Shape from Dynamic Texture for Planes

Yaser Sheikh
SEECS,
University of Central Florida
FL 32816, USA
yaser@cs.ucf.edu

Niels Haering
ObjectVideo
Reston, VA 20191
USA
niels@objectvideo.com

Mubarak Shah
SEECS
University of Central Florida,
FL 32816, USA
shah@cs.ucf.edu

Abstract

We propose a method for recovering the affine geometry of a dynamically textured plane from a video sequence taken by an uncalibrated, fixed, perspective camera. Some instances of approximately planar surfaces that are coated with a dynamic texture include large water bodies, (such as lakes and oceans), heavy traffic, dense crowds, escalators, and foliage in the wind. Under the assumption of translational dynamic textures, we propose a direct algorithm for the estimation of the inter-frame elation that does not require explicit identification of texels or movetons. In addition, we develop a general algorithm for recovering the affine geometry of homogeneous dynamic textures by identifying a constraint on the expected values of motion magnitudes. We report experimental results on several real videos of dynamic texture found in the world.

1. Introduction

Inferring the shape of surfaces and objects from visual cues such as motion, contours, shading and texture has a rich tradition in computer vision literature. While texture and motion have each been examined individually as sources of information for shape, their composite - dynamic texture - has not been investigated for recovery of shape information. In this paper we demonstrate that for planar scenes, the dynamic texture of a surface can be exploited to recover its affine geometry. Planes coated with dynamic textures often arise in the world, in seascapes (such as beaches, ports, lake-sides), dense crowds, highway traffic, foliage in the wind, escalators, and so on. Since parallelism, ratio of areas, ratios of lengths on collinear lines and linear combinations of vectors (such as centroids) are all preserved under affine transformations ([11]), recovering the affine geometry has useful application in surveillance, particularly for seaports and highways.

Inquiry into the use of texture as a cue for shape recovery began with J. J. Gibson's seminal book in 1950, [10]. Shape from texture approaches have traditionally assumed some structure on the textured surface, such as regularity, periodicity, parallelism, homogeneity or isotropy. Different approaches make one or more of these assumptions. Isotropy, for instance, was used by Witkin ([22]) and Brown ([4]), homogeneity by Kanatani and Chou in [13] and Criminisi and Zisserman in [6] and periodicity by Ribeiro and Hancock in [18]. An experimental evaluation of homogeneity and isotropy has been reported by Rosenholtz and Malik in [19]. Homogeneity, in particular, assumes that the texture statistics do not depend on the position of a pixel, only on its neighborhood - a formal definition has been provided by Kanatani and Chou in [13]. Criminisi and Zisserman showed in [6] that by making an assumption of homogeneity rectification up to an affine transformation can be recovered. Other approaches to determining the vanishing line (such as [16], [14] and [3]) involve identifying moving elements and/or tracking them across the plane explicitly. Dynamic or temporal textures were first investigated by Nelson and Polana in [17]. Several statistical models of dynamic textures have since been proposed in literature, such as the the spatio-temporal autoregressive (STAR) model of Szummer and Picard, [20], the multi-resolution scheme of Bar-Joseph in [1] and the AR model of Doretto *et al* in [7]. Motion estimation in scenes containing dynamic textures have also been addressed using stochastic models in [9] and [21]. In our work, we propose the use of dynamic texture for recovery of the affine geometry of a plane, without the need to identify texels, movetons or any other basic element of the dynamic texture. We investigate two models, translational and homogeneous dynamic textures and develop algorithms for both. Our experiments on both real and synthetic data demonstrate accurate estimation on a variety of data.

The rest of the paper is organized as follows. Section 2 introduces our model of a scene and the geometric result required by this work. Section 3 introduces the problem in the context of homogeneous dynamic textures and a special sub-class called translational dynamic textures. We

propose a direct algorithm for estimating the affine geometry for the translational case from two frames, as well as an algorithm for estimation over time for homogeneous dynamic textures. In Section 4 we report results of quantitative experimentation on rendered dynamic textures as well as qualitative demonstration on many real videos. Finally, we conclude with a discussion of our work in Section 5 and muse on future directions of research.

2. Scene Model

The scene is modeled as a plane coated with a dynamic texture which is being viewed by a stationary and uncalibrated perspective camera. Criminisi and Zisserman have shown in [6] that if the vanishing line can be identified in the image then the scene plane can be rectified up to an affine transform using the following transformation

$$\mathbf{H}_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix}, \tag{1}$$

where $\mathbf{l} = (l_1, l_2, l_3)^\intercal$ is the vanishing line. Gibson has observed that "the texture gradient of the ground is orthogonal to the horizon on the retinal image", and using this observation, Criminisi and Zisserman proposed an algorithm to affine rectify a plane using texture information. In our paper, we make a statement, analogous to Gibson's, concerning dynamic texture and propose an algorithm that utilizes the dynamics of the texture to affine rectify the scene plane.

