Motion Segmentation Using Convergence Properties*

Moshe Ben-Ezra and **Shmuel Peleg** and **Benny Rousso**
Institute of Computer Science
The Hebrew University of Jerusalem
91904 Jerusalem, Israel

Abstract

Motion segmentation is traditionally coupled with motion detection, where each image region corresponds to a particular motion model which accounts for the temporal changes in the region. Using the motion model to estimate the second frame from the first frame, for example, should give a very low prediction error in the corresponding region.

To relax the need for accurate motion models, it is proposed to examine the *convergence* of the prediction error, rather than the prediction error itself. In an iterative process of motion computation followed by computing the prediction error, those points for which the prediction error is being reduced are considered as a coherent region. This segmentation approach works well even with approximate motion models that don't eliminate the prediction error.

1 Introduction

Motion segmentation is traditionally coupled with motion detection, where each region corresponds to a particular motion model which explains the temporal changes in that image region [Boult and Brown, 1991; Wang and Adelson, 1993; Irani *et al.*, 1994]. Using the motion model to estimate the second frame from the first frame, for example, should give a very low prediction error in the corresponding region. Under this approach motion segmentation corresponds to selecting a threshold for the prediction error. Regions having below threshold prediction error are considered to belong to the motion model being used.

To remove the need for accurate motion models, and to reduce the dependency on the arbitrarily selected threshold for the prediction error, it is proposed to examine the *convergence* of the prediction error, rather than the prediction error itself. In an iterative process of region-based motion computation [Bergen *et al.*, 1992] followed by computing the prediction error, those points for which the prediction error is being reduced are considered as a coherent region. The proposed approach works well even with approximate motion models that don't eliminate the prediction error.

*This research was sponsored by ARPA through the U.S. Office of Naval Research under grant N00014-93-1-1202, R&T Project Code 4424341—01.

2 Region-Based Motion Analysis

Region motion is computed from spatio-temporal image derivatives [Lucas and Kanade, 1981; Bergen *et al.*, 1992; Irani *et al.*, 1994]. Given two successive frames from an image sequence, $I_1(x,y)$ and $I_2(x,y)$, partial derivatives are denoted as follows:

$$I_x = \frac{\partial I_1}{\partial x}, \quad I_y = \frac{\partial I_1}{\partial y}, \quad I_t = \frac{\partial I}{\partial t} = I_2 - I_1 \quad . \qquad (1)$$

Using the well-known constraint equation [Horn and Schunck, 1981], the 2D motion (u,v) which minimizes the prediction error for the region of analysis $\Re$ [Irani *et al.*, 1992] should also minimize the following error function:

$$\mathcal{E}(u,v) = \sum_{(x,y)\in\Re} (uI_x + vI_y + I_t)^2. \qquad (2)$$

The error minimization is performed over the parameters of one of the following motion models [Irani *et al.*, 1994]:

1. **Translation:** 2 parameters, $u(x,y) = a, \quad v(x,y) = d$. In this model, the entire image is assumed to have a uniform translation.
2. **Affine**: 6 parameters, $u(x,y) = a + bx + cy$, $v(x,y) = d + ex + fy$. [Bergen *et al.*, 1992].
3. **A Moving planar surface** (a pseudo projective transformation): 8 parameters [Adiv, 1985; Bergen *et al.*, 1992] $u(x,y) = a + bx + cy + gx^2 + hxy$, $v(x,y) = d + ex + fy + gxy + hy^2$.

The computation framework is based on multiresolution and iterations, using a Gaussian pyramid, as described in [Bergen and Adelson, 1987; Irani *et al.*, 1994].

One of the major properties of this framework is that it finds the motion parameters of a *single* image region [Burt *et al.*, 1991], even when the image contain several different motions. This region will be called the *Dominant Region* having the *Dominant Motion*. The property of finding a single motion of a single image region appears especially in the case of the translation model, and enables a preliminary image segmentation (Sect. 3) to locate the pixels which belong to that object.

3 Motion Segmentation

When the parameters of a motion model are initially computed for a region that includes several different motions, the resulting parameters are influenced by all motions, and do not correspond to a single motion. However, rarely are all motions perfectly balanced, and there is one region that affects the motion computation more than other regions.

The proposed segmentation method looks for exactly these *more influential*, or *dominant*, pixels. When the two frames are registered using the computed motion parameters, the dominant pixels will be those pixels whose prediction errors are reduced by the registration.

