SANDIA REPORT
SAND2007-1112
Unlimited Release
Printed August 2007

Vision-based Threat Detection in Dynamic Environments

Jeffrey J. Carlson

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2007-1112
Unlimited Release
Printed August 2007

Vision-based Threat Detection in Dynamic Environments

Jeffrey J. Carlson
Intelligent Systems and Robotics Center

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1003

Abstract

This report addresses the development of automated video-screening technology to assist security forces in protecting our homeland against terrorist threats. A prevailing threat is the covert placement of bombs inside crowded public facilities. Although video-surveillance systems are increasingly common, current systems cannot detect the placement of bombs. It is also unlikely that security personnel could detect a bomb or its placement by observing video from surveillance cameras. The problems lie in the large number of cameras required to monitor large areas, the limited number of security personnel employed to protect these areas, and the intense diligence required to effectively screen live video from even a single camera. Different from existing video-detection systems designed to operate in nearly static environments, we are developing technology to detect changes in the background of dynamic environments: environments where motion and human activities are persistent over long periods. Our goal is to quickly detect background changes, even if the background is visible to the camera less than 5 percent of the time and possibly never free from foreground activity. Our approach employs statistical scene models based on mixture densities. We hypothesized that the background component of the mixture has a small variance compared to foreground components. Experiments demonstrate this hypothesis is true under a wide variety of operating conditions. A major focus involved the development of robust background estimation techniques that exploit this property. We desire estimation algorithms that can rapidly produce accurate background estimates and detection algorithms that can reliably detect background changes with minimal nuisance alarms. Another goal is to recognize unusual activities or foreground conditions that could signal an attack (e.g., large numbers of running people, people falling to the floor, etc.). Detection of background changes and/or unusual foreground activity can be used to alert security forces to the presence and location of potential threats. The results of this research are summarized in several MS Power-point slides included with this report.

Acknowledgements

The following individuals are recognized for their contributions to this project and this report:

Wendy Amai, SNL, Department 6644

Brian Henderson, SNL, Department 6634

Colleen Jenkins, New Mexico State University, PSL

Teresa Ko, SNL, Department 8961

Table of Contents

Abstract……....3
Acknowledgements……...4
Table of Contents ……5
Introduction ……..7
Background…….. ……....7
Technical Approach…….……7
Mixture Model for Background Estimation……..8
Investigations for Estimating μ_B when $P_B < 1/(1 + \sigma_F/\sigma_B)$……8
Investigations for Speeding the Estimation of μ_B……9
Investigations for Reducing Nuisance Alarms……13
Presentations……....13
References……... 34
Distribution……....35

Introduction

The problem we are addressing is the development of automated video-screening technology to assist security forces in protecting our homeland against terrorist threats. A prevailing threat is the covert placement and subsequent remote detonation of bombs (e.g., briefcase bombs) inside crowded public facilities. These locations are ideal for terrorist attacks because 1) many facilities do not screen people entering or exiting the facility due to high costs and intense manpower requirements, 2) it is relatively easy to place a bomb unnoticed because of all the surrounding activity, 3) there is a high potential for large numbers of casualties, and 4) the idea of someone bombing a crowded public facility strikes fear into the hearts of nearly all Americans.

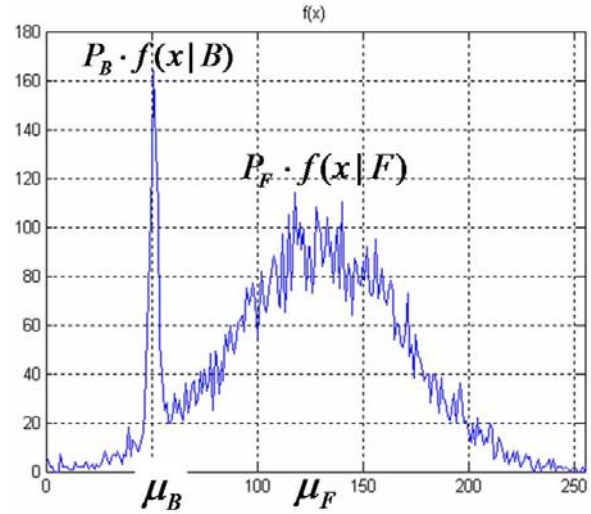
Background

Although video surveillance systems are increasingly common in public facilities throughout the U.S., current systems are unable to detect the placement of bombs. The mere presence of surveillance cameras is assumed to provide some degree of deterrence. It is also unlikely that security personnel could detect a bomb or, someone placing a bomb, by observing live video from surveillance cameras. The problems lie in the large number of cameras required to effectively monitor a large area, the limited number of security personnel employed to protect these areas, and the intense diligence required to effectively screen live video from even a single camera. Automated video motion detection and tracking systems currently exist for detecting intrusions into a monitored, or protected, area (e.g., the perimeter surrounding a nuclear facility). One of the basic underlying assumptions used by algorithm designers of these systems is that the background is free of targets, or motion, most of the time. That is, the camera mostly observes a relatively static background. The performance of these systems is poor in extremely dynamic environments where motion and human activity are persistent (e.g., inside a subway station, airport, or bus depot).