2.1. Dynamic Texture Segmentation

We outline a simple approach to segment out dynamic textures from static sections of an observed scene. Initial segmentation is through a patch-based mechanism that measures autocorrelation. Static areas yield high autocorrelation values across time whereas regions containing dynamic texture yield much lower autocorrelation values. This segmentation is then expanded to include external neighboring pixels whose color statistics fit the statistics of the internal pixels. Several more sophisticated methods exist such as [5], but since we are addressing a simpler case of differentiating dynamic textures from static regions, this approach usually suffices.

3. Recovering the Affine Geometry of Dynamic Textures

In this section we describe two algorithms to recover the vanishing line from the dynamics of textures. The first algorithm can be used when the dynamic texture is *translational*, that is when the flow field can be globally approximated as a translation in the real world. The second algorithm deals more generally with *homogeneous* dynamic textures as defined by Doretto *et al* in [8]. A homogeneous dynamic texture is any texture whose spatiotemporal statistics are homogeneous - a direct analogue to homogeneous texture whose spatial statistics are homogeneous.

3.1. Translational Dynamic Textures

For a special class of dynamic textures, where the motion field of the dynamic texture can be globally modeled to be translational, the vanishing line can be estimated *directly* from image gradients. Examples of dynamic texture that can reasonably be modeled in this way includes escalators, parades, one-way highway traffic, and most importantly videos of seas, oceans and other large bodies of water. The proposed algorithm requires just two frames for accurate computation, although the parameters can be estimated over time across many frames.

3.1.1 Direct Estimation of the Vanishing Line

The transformation induced by the image of a translational flow field along a plane is a conjugate translation and can be expressed as an elation, [11]. An elation is a transformation with an axis (a line of fixed points) and a vertex (a pencil of fixed lines intersecting at that point). All invariant points under an elation lie on its axis[1]. Specifically, the images of any two points on a world plane undergoing translation along the plane, $\mathbf{x}', \mathbf{x} \in \mathbb{P}^2$ are related by an elation, $\mathbf{H}_E$,

$$\lambda \mathbf{x}' = \mathbf{H}_E \mathbf{x} \tag{2}$$

where $\mathbf{H}$ can be parameterized as,

$$\mathbf{H}_E = \mathbf{I} + \mu \mathbf{v}\mathbf{a}^\intercal, \quad \mathbf{a}^\intercal \mathbf{v} = 0 \tag{3}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{a}$ is the axis, $\mathbf{v}$ is the dynamic texture vertex and μ is a constant dependant to the magnitude of motion along the plane. An illustrative example of these constructs is given in Figure 3, where the axis is the horizon and the vertex a point where the motion appears to be emanating from. Equation 3 can be rewritten in non-homogeneous form, giving the pair of equations,

$$x' = \frac{(1 + \mu v_1 a_1)x + (\mu v_1 a_2)y + \mu v_1}{\mu a_1 x + \mu a_2 y + 1 + \mu} \tag{4}$$

$$y' = \frac{(\mu v_2 a_1)x + (1 + \mu v_2 a_2)y + \mu v_2}{\mu a_1 x + \mu a_2 y + 1 + \mu}. \tag{5}$$

The so-called *direct* paradigm for motion estimation, proposed by Horn and Weldon in [12], advocates the estimation of motion parameters directly from image gradients. By making an assumption of brightness constancy over small motion, the well-known optical flow constraint equation is

[1] For further details and for properties of elations see Appendix 7 in [11]

obtained and is often used to estimate the parameters. The optical flow constraint is,

$$I_x u + I_y v + I_t = 0 \tag{6}$$

where $u = x' - x, v = y' - y$ is the motion flow. Equation 6 can be rewritten as,

$$I_x(x' - x) + I_y(y' - y) + I_t = 0 \tag{7}$$

and substituting from Equations 4 and 5 we wish to minimize a function of five variables,

$$\min f_1(a_1, a_2, v_1, v_2, \mu) = 0, \tag{8}$$

but with four degrees of freedom since it is subject to the constraint

$$a_1 v_1 + a_2 v_2 + a_3 v_3 = 0. \tag{9}$$

Thus, given an initial estimate of the μ, $\mathbf{a}$ and $\mathbf{v}$, nonlinear minimization can be performed using the Levenberg-Marquardt algorithm to find an optimal estimate of the vanishing line. Of course, all the conventional strategies that are employed in direct algorithms, such as hierarchical estimation, robust error measures, gradient smoothing, can be employed during estimation to obtain accurate results. It should be noted that for videos containing water bodies, in particular, estimation of the elation using feature based algorithms perform poorly compared to this direct algorithm since it is difficult to locate salient features that do not distort over time. On the other hand, there is clearly a 'global' transformation, that direct algorithms capture accurately. In some cases, (such as the escalator sequence of the results) feature based approaches can be used to estimate the inter-frame elation. But in general, and for dynamic textures that display stochastic behavior (such as water bodies) in particular, direct algorithms are more suited for estimating this type of motion.

