SPATIO-TEMPORAL FILTERS FOR TRANSPARENT MOTION SEGMENTATION

Hongche Liu[†‡]*, Tsai-Hong Hong*[‡]*, Martin Herman*[‡]*, and Rama Chellappa*[†]

[‡]NIST, Intelligent Systems Division, Bldg 220, Rm B124, Gaithersburg, MD 20899 USA
[†]Center for Automation Research, University of Maryland, College Park, MD 20742 USA

ABSTRACT

An image is ideally a projection of the 3-D scene. However, the imaging process is always imperfect and constrained by the physical environment. This paper is concerned with image sequences acquired in such situations, the so-called transparency, for example, viewing through a window with reflections. When such situation occurs, the image sequence contains undesirable transparent motion, for example, of the window reflections. This complicates the already difficult motion estimation problem. We present an algorithm to segment transparent motion based on a spatio-temporal filtering technique—3-D Hermite polynomial differentiation filters. With motion segmentation accomplished, we can then focus on the scene analysis. The implementation of our algorithm is fast and accurate.

1. INTRODUCTION

Transparent motion refers to two superimposed intensity patterns with different motions present in a single image. The resulting image can be hypothesized to be either the addition or multiplication of the two moving patterns depending on the type of transparency. Specular and diaphanous transparencies are additive; film and shadow transparencies are multiplicative [1] .

Fig 1.1-Fig 1.3 give an example of specular transparent motion. Since the moon is rotating, the surface is translating (to the right in Fig 1.3). The moon also revolves around the earth causing the specular reflection of the sun to translate (to the upper left) at the same time.

An image I with transparent motion is formulated as follows:

$$I(x, y, t) = I_1(x, y, t) \oplus I_2(x, y, t) \quad , \tag{1}$$

where $\oplus$ can be addition or multiplication, and

$$I_i(x, y, t) = I_i(x - u_i t, y - v_i t, 0) \quad , i = 1, 2 \, , \tag{2}$$

using the translational motion model.

It can easily be seen that any single motion estimation algorithm would fail to compute either one of the motions unless there is one dominant texture. Transparent motion segmentation needs to be performed before an analysis of individual motion is possible. Although it may seem that the transparent motion scenario rarely occurs, it is actually very common in the real world. The specular reflection depicted in Fig 1.2 is a good example. Also common is the dirty lens scenario where the dirt pattern on the lens is motionless but contaminates the intensity pattern of the scene.

Fig 1.1 Diffuse reflection

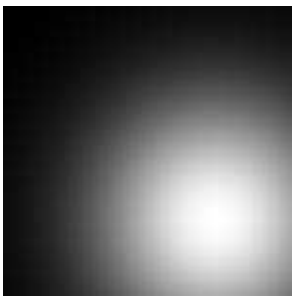

Fig 1.2 Specular reflection

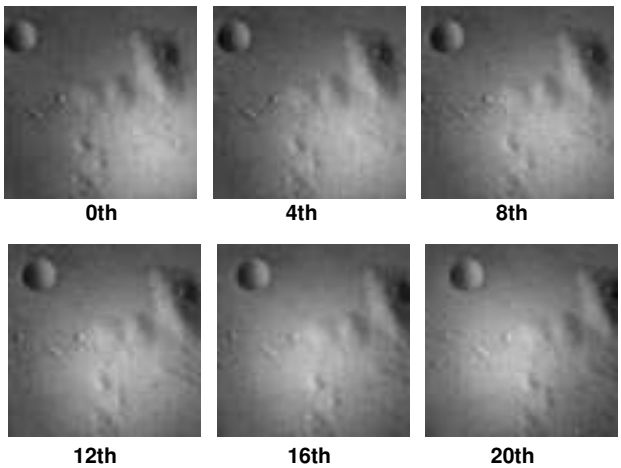

Fig 1.3 Sequential display of image frames

Transparent motion segmentation is difficult to solve. However, it is encouraging to note that when humans view images with transparent motion, they have no problem segmenting different motions.

There are three steps involved in transparent motion segmentation:

1) Determining whether multiple motions exist.

2) Extracting transparent motions.

3) Segmenting transparent motions.

Unlike most previous studies which are primarily focused on 2) and sometimes on 1), this paper address all three steps.

2. PREVIOUS WORK

In this paper, we expand the multiple motion model proposed by Shizawa and Maze[2] [3] . Their basic idea is to apply consecutive linear operators, which are equivalent to single motion constraint equations. For two motions, the model is formulated as:

$$[(u_1, v_1, 1) \bullet \nabla][(u_2, v_2, 1) \bullet \nabla]I = 0 \quad (3)$$

where $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial t}\right)$ and $\bullet$ is the inner product.

