Shape Background Modeling : The Shape of Things That Came

Nathan Jacobs and Robert Pless
Department of Computer Science and Engineering
Washington University, St. Louis, MO, 63117

Abstract

Detecting, isolating, and tracking moving objects in an outdoor scene is a fundamental problem of visual surveillance. A key component of most approaches to this problem is the construction of a background model of intensity values. We propose extending background modeling to include learning a model of the expected shape of foreground objects. This paper describes our approach to shape description, shape space density estimation, and unsupervised model training. A key contribution is a description of properties of the joint distribution of object shape and image location. We show object segmentation and anomalous shape detection results on video captured from road intersections. Our results demonstrate the usefulness of building scene-specific and spatially-localized shape background models.

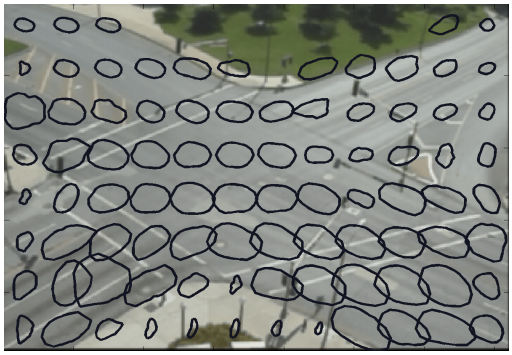


Figure 1: A collection of spatially-localized mean shapes—the mean of shapes with centroids within the surrounding 30 by 30 pixel region—from a video of a vehicular intersection. In this work the dependence of shape statistics on image location is used to improve object segmentation.

1. Introduction

Detecting, isolating and tracking objects as they move through a scene is a common requirement of security and surveillance systems. When security systems are installed for very long terms use (days, months or years), it is feasible to initially spend substantial computing resources to facilitate later processing. While effort has been spent to develop background models to support object and anomaly detection [9, 3, 10], and to characterize motions in a scene [6, 5, 7], little attention has been paid to the fact that other properties are consistent in a scene that is viewed by a static camera.

One such property is the shapes of objects that appear. For example, in a surveillance video of an intersection a pixel location may see either the background road or a car. But all cars at this location are moving the same direction and in the same relative orientation, so the space of shapes seen at this pixel may be quite limited. How can these shape spaces be acquired, represented, and used?

This paper presents a process that includes offline methods for learning and characterizing the distributions of shapes that appear in a scene, and faster methods to use these shape models to support accurate segmentation in challenging conditions. We attempt to characterize the benefits of scene-specific shape models for object localization and anomaly detection. Furthermore, we find that location specific shape models (*i.e.* that give distribution of shapes whose centroid is near an image location) are both easy to compute and provide extremely strong shape constraints.

This is preliminary work on shape background modeling. Therefore we have kept our methods simple to avoid obscuring the core message that a shape background model is useful without the need for complicated methods. In future work we will address more sophisticated methods of building and using shape background models.

1.1. Previous Work

There are many approaches to building pixel-level models of typical, background, properties of a given video. Most approaches focus on building models of densely sampled data such as pixel intensity [9, 3, 10] or motion properties [6, 7].

Anomalous object detection is one goal of surveillance, and most research has focussed on building background models that support real-time classification of atypical scene behavior, based on, for example, statistics of pixel intensity [9], spatio-temporal intensity derivatives [7], or global scene appearance [12]. More power is available in modeling the statistics of foreground objects detected in the

scene. Mostly these models concentrate on object motions. A mixture model is used to define a PDF over the position, velocity, size, and aspect ratio of many blobs tracked over time to support classifying scene activities [9], and similar measurements have been used to derive an HMM over global scene states [1].

The use of statistical shape priors has been shown to improve the ability of object segmentation algorithms to handle ambiguous image data [4, 2] and overlapping objects [11]. Our segmentation approach incorporates energy functionals inspired by these works but takes a simpler approach to energy minization. Previous work has focused on sophisticated methods of energy minimization while ignoring the dependence of shape and location.

The use of spatially-uniform shape priors learned from motion segmentation to improve the accuracy of static object segmentation has been explored by Ross and Kaelbling [8]. The shape prior is combined with image data and a segmentation is determined by performing energy minization on a Markov random field.