Technical Approach

Our approach was to develop an automated video-screening technology that is capable of *quickly* detecting changes in the static background of an otherwise, dynamic environment. Different from existing video-detection systems designed to operate in static environments, the video-screening technology is capable of detecting changes in the static background of a dynamic environment: environments where motion and human activities are *persistent*. Our goal was to quickly detect background changes, even if the background is visible to the camera less than 5% of the time. Our approach employs statistical scene models based on mixture densities. We hypothesized that the static-background component of the mixture has a small variance compared to dynamic components. Our initial experiments show this is true about 90% of the time. We have identified extensions to these models that will enable accurate estimation of the static background over 99.9 % of the time. This requirement is based on manpower estimates for response. We have demonstrated robust-threat-detection capabilities using subsequent comparisons of static-background estimates to detect changes and to alert security to the presence and location of potential threats (e.g., the placement of a briefcase bomb next to a trash can). A guard can then make a visual assessment of any potential threat and plan an appropriate response.

Mixture Model for Background Estimation

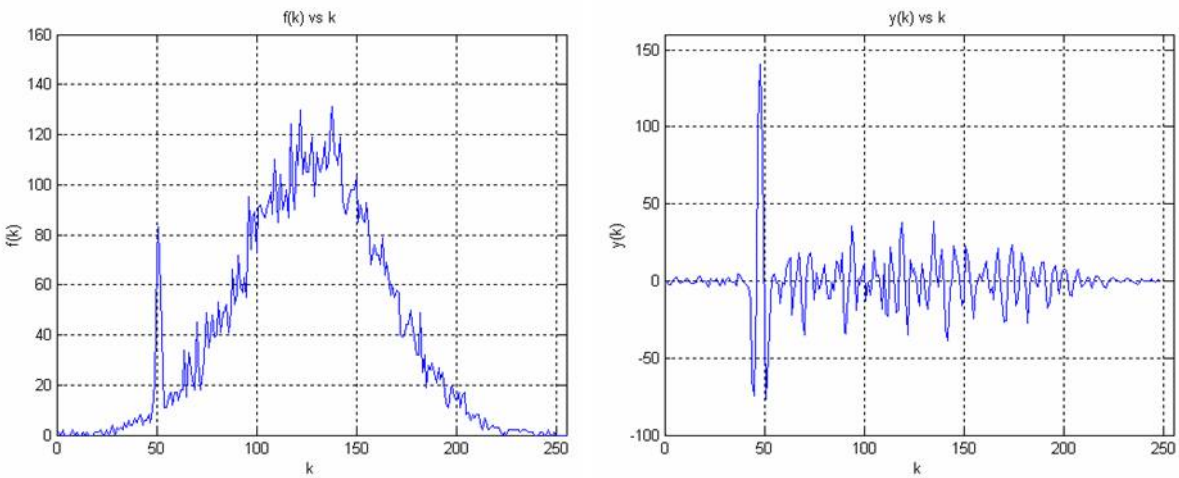


Our baseline approach is based on the results of a 2002 late-start LDRD that funded a study to determine the feasibility of using video-based systems to detect changes in the static background of an otherwise, dynamic environment. The results of that study demonstrated tremendous potential. Our approach employed a statistical scene model based on a mixture of two Gaussian densities. The parameters of the model include the mean and standard deviation of the background component (μ_B, σ_B), the mean and standard deviation of the foreground component (μ_F, σ_F), and the a priori probabilities of both background and foreground, the sum of which add to one ($P_B + P_F = 1$). The parameter of interest (i.e., the one we are trying to estimate) is the mean of the background component, μ_B. We hypothesized that the static-background component of the mixture has a small variance compared to dynamic, or foreground, components. Initial experiments have validated this hypothesis. The previous graph shows an actual histogram of the gray-level values of a single pixel sampled 30 times per second over a 50-second window in an extremely dynamic environment. Notice the standard deviation of background values is much less than the standard deviation of foreground values. Also notice that the sample-mean of the mixture, the statistic traditionally used by video motion detection algorithms to estimate the background, is much closer to μ_F than to μ_B. This shows that the sample mean is not robust for background estimation in a dynamic environment. Instead, as our baseline, we investigated using the sample mode of the mixture to estimate the mean of the background component. The sample mode is the most frequently observed value (i.e., the value indicated by the peak in the histogram). Notice that, for this mixture, the sample mode is a very good estimate of μ_B. A sufficient condition for the mode to be a good estimate of the background mean is $P_B > 1/(1 + \sigma_F/\sigma_B)$. This condition guarantees that the amplitude of the background peak is greater than the amplitude of the foreground peak. If $\sigma_F = 9\sigma_B$, then $P_B > .1$ is sufficient. That is, under this condition, we need to observe the background greater than 10% of the time. A major focus of our technical work was to extend our baseline approach (using the sample mode) to allow accurate background estimation even when $P_B < 1/(1 + \sigma_F/\sigma_B)$. The periods when this condition is true would be the most ideal time for terrorists to conduct an attack. Another focus has been on the development of techniques to reduce the time required to accurately estimate the background and detect changes. Our goal is to give responders plenty of time to assess and engage potential threats.

