Belhumeur, "Real-Time Tracking of Image Regions with Changes in Geometry and Illumination," *Proc. Computer Vision and Pattern Recognition,* 1996.

[21] D. Huttenlocher, J. Noh, and W. Rucklidge, "Tracking Non-Rigid Objects in Complex Scenes," *Proc. Int'l Conf. Computer Vision,* pp. 93-101, 1993.

[22] M. Irani, B. Rousso, and S. Peleg, "Computing Occluding and Transparent Motions," *Int'l J. Computer Vision,* vol. 12, no. 1, 1994.

[23] M. Isard and A. Blake, "Visual Tracking by Stochastic Propagation of Conditional Density," *Proc. European Conf. Computer Vision,* pp. 343-356, 1996.

[24] D. Jones and J. Malik, "A Computational Framework for Determining Stereo Correspondences from a Set of Linear Spatial Filters," *Proc. Second European Conf. Computer Vision,* pp. 395-410, 1992.

[25] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka, "A Video-Rate Stereo Machine and Its New Applications," *Proc. Computer Vision and Pattern Recognition,* 1996.

[26] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiments," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 9, pp. 920-932, Sept. 1994.

[27] M. Kass, A. Witkin, and D. Terzopolous, "Snakes: Active Contour Models," *Proc. Int'l Conf. Computer Vision,* pp. 259-268, 1987.

[28] P. Litwinowicz and L. Williams, "Animating Images with Drawings," *Proc. SIGGRAPH,* pp. 409-412, 1994.

[29] N. Peterfreund, "The Velocity Snake," *Proc. IEEE Non-Rigid and Articulated Motion Workshop,* pp. 70-79, 1997.

[30] M. Shizawa and K. Mase, "Simultaneous Multiple Optical Flow Estimation," *Proc. Computer Vision and Pattern Recognition,* 1990.

[31] E.P. Simoncelli and E.H. Adelson, "Probability Distributions of Optical Flow," *Proc. Computer Vision and Pattern Recognition,* 1991.

[32] D. Terzopolous and R. Szeliski, "Tracking with Kalman Snakes," *Active Vision,* A. Blake, A. Yuille, eds., pp. 3-20, Cambridge, Mass.: MIT Press, 1992.

[33] J. Wang and E. H. Adelson, "Layered Representations for Image Sequence Coding," *Proc. Conf. Computer Vision and Pattern Recognition,* 1993.

[34] L. Wiskott, J. Fellous, N Kruger, C. von der Mals-burg, "Face Recognition and Gender Determination," *Proc. Int'l Workshop Automatic Face and Gesture Recognition,* pp. 92-97, 1995.

[35] W. Yu, K. Daniilidis, S. Beauchemin, and G. Sommer, "Detection and Characaterization of Multiple-Motion Points," *Proc. Computer Vision and Pattern Recognition,* pp. 171-177, 1999.

[36] A.L. Yuille, D.S. Cohen, and P.W. Hallinan, "Feature Extraction from Faces Using Deformable Templates," *Proc. Computer Vision and Pattern Recognition,* pp. 104-109, 1989.

[37] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence," *Proc. Third European Conf. Computer Vision,* 1994.