3.1.2 Initialization

An initial estimate is obtained by approximating the elation as an affine transform, $\mathbf{H_A}$ and computing it using the linear algorithm proposed by Bergen *et al* in [2]. By noting that the vanishing line is that line that does not move under the transformation, $\mathbf{H_A}$, i.e. that $\lambda\mathbf{a} = \mathbf{H_A}^{-\intercal}\mathbf{a}$, and recalling the definition of eigenvectors, an initial estimate of $\mathbf{a}$ can be obtained from the eigenvector of $\mathbf{H_A}^{-\intercal}$ corresponding to the eigenvalue at the greatest distance from unity[2]. Similarly, an eigenvector of $\mathbf{H_A}$ can be used as an estimate of the vertex $\mathbf{v}$. An initial estimate of μ can be obtained by performing a line search that minimizes the sum of squared difference. These estimates can then be used in a nonlinear constrained minimization to estimate the elation that minimizes the sum of squared difference between the target image and the source image warped with the elation,

$$\arg\min \sum_i (\mathbf{I}_t - w(\mathbf{I}_s|\mathbf{a}, \mathbf{v}, \mu))^2. \tag{10}$$

[2] A related result has been obtained using frequency analysis by Ribeiro and Hancock in [18].

Objective

Find the affine rectification matrix $\mathbf{H}_\alpha$ given two consecutive images $\mathbf{I}_s$ and $\mathbf{I}_t$ of the plane coated with a translational dynamic texture.

Algorithm

1. Estimate an affine transformation $\mathbf{H_A}$ between $\mathbf{I}_s$ and $\mathbf{I}_t$ using the direct algorithm of [2].
2. Compute the initial estimate $\hat{\mathbf{a}}$ by taking the eigenvector associated with the eigenvalue of $\mathbf{H_A}^{-\intercal}$ that is at the greatest distance from unity.
3. Compute the initial estimate $\hat{\mathbf{v}}$ by taking the only eigenvector $\mathbf{H_A}$ that is not a point at infinity.
4. Compute the initial estimate $\hat{\mu}$ by performing a line search that minimizes the sum of squared difference between the target and warped source images.
5. Use $(\hat{\mathbf{a}}, \hat{\mathbf{v}}, \hat{\mu})$ as initial estimates for minimization of the SSD between source and target image using the Levenberg-Marquardt algorithm.
6. Compute $\mathbf{H}_\alpha$ from $\mathbf{a}$ (using matrix in Equation 1).

Figure 1. Algorithm for translational dynamic textures

3.2. Homogeneous Dynamic Texture

For homogeneous dynamic textures in general, the motion need not necessarily follow any global parametrization. In [8], Doretto *et al* define a homogeneous dynamic texture as an instance of a weak-sense stationary space-time process $f(\mathbf{x}, t)$. If a process is weak-sense stationarity, the first moment function $\mu(\mathbf{x}, t)$ is independent of $\mathbf{x}$ and t and the second moment function $\gamma(\mathbf{x}, \mathbf{x} + g, t, t + h)$ is independent of $\mathbf{x}$ and t for any h and g. Instead, we define dynamics textures as a strict-sense stationarity process, that $\{\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n\}$ and $\{\mathbf{X}_{1+h}, \mathbf{X}_{2+h}, \ldots \mathbf{X}_{n+h}\}$ have the same joint distribution for all h and $n > 0$. For homogeneous dynamic textures, under the strict stationarity condition, we have,

Proposition 3.1 *The expected value of motion magnitude at any location on a homogeneous dynamic texture is identical.*

Proof. For strict-sense stationarity, the joint distribution $p(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n)$ is the same as $p(\mathbf{X}_{1+h}, \mathbf{X}_{2+h}, \ldots \mathbf{X}_{n+h})$ for all integers h and

$n > 0$. Therefore, for any function $g(\cdot)$, the distribution of $p(g(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n))$ will be identical to $p(g(\mathbf{X}_{1+h}, \mathbf{X}_{2+h}, \ldots \mathbf{X}_{n+h}))$. Since the brightness constancy constraint equation is a function only of the intensities of pixels, i.e. it uses spatial and temporal gradients and does not use location for computation, the distribution of flow and therefore the distribution of flow magnitude $\rho(\cdot)$ (a function, in turn, of flow) will also be stationary. Therefore, the expected value of motion magnitude $E[p(\rho(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n))] = E[p(\rho(\mathbf{X}_{1+h}, \mathbf{X}_{2+h}, \ldots \mathbf{X}_{n+h}))]$ for all integers $n > 0$ and h. It should be noted that since flow is estimated locally, it is sufficient to require only that the joint distribution of local neighborhoods remain the same at any location.