Investigations for Estimating μ_B when $P_B < 1/(1 + \sigma_F/\sigma_B)$

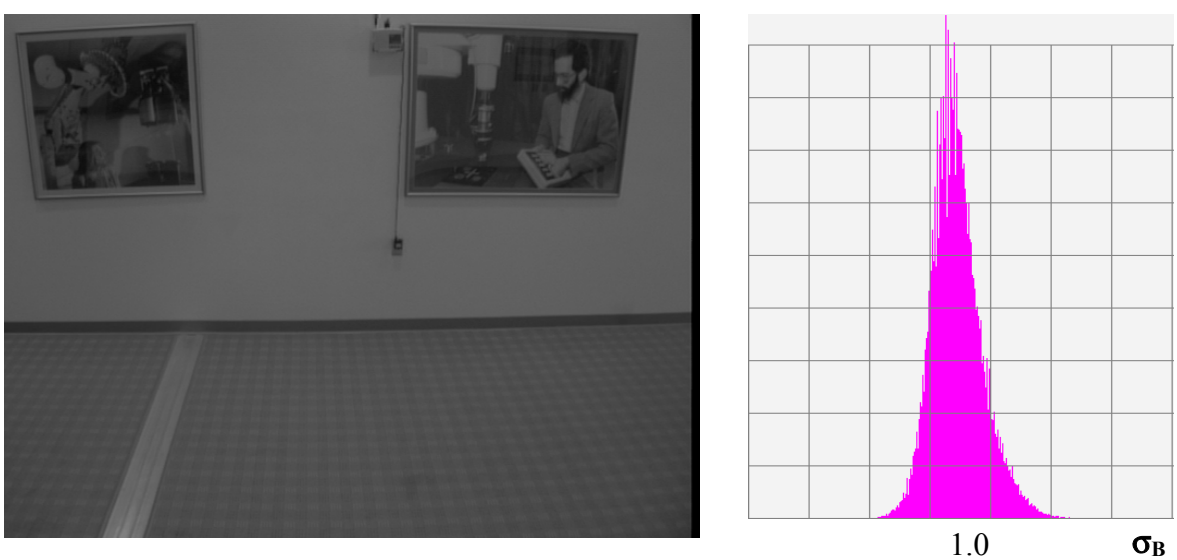
One approach we investigated to achieve this goal was to high-pass filter the histograms for each pixel. Since the background component resembles an impulse, it will readily be passed to the output by the filter. Since the foreground component of the mixture is much

broader, it will be greatly attenuated by the filter. The location of the peak value of the high-pass filtered histogram will identify μ_B. The following figures demonstrate how this approach adequately estimates μ_B even when $P_B < 1/(1 + \sigma_F/\sigma_B)$. In this case, the sample mode of the original histogram on the left would have identified μ_F. The location of the peak value in the filtered histogram correctly identifies μ_B. As will be explained later, this approach can also be exploited to reduce nuisance alarms in the detection algorithm.



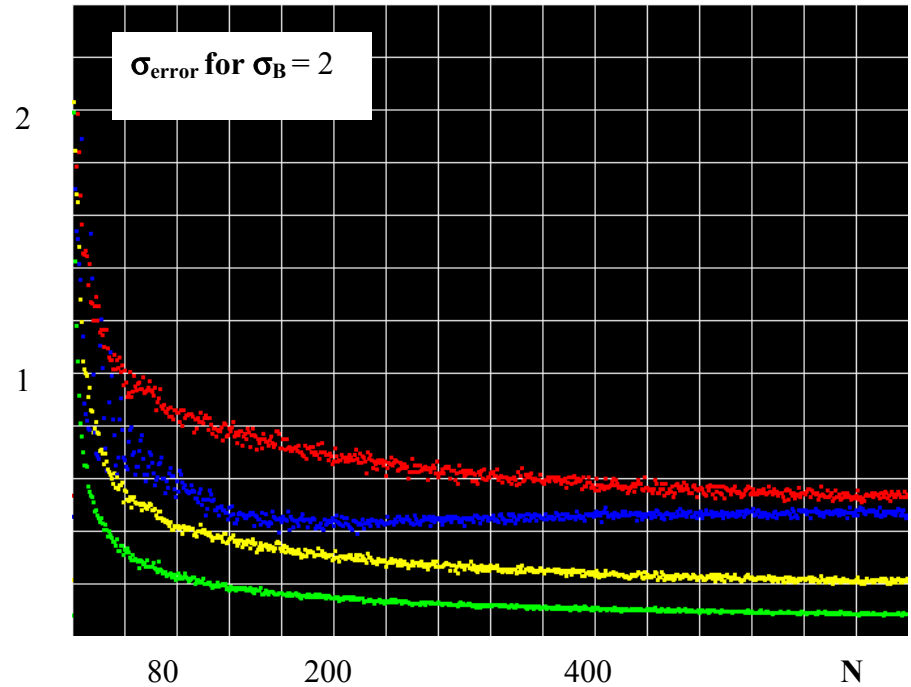
Investigations for Speeding the Estimation of μ_B