With this model, Langley, et al. [4] used a phase-based method to extract multiple motion.

Our expansion of the model is based on the gradient constancy assumption, which is true when brightness constancy is true. It is formulated as

$$[(u_1, v_1, 1) \bullet \nabla^i][(u_2, v_2, 1) \bullet \nabla^j]I = 0 \quad (4)$$

where $\nabla^i = \left(\frac{\partial^i}{\partial x^i}, \frac{\partial^i}{\partial y^i}, \frac{\partial^i}{\partial t^i}\right)$.

With different combinations of i and j, we get multiple motion constraint equations to solve the problem.

We adopt the conventional gradient-based method instead of using frequency information because our previous work [5] [6] has shown that spatio-temporal filtering based on 3-D Hermite polynomial differentiation filters is excellent in estimating image gradients.

The orthogonality and Gaussian properties of the filters insure numerical stability up to very high order. Numerous physiological models support the theory that visual receptive fields are modeled by Gaussian derivative masks of various widths. Also, the separability of the filters in three dimensions offers a tremendous speed advantage over the frequency domain filters of Langley, et al. and Shizawa & Mase.

Bergen et al. [7] proposed an iterative scheme and image registration to minimize their multiple motions estimation error. Darrell and Pentland [8] used robust estimation techniques and a layered representation of the images to segment multiple flows. Although these algorithms are better at handling occluding boundaries, their iterative nature does not guarantee convergence to the global minimum and their computational cost is very high.

3. THE ALGORITHM

To facilitate the accurate estimation of image gradients in the framework of the expanded transparent motion model depicted in (4), an accurate and stable image differentiation filtering scheme is needed. The set of orthogonal 3-D Hermite polynomial filters is excellent for this task.

3.1 Hermite polynomials

The nth Hermite polynomial $H_n(x)$ is a solution of

$$\frac{d^2 H_n}{dx^2} - 2x\frac{dH_n}{dx} + 2nH_n = 0. \quad (5)$$

The $H_n(x)$ are derived by Rodrigues' formula [9]

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2} . \quad (6)$$

By substituting $G(x)$ (with variance σ^2) for e^{-x^2} in (6), we generalize to Hermite polynomials with respect to the Gaussian function. Let these Hermite polynomials be denoted by $\bar{H}_n(x)$. Then

$$\bar{H}_n(x) = \left(\frac{1}{2^{1/2}\sigma}\right)^n H_n\left(\frac{x}{2^{1/2}\sigma}\right) . \quad (7)$$

The scalar product of two functions and the L_2-norm of a function with *G(x)* as a weight function are defined as

$$\langle a, b\rangle \equiv \int_{-\infty}^{\infty} G(x)a(x)b(x)dx \quad \text{and } \|a\| \equiv \langle a, a\rangle^{1/2}. \quad (8)$$

The orthogonality of $\{\bar{H}_n(x)\}$ can be expressed in the following way[9] :

$$\langle \bar{H}_m, \bar{H}_n\rangle = \sigma^{-2n} n!\,\delta_{mn}. \quad (9)$$

The 3-D case of Hermite polynomials is especially simple because they are separable:

$$\bar{H}_{ijk}(x, y, t) = \bar{H}_i(x) \cdot \bar{H}_j(y) \cdot \bar{H}_k(t) \quad (10)$$

The estimate of image gradients using 3-D Hermite polynomials is $I_{ijk} = \langle I, \bar{H}_{ijk}\rangle$, where (i, j, k) is the order of differentiation in the (x, y, t) direction, respectively. The computation of I_{ijk} amounts to three 1-D convolutions due to the filters' separability, so the implementation is efficient. Later in this section, we use a more conventional notation for image gradients. For example, I_{211} is expressed as I_{xxyt}.

3.2 Determining Multiple Motions

First, to determine whether multiple motions exist, we adapt the idea proposed by Yamamoto [10] . We first try to model the image sequence with a single motion model (refer to [6] [5] for details) and use the linear squared error (residual) as an indicator of multiple motion:

$$\|Ax_0 - b\| > k_1\|Ax_0\| \text{ and } \|Ax_0 - b\| > k_2\|A\|, \quad (11)$$

where A is the matrix composed of spatial gradients of multiple orders computed by Hermite polynomial differentia-

tion filters and b is the corresponding temporal gradients.

Once it is determined that multiple motion exists at a point, we extract and segment the multiple motions as follows.

3.3 Extracting Transparent Motions

If the transparency is multiplicative, we simply take the logarithm of the image intensity and follow the same procedure as with additive transparencies.