Then, analogous to Gibson's statement on texture gradients, we have

Proposition 3.2 *In perspective images of a plane with a homogeneous dynamic texture, the gradient of imaged motion magnitude is perpendicular to the vanishing line in the image coordinate.*

Proof. From Proposition 3.1, the expected value of motion magnitude is equal all over the 3D world plane, and thus when projected onto an image plane it captures the perspective effects of image projection. Criminisi and Zisserman provide a geometric proof that the gradient of any perspective effect is perpendicular to the vanishing line in [6].

3.2.1 Estimation of the Vanishing Line

We model the motion magnitude at a pixel as a univariate Gaussian distribution, and sequentially estimate the expected value (mean) of motion magnitudes per pixel over time, i.e. that the magnitude of motion at a pixel is distributed $\kappa(i,j) \sim N(\mu(i,j), \sigma(i,j))$. At the incidence of each frame, pixel-wise optical flow is computed and used to update the current estimates of μ and σ. To recover the vanishing line from the field of motion magnitudes, we need to fit a line to the projections of points onto the $x-y$ plane (image coordinates) along the motion magnitude gradient direction. This is equivalent to solving the following linear system of equations,

$$\mathbf{Ya} = \mathbf{s} \tag{11}$$

where

$$\mathbf{Y} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ & \vdots & \\ x_n & y_n & 1 \end{bmatrix}, \mathbf{s} = \begin{bmatrix} \kappa(x_1, y_1) \\ \kappa(x_2, y_2) \\ \vdots \\ \kappa(x_n, y_n) \end{bmatrix} \tag{12}$$

Objective

Find the affine rectification matrix $\mathbf{H}_\alpha$ given a sequence of k images $\{\mathbf{I}_1, \ldots \mathbf{I}_k\}$ of a plane coated with a homogeneous dynamic texture.

Algorithm

1. For each frame,
   - Estimate the motion magnitude $\bar{\kappa}(i,j)$ at each pixel location using an optical flow algorithm (e.g. Lucas-Kanade approach, [15]).
   - Update motion magnitude mean estimation at each pixel location, $\kappa(i,j)$ using a sequential update.
2. Estimate the location of the vanishing line $\mathbf{a}$ by solving the linear system of equations of in 11.
3. Compute $\mathbf{H}_\alpha$ from $\mathbf{a}$ (using matrix in Equation 1).

Figure 2. Algorithm for homogeneous dynamic textures

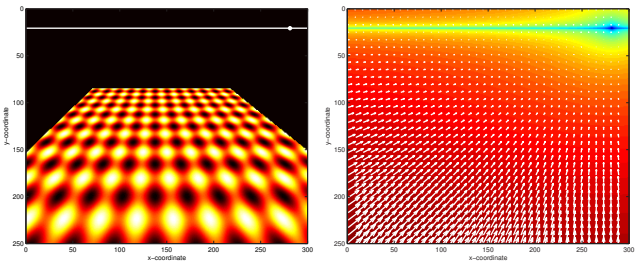
and $\mathbf{a} \in \mathbb{P}^2$ is the vanishing line in homogeneous coordinates. An intuitive interpretation of this result is as follows. A plane $\mathbf{p} = [p_x \ p_y \ 1 \ p_c]$, where (x,y) are the image spatial indices, can be fit to the data points, $[\kappa(x,y) \ x \ y] \in \mathbb{R}^3$, and the intersection of this plane with the $x-y$ plane gives us the vanishing line. Since the equation of the $x-y$ plane is $\kappa = 0$, simple substitution shows that the equation of the intersection line (also the vanishing line) is $\mathbf{a} = [p_x \ p_y \ p_c]$.

4. Experimentation

We have performed experiments both qualitatively using real-world videos as well as quantitatively with rendered data to test sensitivity to noise. For homogeneous textures we found that around 2 to 3 minutes of video was usually required to obtain stable results. All results for translational dynamic textures were performed using only a pair of frames. The performance of the proposed approaches is tested on real videos including, surveillance videos of ports, heavy traffic, escalators, dense crowds and a treadmill.

4.1. Qualitative Testing

The algorithms were extensively tested on was port surveillance video, with a fixed perspective camera observing the port. Figure 4 shows the results on a number of such sequences - all results were obtained between a pair of frames. The dashed line represents the estimate yielded by the affine transformation, which was used as an initialization to the computation of the elation. The solid line represents the final estimate of the elation. In each sequence the affine transform gives us results in the correct vicinity, and the elation latches exactly onto the correct horizon. In Figure 5 we show two images of a boat passing along a scene where we had already estimated the vanishing line. We manually



(a) (b)

Figure 3. Translational dynamic texture as an elation. (a) The horizon, shown as a white line, is the axis of the elation, and the white point from which the motion appears to emanate is the vertex. Both quantities were estimated using the proposed approach. (b) Magnitude of the motion field. The magnitude of motion is zero (minimum) at the horizon.