Another key focus of our work was to speed up the estimation process. The two goals (estimate more quickly and observe less frequently) actually conflict. The number of samples required to estimate μ_B depends on the standard deviation of background pixels. Empirical results indicate that σ_B is generally 2 or less. We tested this by observing different static backgrounds over several hundred frames and computed the standard deviation of the gray level for each pixel in the image. Shown is a sample image from the static background sequence and a histogram of the standard deviation values computed for each pixel in the image.



We found the results vary somewhat from camera to camera and also depend on the automatic gain control (AGC) setting of the camera. Under low-light conditions, the AGC automatically increases the gain and the pixel standard deviation increases (i.e., becomes noisier). Under daylight and normal room lighting, however, the results shown in the previous histogram are typical.

With knowledge of σ_B, we conducted experiments to determine how many samples were required for the sample mode to be a good estimate of the population mean given we only observe representative background samples. We used the standard error of the estimate as a measure of goodness.

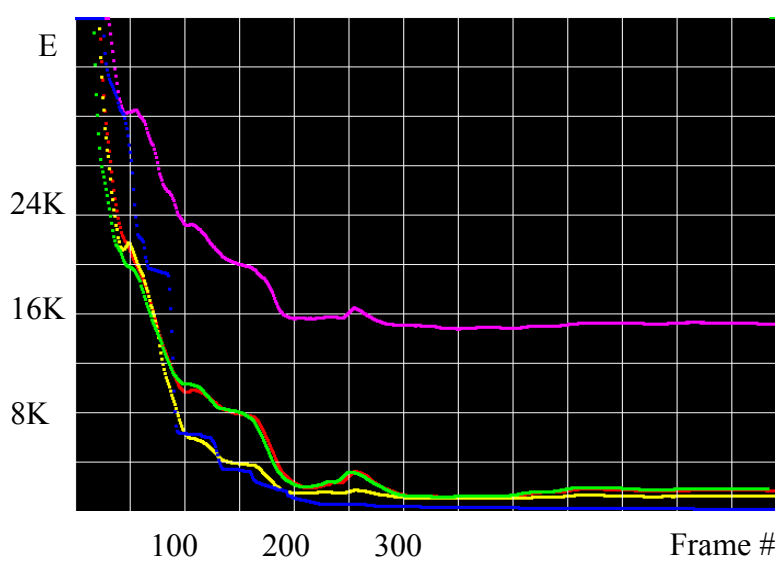


Defining the error of the estimate as error = estimate - μ_B, the standard error is simply the standard deviation of the error. An estimator is said to be consistent if the standard error tends towards zero as the number of samples used to formulate the estimate tends towards infinity. When comparing different estimators, the estimator with the smallest standard error is best. It is well known from statistics that the sample mean is the maximum likelihood (ML) estimate of the population mean. It is also well known that if a ML estimate exists, then no other estimator will have a smaller standard error (i.e., it is the minimum variance estimator). The ML estimate, however, is often not robust in the presence of outliers and, as was shown earlier, is not robust for background estimation in dynamic environments. We can, none the less, use the standard error of the sample mean as a bound on estimator performance. We desire a robust estimator whose standard error rapidly approaches zero, or at least approaches the standard error of the sample mean, as the number of samples increase. The following graph plots the standard error, σ_{error}, as a function of the number of samples, N, for three different estimators of μ_B when $\sigma_B = 2$. Note in this experiment, the estimate is based solely on observations of background (i.e., no foreground activity present). The red curve shows results for our baseline estimator, the sample mode statistic. The blue curve shows results obtained using the sample mode of low-pass filtered (i.e., smoothed) pixel histograms. The yellow curve shows results using a trimmed mean statistic. Finally, the green curve demonstrates the performance bound obtained using the sample mean. These results indicate that between 10 and 40 samples are

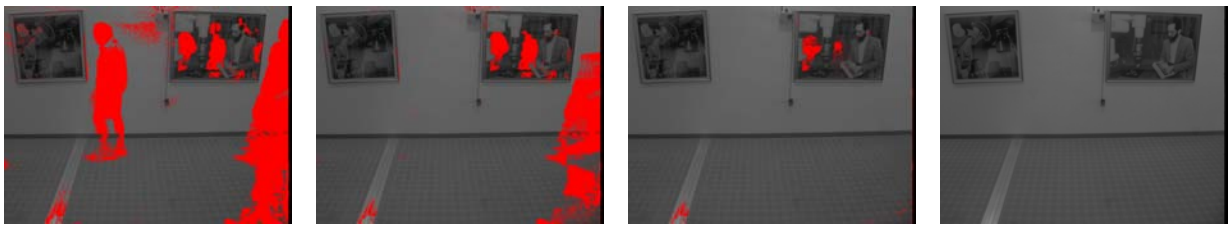
necessary to reduce the standard error below 1. For a standard deviation of 1 gray level, the true value of the background will be within +/- 3 gray levels of the estimate 98% of the time (assuming a Gaussian distribution). For change detection purposes, this would require a detection threshold setting greater than 4. Note that the performance of these estimators tends to level off after approximately 200 observations of background have been obtained. Sometimes even more than 200 samples could be required depending on if the background is relatively static, or also contains dynamic components (e.g., leaves rustling due to wind). The estimator based on the trimmed mean statistic gave the best performance. This statistic is formulated by computing the center of mass of histogram bins that are within a specified radius of the sample mode. The trimmed mean statistic has also proven robust for background estimation. Note that since the trimmed mean converges to the background faster than the sample mode, it provides a speed up in background change detection.