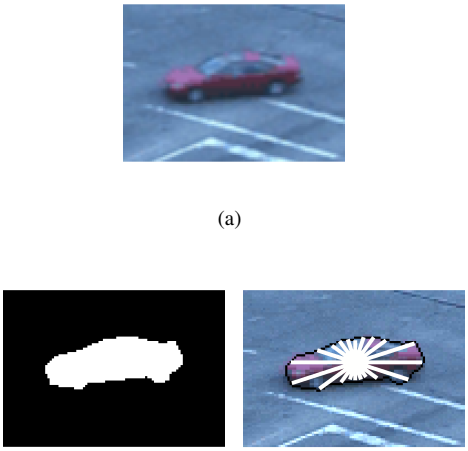
We believe the results in this paper argue that including richer models of foreground shapes in scene modeling would support more specific anomaly detection in addition to improving the initial segmentation of objects from the background.

2. Shape Background Modeling

In this section we describe a simple shape descriptor and a method for estimating a scene-specific and spatially-localized probability distribution over the space of shapes. As with most intensity background modeling approaches, a density estimate is constructed from a set of training examples.

2.1. Building a Training Set

During the training phase, foreground blobs are extracted and a shape descriptor is built for each blob. For a given image we extract foreground blobs by sutracting a mean image, thresholding the difference, and a performing a morphological close operation but any method that produces a binary foreground mask could be used. A given blob is described by its centroid x_i and a set of shape parameters θ_i. The shape descriptor $\theta_i = l_{i,1}, \ldots, l_{i,n}$ describes a polygon where $l_{i,j}$ is the distance from x_i to the edge of the blob in direction $\alpha_j \in [0, \ldots, 2\pi)$, for all results in this paper $n = 20$ and values are evenly sampled. See Figure 2(c) for an example of the descriptor. We use a simple finite-dimensional shape space but other shape descriptors are possible including implicit shape descriptors, such as signed distance functions.



(a)

(b) (c)

Figure 2: **Shape descriptors are extracted from image data using background subtraction.** (a) An example image region containing a foreground object. (b) The image region in (a) with each pixel labeled as either foreground, the blob, or background using a background model. (c) The approximation shape descriptor for the blob (b).

2.2. Building the Shape Eigenspace

In order to improve the robustness of density estimation it is desirable to reduce the dimensionality of our shape descriptor. Our approach is to perform PCA on shapes *near* a pixel location x to determine a low-dimensional shape eigenspace. We begin by formally defining a spatial-support function that defines the meaning of near in our approach.

The influence of shape θ_i on the shape background model at pixel x is determined by a support function $w(x, x_i)$. We primarily use a simple support function w_r that includes only shapes within a radius r of a given point:

$$w_r(x, x_i) = \begin{cases} 1 & \text{if } ||x - x_i|| \leq r \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $||.||$ is the Euclidean distance between pixels and r is adjusted to give a sufficient number of training examples. Note that w_∞ is 1 for all shapes, wherever they occur.

Given an image location x and a support function w_r the next step is creating a low-dimensional subspace that is useful for shapes centered at x. We use a PCA decomposition of all shapes within a radius r of x, *i.e.* the subset of the training set $T = \{\theta_i | w_r(x, x_i) \neq 0\}$. Using the PCA decomposition, the shape descriptor θ_i relative to pixel x can be written as:

$$\theta_i = \theta_\mu + \sum_{j=1}^{n} \lambda_j \beta_{i,j} \psi_j \quad (2)$$