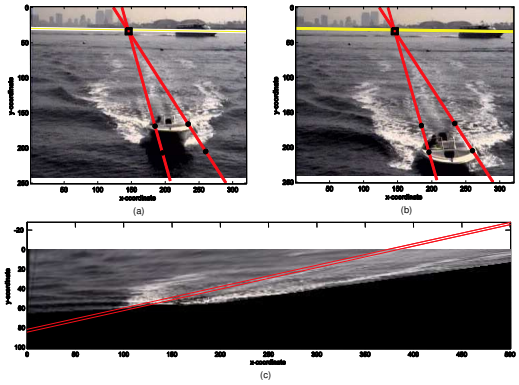


Figure 5. Affine Rectification. (a,b) Lines through corresponding points selected (manually) on the boat over two frames converge on the estimated vanishing line. (c) After rectification the lines are parallel (meet at the line at infinity).

identified two points on the boat across the images, and verified that the lines passing through the correspondences did indeed intersect on the vanishing line. In addition we rectified the images using the affine rectification matrix and verified that the transformed lines were parallel. Figure 6 shows the result of the direct algorithm on an escalator sequence. Since the camera is already almost fronto-parallel, the vanishing line is far from the image. Finally, a treadmill image is rectified using the direct method in Figure 7.

Next, we tested the algorithm for homogeneous textures. Figure 8 shows a plot of the mean motion magnitude for an IR sequence looking across a lake. The motion magnitudes were learnt over 3000 frames, and it is clear through inspection that the points lie approximately on a plane. A video of the learning process is included in the supplementary material. The white line marks the vanishing line estimated from the motion magnitudes. Although this approach is not as exact as the direct algorithm for translational motions, the example validates our model. The constraint of strict spatial homogeneity can be relaxed as long as the motion is approximately uniform throughout the plane. Figure 9 shows a much shorter sequence of 117 frames with a dense crowd of people. Despite the relatively short learning period, the estimate of the vanishing line is reasonable. Figures 10 shows the result on a highway sequence sequence of 1000 frames. The estimate in this case is not as accurate as could be expected since the cars do not move at similar speeds and the scene is comparatively sparse. Despite this the affine rectification of the sequence, shown alongside the images, is still satisfactory. Once the affine geometry has been recovered, higher-level analysis of objects and their behaviors is more meaningful since parallelism, ratio of areas, ratios of lengths on collinear lines and linear combinations of vectors (such as centroids) are all affine invariants.

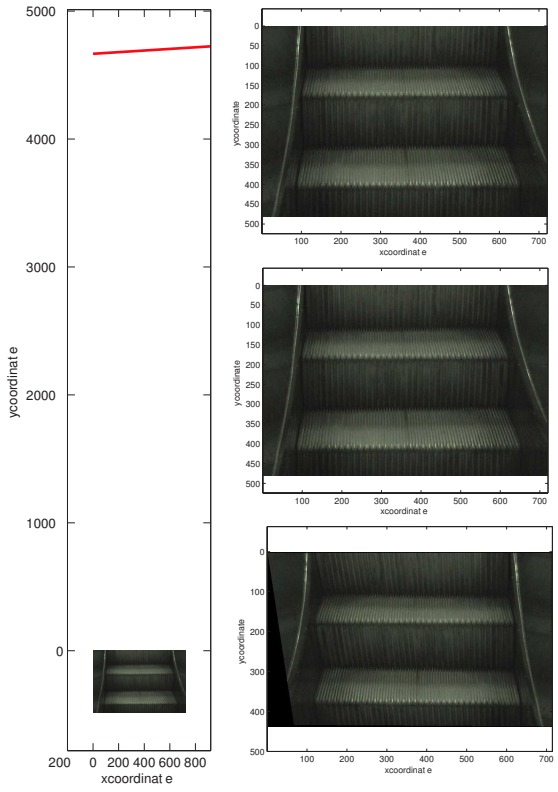


Figure 6. Escalator Sequence. The line in the left image is the vanishing line of the plane, the top right and middle right images are the two images used for computing the line and the bottom right is the rectified image.

4.2. Quantitative Testing

The algorithm for translational dynamic textures was tested quantitatively using the rendered data. A periodic texture was synthetically generated and translated along a plane in 3D space. A perspective camera was defined to