This definition of *dominant* pixels avoids the need for arbitrary thresholds of the prediction error itself, and it gives good results even when inaccurate simple motion models are being used (Figs. 2,3). The preliminary segmentation using a simple motion model is being used for another motion computation process, this time focusing only on the dominant region. This process is repeated while upgrading the motion model to a more accurate one, and in our experiments the most elaborate motion model used was the image motion of a 3D moving planar surface with its 8 motion parameters.

3.1 Changes in the Prediction Error

Given two frames I_1 and I_2, let I_2^w be the frame I_2 warped back towards frame I_1 using the motion parameters computed in the registration step. I_t and I_t^w will be the temporal derivatives *before* and *after* the registration process :

$$I_t = I_2 - I_1 \quad , \quad I_t^w = I_2^w - I_1 \quad . \tag{3}$$

The *Improvement Measure* M, measuring for each pixel (x,y) the improvement in the prediction error after applying the registration, is defines as:

$$M(x,y) \;=\; \frac{|I_t| \;-\; |I_t^w|}{|I_t| \;+\; |I_t^w|} \; . \tag{4}$$

When $|I_t| = |I_t^w| = 0$, no motion has been detected, and M is set to 1. This way, regions which seem stationary both before and after the warping are considered to be part of the dominant object.

The measure M has the following properties:

1. $-1 \leq M \leq 1$.
2. $M = 0$ when the registration has no effect on the prediction error.
3. When $|I_t| \gg |I_t^w|$ then $M \rightarrow 1$, indicating that the prediction error *decreases* by the registration process, and the pixel therefore belongs to the dominant object.
4. When $|I_t| \ll |I_t^w|$ then $M \rightarrow -1$, indicating that the prediction error *increases* by the registration process, and the pixel therefore does not belong to the dominant object.

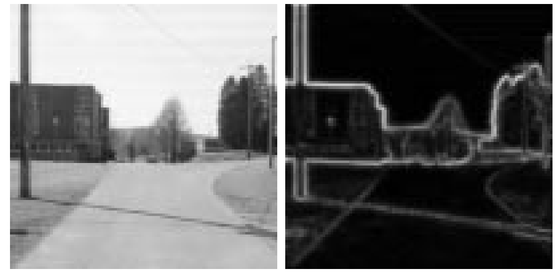

Figure 1: The Reliability Measure of an image.

Note that since $\frac{|I_t|}{\sqrt{I_x^2+I_y^2}}$ is actually the magnitude of the *Normal Flow*, the measure M (Eq. 4) can also be interpreted as the improvement in the *Normal Flow*.

3.2 Membership in the Dominant Region

A point will be considered as part of the dominant region if it has an improvement in the prediction error ($M(x,y) > 0$), subject to the reliability of the motion information at this point. The *Reliability Measure* R at pixel (x,y) is defined as

$$R(x,y) = \sqrt{I_x(x,y)^2 + I_y(x,y)^2}. \tag{5}$$

Figure 1 displays this measure for a given image.

The membership function in the dominant region is computed using images in several resolution levels, $0 < l \leq N-1$. For each resolution level l both measures $R_l(x,y)$ (Eq. 5) and $M_l(x,y)$ (Eq. 4) are computed. The membership function in the dominant region is defined as the weighted sum over all resolution levels:

$$S(x,y) \;=\; \frac{1}{\sum_{l=0}^{N-1} R_l(x,y)} \sum_{l=0}^{N-1} R_l(x,y) * M_l(x,y). \tag{6}$$

The measure S, whose values are always in the range $-1 \leq S \leq 1$, indicates the membership in the dominant region, where $S = 1$ indicates absolute membership, and $S = -1$ is absolute non-membership. It is interesting to observe uniform regions with no gradients. In this case $S = 0$ since $M = 1$ (Eq. 4) and $R = 0$ (Eq. 5), and the uniform regions will not be considered as part of the dominant region.

3.3 Steps of Segmentation Algorithm

1. Given two frames $I_1(x,y)$ and $I_2(x,y)$, the global *translation* of the entire image is calculated. This step converges to the *dominant translation* even without segmentation [Burt *et al.*, 1991].
2. Frame I_2 is warped towards frame I_1 using the motion parameters that were computed in the previous step.
3. The *Segmentation Measure* $S(x,y)$ is computed (Eq. 6).
4. The above process is repeated for higher order motion models, where the *translation* motion model is followed by the *affine* model, which is followed by the *moving plane* model. For all motion models,