Another important performance indicator for background estimation is the number of observed video frames required to formulate an accurate background estimate. Based on the previous results, this number depends on N (the number of background observations required to estimate μ_B). Empirical results suggest that up to 40 samples are required if we continuously observe background. If we only observe background P_Bx100% of the time, then the minimum number of video frames required for background estimation is equal to N/P_B. For example, if background is observed only 10% of the time, then 40/.1 = 400 frames are required. We investigated several approaches to reduce the time required to obtain an accurate background estimate and to detect changes. Based on the previous equation, reducing the number of frames requires that either we decrease N or we increase P_B. Although this might not seem possible, we have employed various innovative techniques to do this.

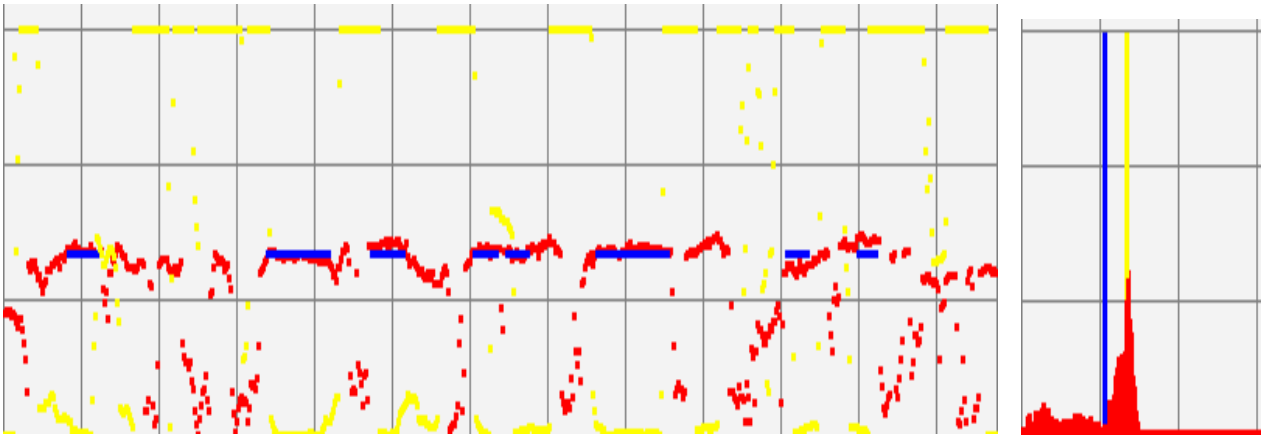
In addition to the sample mode, the sample mode of the smoothed pixel histograms, and the trimmed mean statistic, we examined 3 additional techniques to speed up the estimation process. One technique leverages the statistical similarities of spatially proximate pixels – allowing us to reach N in fewer frames by combining multiple pixel values in each individual frame. Another technique looks for short periods of extended stationarity in the pixel statistics. This technique uses a statistic that is not related to the mode. A major advantage of this approach is that it doesn't require memory allocation for histograms for every pixel (approximately 150Mbytes for a 640x480 pixel image). A third technique uses multiple cameras observing a



common background from different viewpoints. The idea is that while part of the background might be occluded from one camera, the chances are that it might be visible to other cameras (i.e., with multiple viewpoints, we are actually increasing the probability that we will observe the background). The previous graph demonstrates the performance of these estimators in an extremely dynamic environment by comparing estimates at each frame with a background "truth" image acquired under static conditions. For each pixel, the absolute difference between the estimate and the truth value is compared to a threshold of 6. The vertical axis indicates the number of differences that exceed the threshold in each frame. Ideally, this number should go to zero. The red curve shows the performance of the baseline sample mode estimator. The yellow curve shows the performance of the trimmed mean estimator. The green curve shows the performance using spatially proximate pixels to build the histograms more quickly. The blue curve illustrates the performance based on the extended stationarity concept. Of all the techniques investigated, the extended stationarity approach exhibited the best convergence. Note that after just 300 frames, the number of misclassified pixels was zero for this method. The following sequence of images illustrates the classification error based on this approach after 1, 50, 150, and 300 frames (red pixels indicate errors).