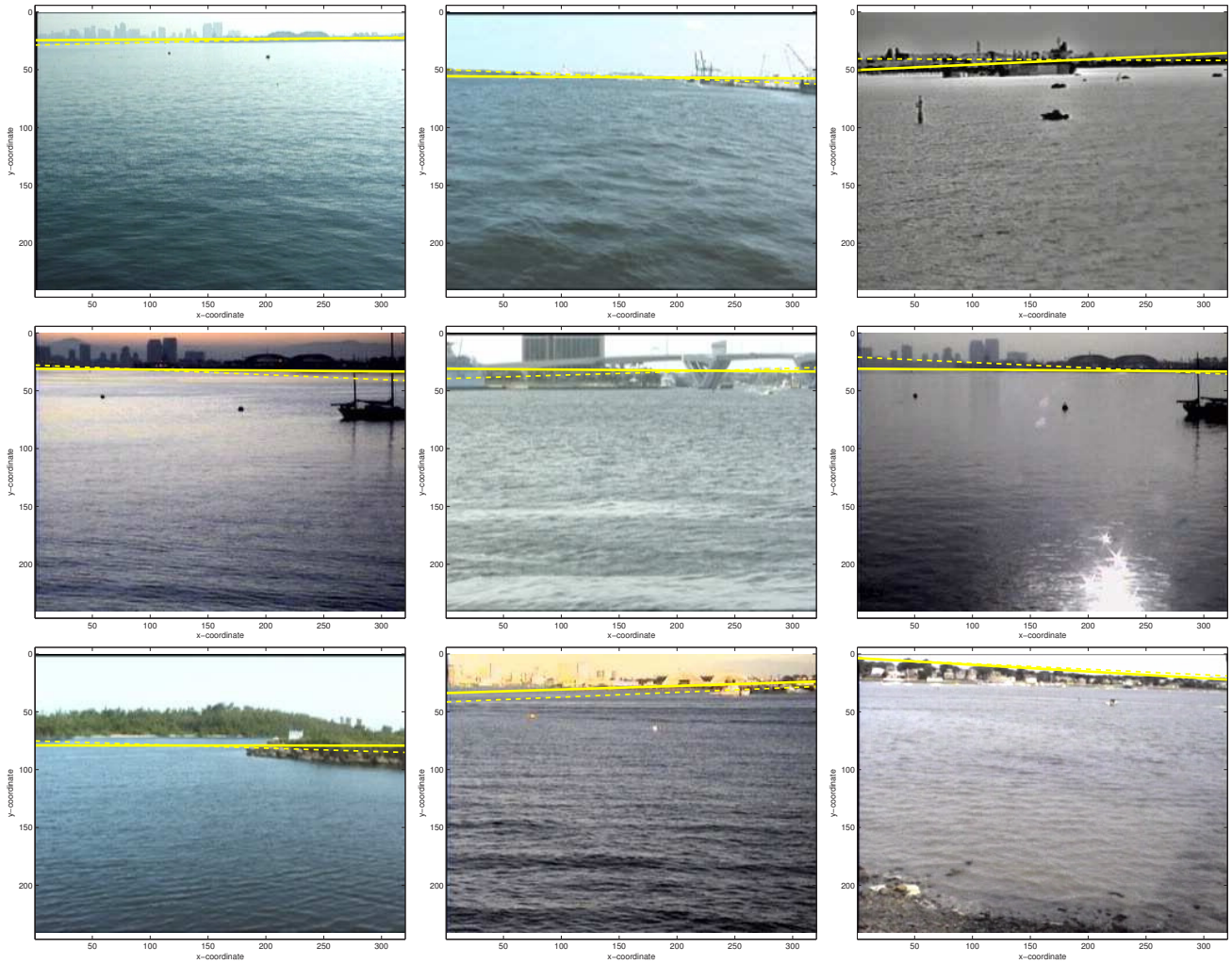


Figure 4. Recovering the vanishing lines for translational dynamic textures. The dashed line is the estimate of the horizon provided by the affine approximation, and the solid line is the refinement using an elation.

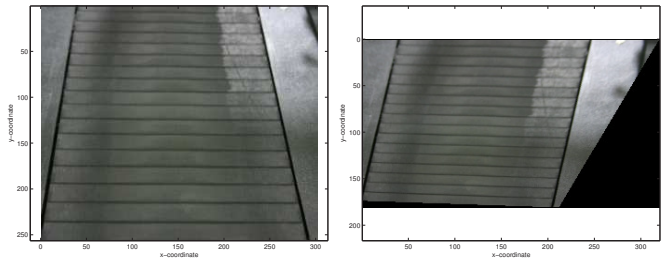
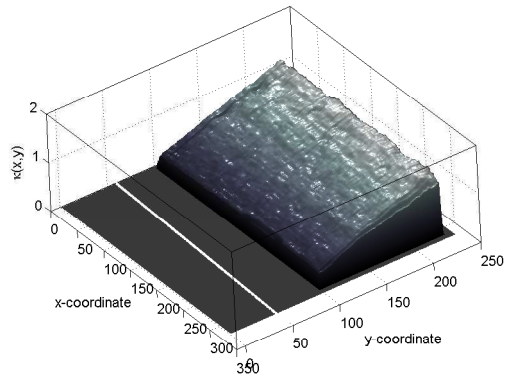


Figure 7. Treadmill images. (a) The source image. (b) Rectified source image.