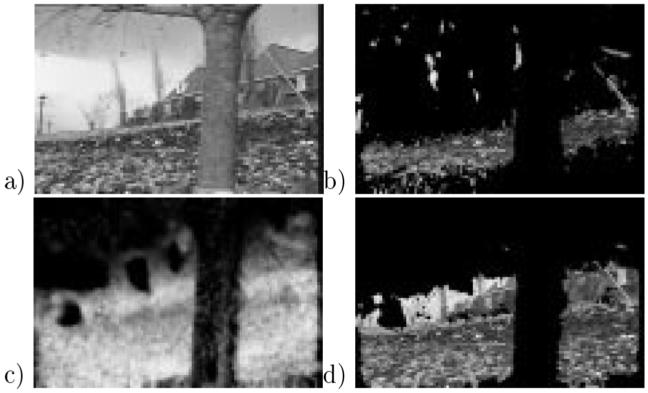

Figure 2: Motion segmentation - first sequence.
a) An original frame.
b) Binary Segmentation mask using a translation motion model. Black is excluded from the dominant region.
c) The fuzzy segmentation mask S using a moving-plane motion model.
d) Binary segmentation mask using a moving-plane motion model.

except the translation model, motion computation was done when points were weighted by the segmentation mask S, so that only points in the dominant region are influencing the computation.

4 Examples

Results of applying the proposed motion segmentation method on two publicly available image sequences are displayed in Figures 2-3. In addition to an original frame, binary segmentation masks are displayed when the *translation* motion model was used, and when the *moving plane* motion model was used. The fuzzy segmentation mask S is displayed for the last case of the *moving plane* motion model. Note that the binary mask was created for visualization only, and is not used by any part of the algorithm. Only the fuzzy segmentation mask S is used.

5 Concluding Remarks

The motion segmentation approach described in this paper has two contributions. First, the segmentation can be performed even with an approximate motion model. In addition, identifying regions with an improved prediction error does not depend on arbitrary thresholds as do existing methods, which examine whether the prediction error falls below the threshold. We find this approach very helpful in the analysis of multiple image motions.

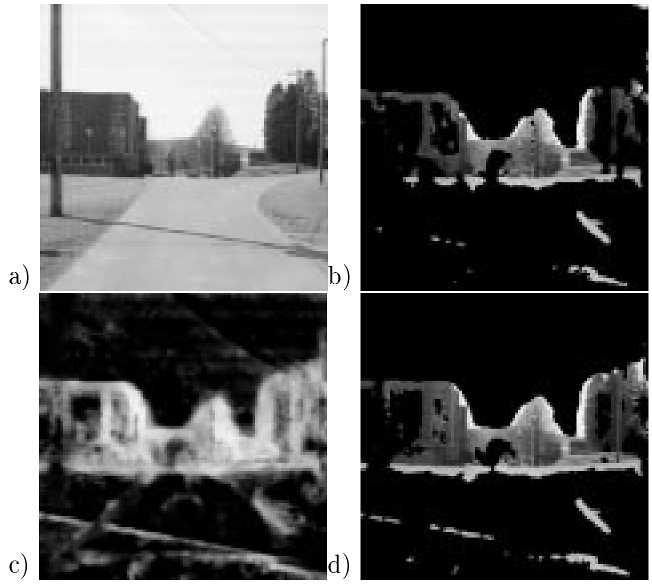

Figure 3: Motion segmentation - second sequence.
a) An original frame.
b) Binary Segmentation mask using a translation motion model. Black is excluded from the dominant region.
c) The fuzzy segmentation mask S using a moving-plane motion model.
d) Binary segmentation mask using a moving-plane motion model. Note that the walking person was excluded from the dominant region.

References

[Adiv, 1985] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(4):384–401, July 1985.

[Bergen and Adelson, 1987] J.R. Bergen and E.H. Adelson. Hierarchical, computationally efficient motion estimation algorithm. *J. Opt. Soc. Am. A.*, 4:35, 1987.

[Bergen *et al.*, 1992] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conf. on Computer Vision*, pages 237–252, 1992.

[Boult and Brown, 1991] T.E. Boult and L. Gottesfeld Brown. Factorization-based segmentation of motions. In *IEEE Workshop on Visual Motion*, pages 179–186, Princeton, 1991.

[Burt *et al.*, 1991] P.J. Burt, R. Hingorani, and R.J. Kolczynski. Mechanisms for isolating component patterns in the sequential analysis of multiple motion. In *IEEE Workshop on Visual Motion*, pages 187–193, 1991.

[Horn and Schunck, 1981] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[Irani *et al.*, 1992] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *European Conf. on Computer Vision*, pages 282–287, 1992.

[Irani *et al.*, 1994] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *Int. J. of Computer Vision*, 12(1):5–16, January 1994.

[Lucas and Kanade, 1981] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*, pages 121–130, 1981.

[Wang and Adelson, 1993] J. Wang and E. Adelson. Layered representation for motion analysis. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 361–366, New York, June 1993.