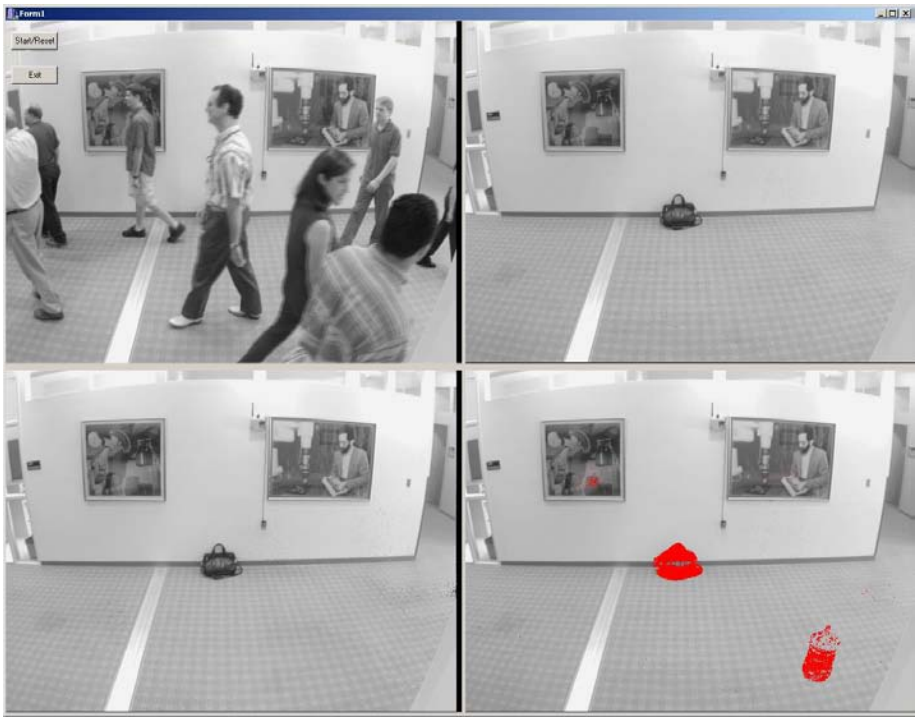


The following graph represents the concept of extended stationarity for just one pixel. The standard deviation of pixel values over the past 15 frames is computed. If the standard deviation was less than 1, then those pixels are declared stationary. Detected stationary pixel regions are marked in blue. These values are then recursively averaged to compute the background estimate. The red curve indicates actual pixel values. The yellow curve is the measured standard deviation of pixel values. The histogram shown to the right shows the sample mean estimate of the background (blue line) and the estimate based on extended stationarity (yellow line). Note how accurately the estimate based on extended stationarity predicted the true background value (in this case, in agreement with the sample mode).



Investigations for Reducing Nuisance Alarms

We successfully demonstrated robust threat detection capabilities using subsequent comparisons of static-background estimates to detect changes (e.g., the placement of a briefcase bomb). The following picture illustrates results based on the mode statistic. The upper left picture is from the actual input stream. The lower left shows the current background estimate. The upper right is the previous background estimate. The lower right shows detected changes. The red blob in the lower right hand corner was the result of a water bottle that was removed from the scene. Although the system performed quite well, major obstacles exist in keeping nuisance alarms at acceptable levels.



Nuisance alarms can be caused by benign changes in the background due to shadows and other lighting effects. Nuisance detections are visible in the detected image (lower-right image) in the left-hand picture hanging on the wall. We explored various approaches to minimize nuisance alarms. One approach exploited the estimation algorithm directly. Our background estimation technique is unique in that it can detect if it is unable to formulate a background estimate. The presence of background is indicated by an impulse, or narrow peak, in the pixel histogram. If this peak is absent for a particular pixel, we simply ignore that pixel for detection purposes. We also explored another approach more suited to the extended stationarity technique. Here we just counted the number of detected stationary pixel intervals over a specified time period. If the total number of pixels included in the stationary intervals is too small, then that pixel is ignored for detection purposes. We also successfully demonstrated the benefit of high-pass filtering the input video stream to reduce the sensitivity to lighting variations. This approach has been successfully applied in more traditional video motion detection algorithms and has proved just as suitable for our background estimation algorithms.

Presentations

The following power-point slides summarize the results of this research.