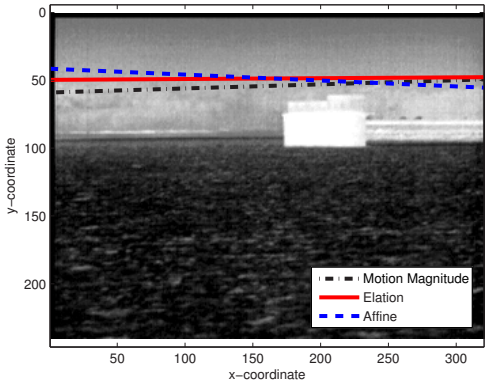
view the plane, in terms of a homography, $\mathbf{H}$, producing a conjugate translation. Using the direct algorithm, we recovered an estimate of the vanishing line and vertex. Figure 11 shows how the accuracy of the estimate is affected as noise strength is increased in the images. The noise was increased from 0% to 21% of the maximum signal strength, at unit increments, and the accuracy of the estimate of vanishing line was estimated by the norm of the difference between the ground truth and the estimated line. At each noise strength, 20 runs were executed (different noise) and the line-scatter plot is also shown. Figure 11 (a) shows the error norm of the estimated vanishing line to the true vanishing line, and Figure 11 (b) shows the error norm of the estimated vertex to the true vertex. The solid plots show the mean error norms for both tests, and the square markers denote each run at that noise strength.

5. Conclusion and Future Work

We propose the use of dynamic texture as a cue for recovering the affine geometry of a plane. We discuss two types of dynamic textures, translational and homogeneous, and propose algorithms for each case. For the translational case a direct algorithm is presented that relates the vanish-

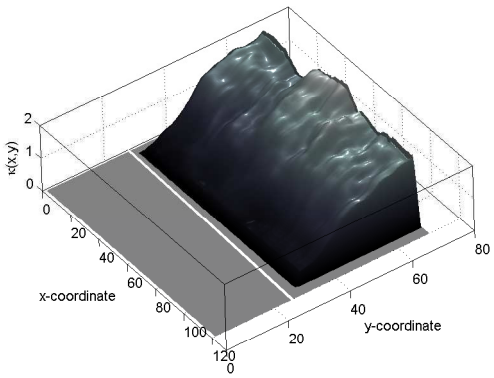


(a)

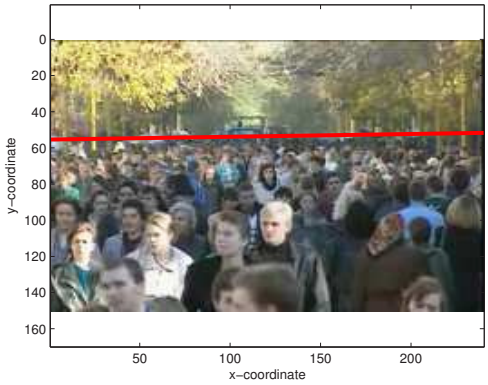


(b)

Figure 8. IR Lake Sequence. (a) The estimated mean of the motion magnitudes after 3000 frames. (b) Estimates of the vanishing line from the affine transform, the elation, and the motion magnitude.



(a)



(b)

Figure 9. Crowd Sequence. (a) The estimated mean of the motion magnitudes after 117 frames. (b) Estimates of the vanishing line.

ing line directly to the image gradients, and for the homogeneous case relates the vanishing line directly to the expected values of motion magnitude. We test our algorithms on a variety of data, including port surveillance videos, highways, dense crowds, and escalators. There are many future directions that can be pursued. First, methods can be developed to investigate shape from dynamic texture for curved surfaces, relating texture dynamics at a pixel location to the surface normal at that point. Second, in this paper we have utilized the motion characteristics of dynamic textures and an interesting direction may be to investigate using both spatial texture and temporal behavior concurrently. Finally, the use of isotropy to recover a metric upgrade seems promising.

Acknowledgments

This research was partially supported by Homeland Security Advanced Research Project Agency Award N00014-05-C-0112.

References

[1] Z. Bar-Joseph, R. El-Yaniv, D. Lischinski and M. Werman, "Texture Mixing and Texture Movie Synthesis using Statistical Learning", *IEEE Transactions on Vision and Computer Graphics*, 2001.

[2] J. Bergen, P. Anandan, K. Hanna, R. Hingorani, "Hierarchical Model-Based Motion Estimation," *European Conference on Computer Vision*, 1992.

[3] B. Bose and E. Grimson, "Ground Plane Rectification by Tracking moving Objects," *IEEE International Workshop on Visual Surveillance and PETS*, 2003.

[4] L. Brown and H. Shvaytser, "Surface Orientation from Projective Foreshortening of Isotropic Texture Autocorrelation," *IEEE Transactions on Patten Analysis and Machine Intelligence*, 1990.

[5] D. Cremers, G. Doretto, P. Favaro, and S. Soatto, "Dynamic Texture Segmentation", *IEEE International Conference on Computer Vision*, 2003.