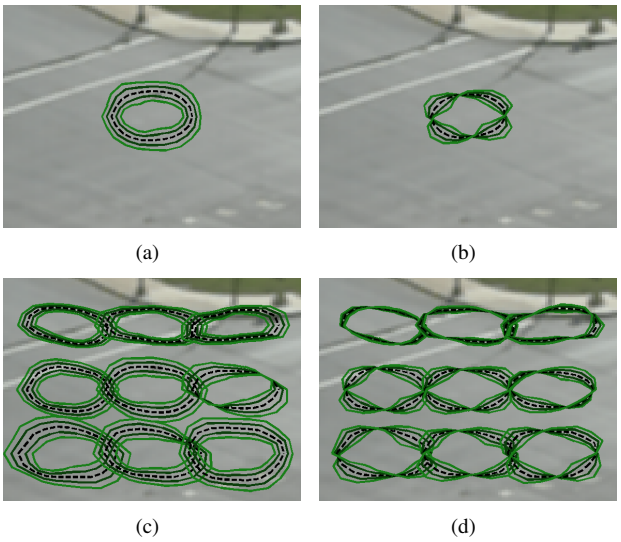


Figure 3: **Shapes reconstructed using PCA by varying the coefficients of the first (a) and second (b) principal components.** From these images we can see that shape subspaces in the top row have much lower variance than those in the bottom two rows. This spatial dependence is the natural result of traffic patterns in the intersection. The shape priors in the top row describe cars all traveling in the same direction, either around a corner or in a lane, while the shape priors in the bottom two rows describe cars traveling in two different directions.

where ψ_j is a principal component and λ_j is the corresponding eigenvalue. The projection of each shape θ_i along the principal direction ψ_j is the coefficient $\beta_{i,j}$.

Before discussing ways to use this decomposition for segmentation and anomalous shape detection, we will first describe properties of the principal directions ψ and the shape parameters β for eigenspaces constructed from real-world videos.

3. Properties of Shape Eigenspaces

In this section we describe properties of automatically generated shape eigenspaces, *i.e.* we examine the values of the coefficients in Equation 2. We begin by directly examining several shape eigenspaces and the coordinates of the training shapes within these eigenspaces.

3.1. Eigenspace Basis Vectors

Figure 3 contains visualizations of the shape eigenspaces built with shapes extracted from a video of a vehicular intersection using either support function w_∞ (global) or w_{20} (local).

The eigenspaces built with support function w_{20} show the dependence between shape eigenspace and image location. The first thing to notice is that the mean shapes (black-dashed line) are different at different image locations. In addition the principal modes of variations are dependent upon image location. In the middle of the intersection the second principal direction encodes for the orientation of the each shape. However, near the top of the image, where cars are primarily traveling in the same direction, the variation due to orientation is much smaller. Note that although a subspace is constructed for each pixel only a subset, sampled every 20 pixels, of the subspaces is shown.

The eigenspace built using w_∞, *i.e.* all shapes in the training set, show that the average shape is approximately an oval with a horizontal long axis, the principle mode of variation is size of the shape, and the second principle mode of variation is the orientation of the shape. This eigenspace, computed over the whole image, provides a data-driven scene specific model of foreground shapes.

3.2. Shape Coordinates in Eigenspace

We now explore the coordinates of the shapes in the training set with the scene-specific, but spatially-uniform, shape eigenspace constructed from the entire training set, *i.e.* using spatial support function w_∞. Figure 4 contains two plots that show a randomly chosen subset of the training shapes. Each point represents a shape in the training set and the coordinates of each point corresponds to the pixel location of the shape centroid. Figure 4(a) shows that there is a strong dependence between the first eigenspace coordinate, the color of the point, and the y-axis location of the shape centroid. Figure 4(b) shows that the dependence of the second eigenspace coordinate and location is not linear. The second eigenvector largely encodes for the orientation of the shape, *i.e.* which road is it on.

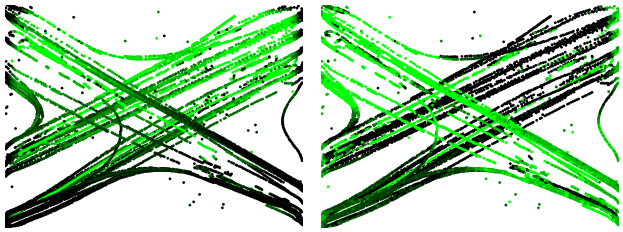
We have shown much can be learned about the foreground objects in a scene by manual inspection of shape eigenspace bases and shape coordinates within the eigenspaces. Understanding the dependence of shape coordinates and image location is key to building accurate shape background models. Next we describe methods for using shape background models to improve the accuracy of segmentation and to enable anomalous shape detection.