Background Estimation & Change Detection in Extremely Dynamic Environments

8-16-2006

Team Members

Jeff Carlson, 6646 (Overview)

Teresa Ko, 8961 (Extended Stationarity)

Colleen Jenkins, NMSU PSL (Pearson's Method of Moments)

Wendy Amai, 6644 (Baseline Results)

Brian Henderson, Grad Student Intern (Funded Application)

Sandia National Laboratories

Project Goal

- **Develop video-based technology to rapidly detect changes in the background of extremely dynamic environments: environments where motion and human activities are persistent over long periods – possibly never free from foreground activity**

Sandia National Laboratories

Intended Application

- **Monitor crowded areas for placement of Improvised Explosive Devices (IEDs)**

(1) Urban Scene

(2) IED Placement

(3) Urban Scene Continues

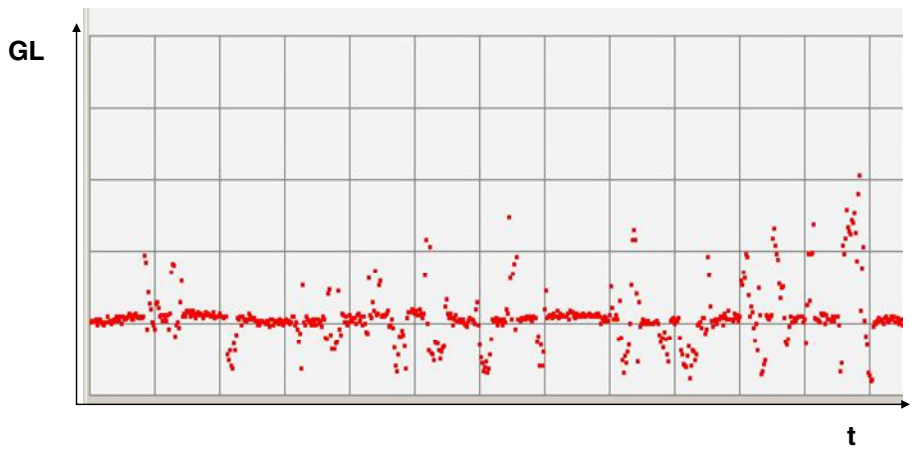
(4) Automatic Detection

Sandia National Laboratories

Video Clip

Sandia National Laboratories

Pixel Intensity Characteristics



Statistical Scene Model

- The pdf of each pixel can be modeled as a Gaussian mixture:

$$f(x) = P_B \bullet f(x \mid B) + P_F \bullet f(x \mid F)$$

Where *x* is the gray-level or feature value for each pixel,

$$f(x \mid B) \sim N(\mu_B, \sigma_B)$$
$$f(x \mid F) \sim N(\mu_F, \sigma_F)$$

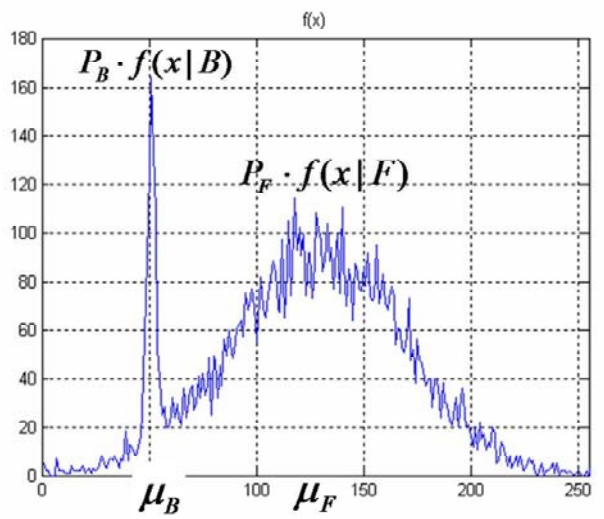
(B = Background, F = Foreground)

And we assume

$$\sigma_F >> \sigma_B \quad , \quad P_F >> P_B$$

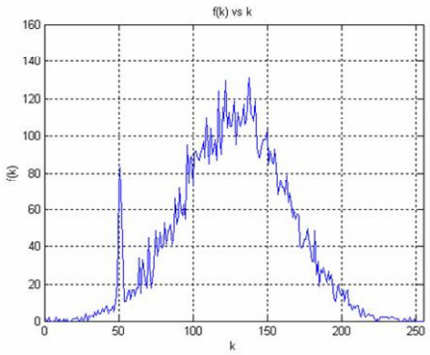
- The scene model consists of a 2D array of mixtures - one for each pixel.
- The background estimate is determined from an estimate of μ_B
- The statistic we use for estimating the mean background level at each pixel is the sample mode of each mixture density.

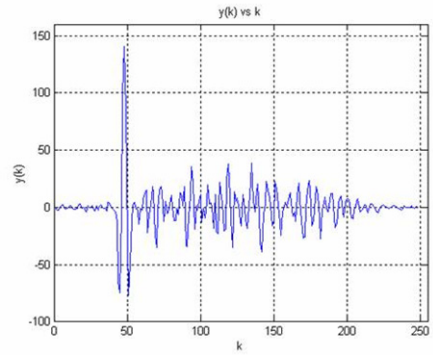
Estimating μ_B when $P_B > 1/(1 + \sigma_F/\sigma_B)$



Sandia National Laboratories

Estimating μ_B when $P_B < 1/(1 + \sigma_F/\sigma_B)$





Sandia National Laboratories

Other Approaches & Limitations

- Compare scenes before and after detected activity
  - Sometimes activity is persistent over long periods
- Look for new foreground objects that remain stationary (e.g., Lockheed Martin NY Subway system)
  - Doesn't work well in very dynamic conditions
- Current video motion detection and tracking systems assume the scene is free of motion, or foreground activity, most of the time.
  - This assumption allows pixel averaging to be a good estimate of the background.
  - Averaging does not provide robust background estimates in extremely dynamic scene environments.