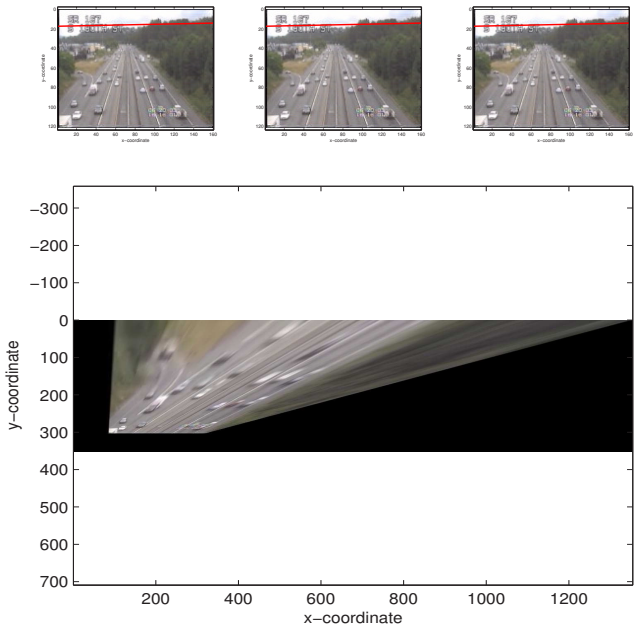
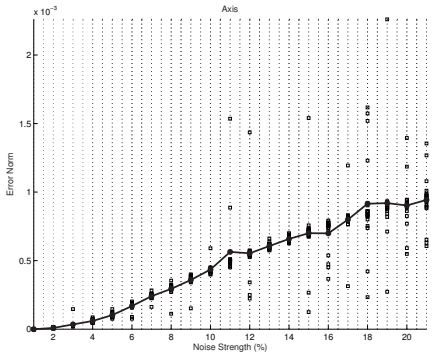
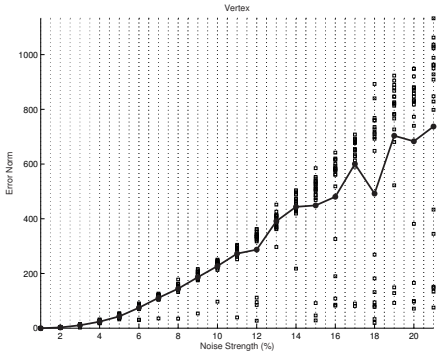


Figure 10. Highway Sequence. Top: Frames of the highway sequence. The vanishing line after 1000 frames is marked. Bottom: Image after rectification.



(a)



(b)

Figure 11. Tolerance to noise.(a) The error norms for the estimates of the axis w.r.t the ground truth, (b) The error norms for the estimates of the vertex w.r.t. the ground truth.

[6] A. Criminisi and A. Zisserman, "Shape from texture: homogeneity revisited", *British Machine Vision Conference*, 2000.

[7] G. Doretto, A. Chiuso, Y. Wu, S. Soatto,"Dynamic Textures", *International Journal on Computer Vision*, 2003.

[8] G. Doretto, E. Jones, S. Soatto,"Spatially Homogeneous Dynamic Textures", *European Conference on Computer Vision*, 2004.

[9] A. Fitzgibbon, "Stochastic Rigidity: Image Registration for Nowhere-static Scenes," *IEEE International Conference on Computer Vision*, 2001.

[10] J. Gibson, "The Perception of the Visual World", Houghton Mifflin, Boston, 1950.

[11] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.

[12] B. Horn and E. Weldon Jr., "Direct Methods for Recovering Motion," *International Journal of Computer Vision*, 1988.

[13] K. Kanatani and T-C Chou, "Shape from Texture: General Principle", *Artificial Intelligence*, 1989.

[14] N. Krahnstoever and P. Mendonça, "Bayesian Autocalibration for Surveillance," *IEEE International Conference on Computer Vision*, 2005.

[15] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *IJCAI*, 1981.

[16] F. Lv, T. Zhao, and R. Nevatia, "Self-Calibration of a Camera from Video of a Walking Human," *IAPR International Conference on Pattern Recognition*, 2001.

[17] R. Nelson and R. Polana, "Qualitative Recognition of Motion using Temporal Texture", *CVGIP Image Understanding*, 1992.

[18] E. Ribiero and E. Hancock, "Shape form Periodic Texture Using Eigenvectors of Local Affine Distortion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.

[19] R. Rosenholtz and J. Malik, "Surface Orientation from texture: Isotropy or Homogeneity (or Both?)", *Vision Research*, 1997.

[20] M. Szummer and R. Picard, "Temporal texture Modeling", *IEEE International Conference on Image Processing*, 1996.

[21] R. Vidal and A. Ravichandaran, "Optical Flow Estimation and Segmentation of Multiple Moving Dynamic Textures," *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[22] A. Witkin, "Recovering Surface Shape and Orientation from texture", *Artificial Intelligence*, 1981.