4. Applications

In this section we show how to use shape background models to improve object segmentation and enable detection of anomalous shapes.

4.1. Datasets

Our dataset consists of two videos, both of which were captured by a static camera with an overhead view of a road intersection. The majority of foreground objects in both scenes are vehicles, although some pedestrians are present.



(a) PC 1 Coefficient (b) PC 2 Coefficient

Figure 4: **The principal component coefficients of shapes are dependent on location of the shape in the image.** These scatter plots show a set of centroids of training shapes colored by principal component coefficient values using the global eigenspace from Figure 3. (a) The first principal component coefficient, which encodes mostly for size, has a strong linear correlation with the y-axis. (b) The second principal component coefficient, shows a more interesting dependence on location. This coefficient, which encodes primarily for orientation, is dependent on which road the shape was found.

The first video, which we label A, is 9:37 minutes long and has relatively high contrast. The background subtraction algorithm found 183,449 shapes in this video. The second video we will use, labeled B, is 2:05 minutes long with relatively low contrast. The background subtraction algorithm found 21,806 shapes in this video.

4.2. Foreground Object Segmentation

This section describes our integration of scene-specific shape priors into the problem of foreground object segmentation. We first describe our approach to segmenting a single foreground object and, subsequently, describe our approach for segmenting multiple objects.

We describe a segmentation as a curve which minimizes an energy functional $E(\theta_i, x) = E_I(\theta_i, x) + E_S(\theta_i, x)$ that combines image energy E_I and shape energy E_S which we define:

$$E_I(\theta_i, x) = \int_{\phi^+} (c - p_f(x)) d\phi \quad (3)$$

$$E_S(\theta_i, x) = \sum_{j=1}^{n} |\beta_{i,j}| \quad (4)$$

In Equation 3 the area ϕ^+ is the foreground area. The constant c (in our experiments $c = .2$) penalizes area inside the curve that are of low probability. The function $p_f(x)$ is the probability, based on our background model, that the pixel x is foreground. In Equation 4, which penalizes deviations from the mean shape, the $\beta_{i,j}$ is a principal component coefficient. We use the simplex search method to minimize E in the shape subspace defined by the first three principal components; the search is initialized with the mean shape at pixel x.

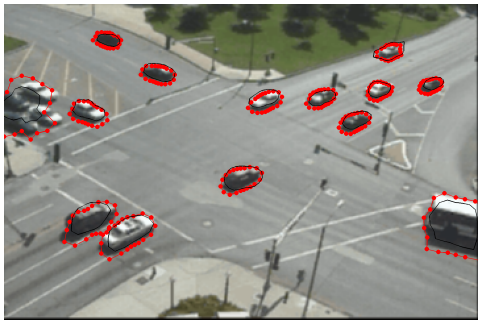


Figure 5: Segmentation results using our spatially-localized shape prior. The dotted lines (red) correspond to segmentation results. The solid lines (black) correspond to segmentation initialization (the mean shape at the centroid of a blob). The large blob on the left side of the image includes many cars. This is the result of the large number of training examples in this area that include multiple cars.

We take an iterative approach to segmenting multiple objects. First, a set of object centroid starting locations is initialized. In this work we initialize with the centroids of the foreground blobs extracted using the method described in Section 2.1. We then perform the single object segmentation approach described above for each centroid with one modification. The foreground function p_f is modified to return 0 for image locations already determined to be foreground within the current frame; this modification helps prevents two starting locations from converging to the same image location.

In order to evaluate the importance of the shape priors, we compare segmentation results for variations of the energy functional E. The first variation is that two different shape subspaces are used: one defined by spatial support function w_∞ and another defined by spatial support function w_{20}. The second variation is whether or not to include the shape energy term E_s in the energy functional. Without this term the shape background model is used to constrain the shapes to a subspace but are not biasing the shape toward the mean shape.

Figure 5 shows results from an example frame from video A using the the w_{20}-subspace and prior. This particular frame is easy to segment because there is little overlap between objects. Figure 6 shows that the strong constraints provided by the w_{20}-subspace significantly improve the segmentation results with overlapping vehicles.