From equation (4), we derive enough equations to solve for the unknowns (x in (13)). The unknowns are nonlinear functions of (u_1, v_1) and (u_2, v_2). But we solve the linear system for the unknowns first and use the relations (x in (13)) to solve for the flow. For double transparent motion, we have

$$min\|Ax - b\| \text{ where} \tag{12}$$

$$A = \begin{bmatrix} I_{xx} & I_{xy} & I_{yy} & I_{xt} & I_{yt} \\ I_{xxx} & I_{xxy} & I_{xyy} & I_{xxt} & I_{xyt} \\ I_{xxy} & I_{xyy} & I_{yyy} & I_{xyt} & I_{yyt} \\ I_{xxxx} & I_{xxxy} & I_{xxyy} & I_{xxxt} & I_{xxyt} \\ I_{xxxy} & I_{xxyy} & I_{xyyy} & I_{xxyt} & I_{xyyt} \\ I_{xxyy} & I_{xyyy} & I_{yyyy} & I_{xyyt} & I_{yyyt} \end{bmatrix}, \quad b = \begin{bmatrix} I_{tt} \\ I_{xtt} \\ I_{ytt} \\ I_{xxtt} \\ I_{xytt} \\ I_{yytt} \end{bmatrix} \quad \text{and}$$

$$x = \begin{bmatrix} u_1 u_2 \\ u_1 v_2 + u_2 v_1 \\ v_1 v_2 \\ u_1 + u_2 \\ v_1 + v_2 \end{bmatrix}. \tag{13}$$

Once the value of x is computed from the least square system, (u_1, v_1) and (u_2, v_2) can be easily computed and properly aligned according to [2] .

3.4 Segmenting Transparent Motions

Now that every image point has two flow vectors, we need to separate them into two coherent flow fields. For this, we assume that the foreground flow is constant. Thus we compute a 2-D image flow histogram and find a peak in the histogram, which corresponds to the constant motion. Then at every point, the flow closer to this peak is assigned to the foreground motion, while the other flow is assigned to the background motion. This assumption is often true when the foreground motion is induced by window reflections or a dirt patterns on the lens.

4. EXPERIMENTS

We have tested our algorithm on a synthesized image sequence (Fig. 1) and on a real image sequence (Fig. 4). The 3-D window size used in both instances is 25x25x17.

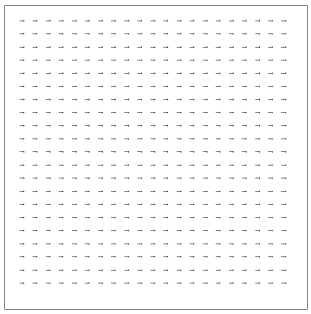

Fig 2.1 True moon surface flow

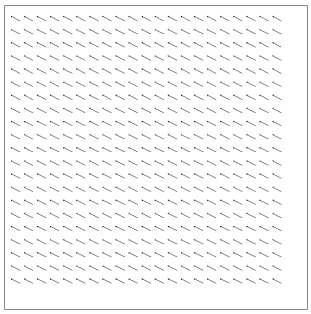

Fig 2.2 True specular flow

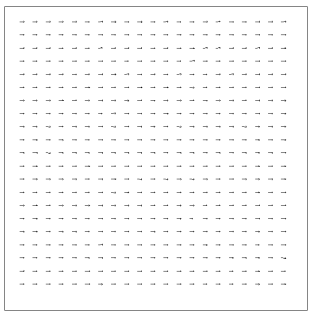

Fig 2.3 Computed moon surface flow

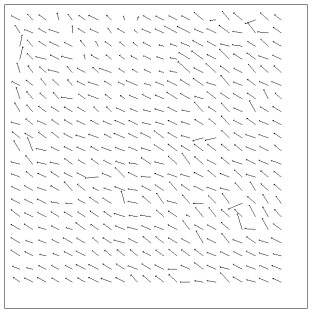

Fig 2.4 Computed specular flow

In Fig. 2, we show the true and computed flow for the moon surface and the specular reflection. As can be seen, the computed background motion corresponding to the moon surface is very accurate. The computed specular motion is less accurate because the specular component has significantly less texture. This is the general aperture problem present in motion algorithms. In the extreme case where one moving pattern has no texture, only the other motion can be detected.

We then separate the two intensity patterns by the following technique. In Fig 3.1, we warp the first frame with flows from the flow field associated with specular reflection motion and then subtract the second frame, resulting in the cancellation of specular reflection pattern. The output is the temporal difference of the moon surface pattern. Fig 3.2 is derived conversely.