The segmentation results demonstrate that, even using simple methods, the inclusion of location-specific automatically-generated shape priors is advantageous for foreground object segmentation. In the next section we show how to detect anomalous shapes by estimating the

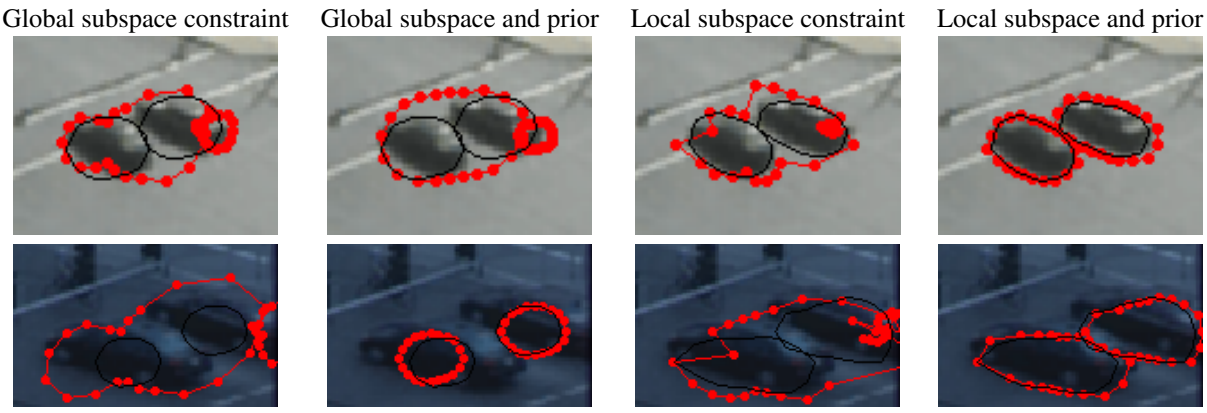


Figure 6: **Segmentation results containing two nearby cars that demonstrate the effectiveness of using a location-dependent shape background model.** In all images the dotted line (red) corresponds to the final segmentation results and the solid line (black) corresponds to the segmentation initialization. The objects in the bottom row of images are difficult to see because of the low contrast in the original video. These figures demonstrate two features of using a location-dependent shape background model: better initialization and better final segmentation.

joint probability of shape and location.

4.3. Anomalous Shape Detection

In this section we describe a method for detecting anomalous shapes using a location-specific shape background model and show promising detection results.

We define anomaly detection in a probabilistic framework by labelling shapes as anomalous if they are of sufficiently low probability with respect to the conditional probability $p(\theta|x)$ of object shape given object centroid image location. Estimation of the density is problematic given the sparsity of training examples and the high-dimensionality of the space—recall that θ is a 20-dimensional shape descriptor. In order to improve the robustness of model parameter estimation we make a number of modifications:

- We assume that $p(\theta|x)$ is constant within a small region defined by support function w, therefore $p(\theta|x) \approx p_x(\theta)$. This notational change emphasizes the fact that we now estimate model parameters using all shapes within the region defined by w.

- We reduce dimensionality by using a subset of the local eigenspace coordinates of θ, more formally $p_x(\theta) \approx p_x(\beta_1, \ldots, \beta_m)$ where β_i is a coordinate in a shape eigenspace built from shapes in the support region defined by w.

- We assume shape eigenspace coordinates are drawn independently from a multivariate Gaussian with mean μ and covariance Σ.