Fig 3.1 Temporal difference of the moon surface pattern

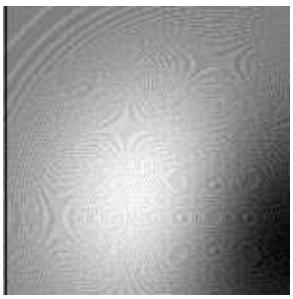

Fig 3.2 Temporal difference of the specular reflection pattern

The real image sequence used in our experiments is shown in Fig 4. In this sequence, the scene is composed of

a picture in a frame which reflects the ceiling light. Note that the picture is moving toward the upper right corner and the reflection is moving upward.

From the results in Fig 5.1, we see that the motion of the picture is extracted accurately in most areas. However, the motion of the light reflection is noisy because of the lack of texture in this pattern. Nonetheless, a close inspection of the flow field in Fig 5.2 reveals that it is indeed moving upward. It should be noted that the object motion, rather than the reflection motion, is usually the desirable information to extract and the accuracy associated with the former is more important

1st 10th 19th

Fig 4.Sequential display of real images

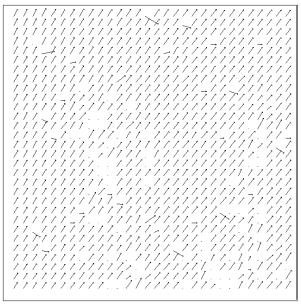

Fig 5.1 Computed picture flow

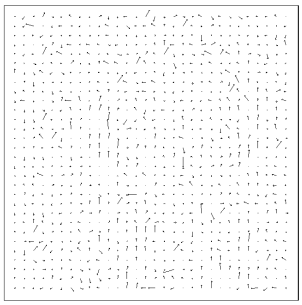

Fig 5.2 Computed reflection flow

Our transparent motion segmentation algorithm was implemented on a Sun Hyper Sparc 10 Themis board. For the window size specified above, it ran on a 150x150 image sequence in less than 1 minute. The code is available through our ftp site at giskard.cme.nist.gov, in directory ftp/pub/motion/transparent.

5. CONCLUSION

Transparent motion segmentation has received relatively little attention due to its numerical cost. The same reason appears to have led recent approaches away from gradient-based methods. In this paper, however, we show a gradient-based method for transparent motion segmentation which is facilitated by a spatio-temporal filtering technique based on 3-D Hermite polynomials. We achieve a fast and accurate algorithm. This work offers a comprehensive solution to transparent motion segmentation by addressing all three issues involved in transparent motion segmentation: determining the existence of multiple motions, extracting transparent motions and segmenting transparent motions. In the future, a more sophisticated assumptions about the motion will be developed to adapt the algorithm to the general case.

REFERENCES

[1] Kersten, D. "Transparency and the Cooperative Computation of Scene Attributes", in Computational Models of Visual Processing, ed. by Landy, M. and Movshon, J.A., pp. 209-228, MIT Press, 1991.

[2] Shizawa, M. and Mase, K., "Principles of Superposition: A Common Computational Framework for Analysis of Multiple Motion", Proceedings of the IEEE Workshop on Visual Motion, pp.289-295, 1991.

[3] Shizawa, M. and Mase, K., "A Unified Computational Theory for Motion Transparency and Motion Boundaries Based on Eigenenergy Analysis", Proceedings of the IEEE Workshop on Visual Motion, pp.164-172, 1991.

[4] Langley, K., Fleet, D. and Atherton, T., "Multiple Motion from Instantaneous Frequency", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.846-849, 1992.

[5] Liu, H., Hong, T., Herman, M. and Chellappa, R., "A Generalized Motion Model for Estimating Optical Flow Using 3-D Hermite Polynomials", Proceedings of the IEEE International Conference on Pattern Recognition, pp.360-366, 1994.

[6] Liu, H., Hong, T., Herman, M. and Chellappa, R., "A General Motion Model and Spatio-temporal Filters for Computing Optical Flow", NIST-IR 5539, January, 1995, to appear in the International Journal of Computer Vision.

[7] Bergen, J.R., Burt, P.J., Hingorani, R. and Peleg, S., "A Three-Frame Algorithm for Estimating Two-Component Image Motion", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 9, pp. 886-896, 1992.

[8] Darrel, T. and Pentland, A., "Robust Estimation of a Multi-Layered Motion Representation", Proceedings of the IEEE Workshop on Visual Motion, pp.173-178, 1991.

[9] Hashimoto, M. and Sklansky, J. "Multiple-Order Derivatives for Detecting Local Image Characteristics", Computer Vision, Graphics, and Image Processing, vol. 39, pp. 28-55, 1987.

[10] Yamamoto, M., "A Segmentation Method Based on Motion from Image Sequence and Depth", Proceedings of the IEEE International Conference on Pattern Recognition, pp. 230-232, 1990.