Using these assumptions we consider a shape θ anoma-

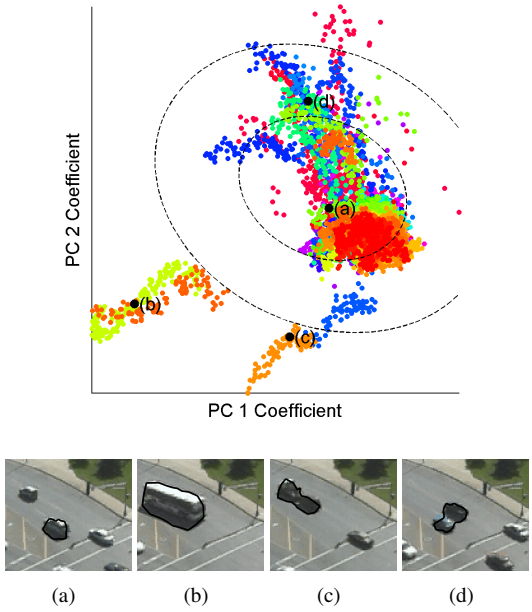


Figure 7: **Detecting anomalous foreground objects using shape priors.** (top) A scatter plot of the first and second principal component coefficients of all shapes with centroids in a small portion of the image. The color of each point reflects when the shape was seen. The dashed ellipses are the first and second standard error ellipses. (a) A typical car shape. (b) An example from a collection of outlier shapes; in each case the underlying object is a bus. From the scatter plot (top) it is evident that two buses passed through the image at different timestamps. (c) An example of a vehicle towing another vehicle. (d) An example of a incorrect segmentation caused by overlapping.

lous if the Mahalanobis distance

$$d(\theta, \mu) = \sqrt{(\theta - \mu)^{\mathsf{T}} \Sigma^{-1} (\theta - \mu)}$$

between the shape and the mean shape μ is greater than some constant k, we use $k = 2$.

Figure 7 shows anomaly detection results for a given sub-window of the video A. The axes are the first and second eigenspace coefficients of all shapes with centroids in the sub-window. Using these simple methods we easily detect two types of anomalous shapes: buses and vehicles towing other vehicles.

5. Future Work

We have intentionally kept algorithms simple to emphasize the strength of the shape background models. This approach leaves open several directions for future work.

Our current density estimation approach assumes that shapes are independent random draws from a Gaussian distribution. We have shown that this assumption is reasonble, however, there is room for improvement. An example of an image location that violates this assumption is in a traffic intersection where vehicles are clustered in several discrete orientations. This points to the need for a local mixture model to better describe the set of previously seen shapes.

We also make the assumption that the distribution of shapes is independent of time but this assumption is often incorrect. For example, the distribution of shapes in the middle of an intersection is dependent on the state of the traffic signals. Anomaly detection results, for example, would be more accurate if this state was known because less variation would be caused by object orientation.

Lastly, we have chosen a simple spatial support function which is binary valued and location independent. We plan to investigate spatial support function that are based on similarity of image regions.

6. Conclusion

This work shows that scene-specific and spatially-localized shape background models can be used to improve the accuracy of object segmentation and enable anomalous shape detection. In addition, we have shown that automatic generation of a set of training shapes using pixel intensity background modeling is sufficient for model parameter estimation.

Our focus in this work has been to explore the capabilities of such a background model while still using straightforward methods. We believe that these initial results argue for further investigation of local shape background models.

References

[1] M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):844–851, 2000.

[2] D. Cremers, F. Tischhäueser, J. Weickert, and C. Schnörr. *International Journal of Computer Vision*, 50(3):295–313, December 2002.

[3] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis. Background and foreground modeling using nonparametric kernel density for visual surveillance. In *Proceedings of the IEEE*, volume 90, pages 1151–1163, July 2002.

[4] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 316–323, June 2000.

[5] A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 302–309, 2004.

[6] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh. Background modeling and subtraction of dynamic scenes. In *Proc. International Conference on Computer Vision*, pages 1305–1312, 2003.

[7] R. Pless, J. Larson, S. Siebers, and B. Westover. Evaluation of local models of dynamic backgrounds. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 73–78, 2003.

[8] M. G. Ross and L. P. Kaelbling. Learning static object segmentation from motion segmentation. In *Proc. National Conference on Artificial Intelligence*, 2005.

[9] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2246–2252, 1999.

[10] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. International Conference on Computer Vision*, pages 255–261, 1999.

[11] Q. Zhang and R. Pless. Segmenting multiple familiar objects under mutual occlusion. In *Proc. IEEE International Conference on Image Processing*, October 2006.

[12] J. Zhong and S. Sclaroff. Segmenting foreground objects from a dynamic textured background via a robust kalman filter. In *Proc. European Conference on Computer Vision*, pages 44–50, 2003.