Sandia National Laboratories

Programmatic Accomplishments

- Presentation to Pace VanDevender
  - Pace recommended application for R&D 100 award
- Initiated partnership with security personnel at the Albuquerque International Sun Port for the purpose of extended algorithm development, testing, and evaluation in a real operational setting
- Invited presentation on research developments to DHS in regards to the America's Shield Initiative (June 7th)
- Provided several demonstrations/presentations to potential customers (DARPA, Port Security Authorities, U.S. Army, Lockheed Martin)
- Technical Advance prepared

Sandia National Laboratories

Sun Port Installation

Sandia National Laboratories

Technical Accomplishments

- Developed capability to rapidly estimate background even when background is observed less than 5%
- Developed robust techniques to significantly speed up background estimation and change detection
- Developed new, robust algorithm for background estimation that eliminates the mass-memory requirements of our baseline approach based on the sample-mode statistic
- Developed robust algorithms for reducing nuisance alarms
- Investigated traditional approaches for probability mixture estimation (Pearson's method of moments)

Sandia National Laboratories

Marketing Troubles

- The terrorist attacks involving London's mass transit system demonstrated the extremely high value of video surveillance for post-attack assessment.
- Subsequent to the London attack, Lockheed Martin was awarded a $212M, 3 year contract to install a video surveillance system with 1000 cameras in the New York Subway system.
- Lockheed advertised the ability to instantly detect background changes (e.g., someone leaving a brief case unattended).
- A demonstration of this capability, not mentioning any of the limitations, was broadcast on public news channels world wide.
  - Their approach detects new foreground objects that remain stationary.
  - This approach is inadequate in extremely dynamic environments – ones where terrorists would prefer to carry out an attack.
- CUSTOMERS (E.G., DARPA, DoD, DOE) HAVE THE PERCEPTION THAT THIS IS NOW A SOLVED PROBLEM.

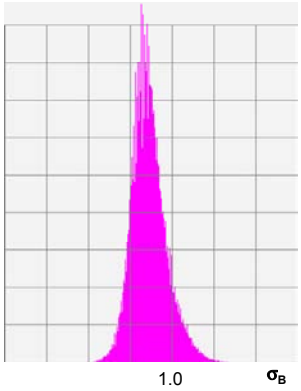
Sandia National Laboratories

Results

- This research has enormous potential for protecting our Nation from terrorist attacks.
- Depending on foreground statistics, we found that the sample-mode was adequate if we could expect to observe the background greater than 5% of the time (i.e., $P_B > .05$).
- Depending on the characteristics of foreground activity, detection can take as long as several minutes.
- High-pass filtering pixel histograms enabled robust background estimation even when $P_B < .05$.
- Extended Stationarity algorithms enabled significant increases in the speed of background estimation.

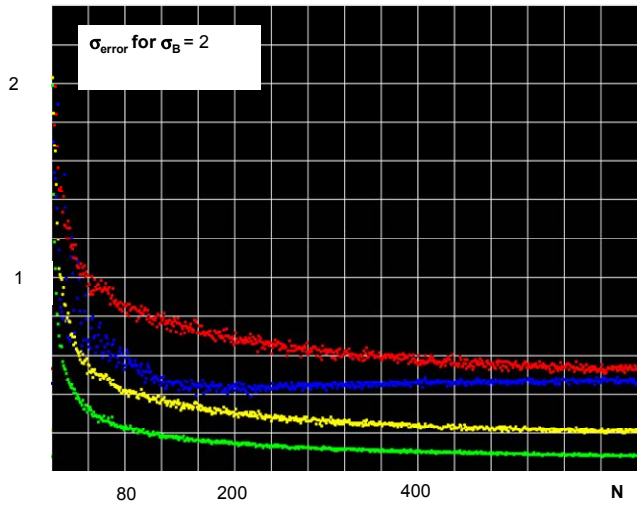
Sandia National Laboratories

Empirical evaluation of σ_B



σ_B is generally < 1 and is most always < 2.

How many observations of B are required to estimate μ_B?



Red = sample mode
Blue = sample mode of smoothed pixel histograms
Yellow = trimmed mean
Green = sample mean

Error = estimate – μ_B

σ_{error} = standard deviation of error.

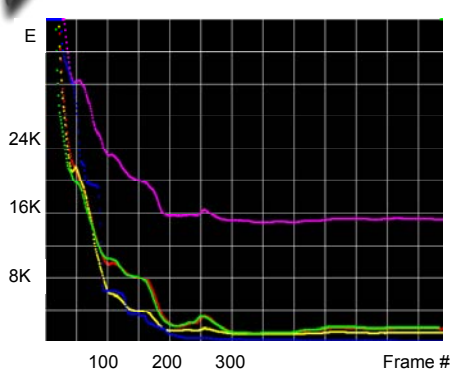
For σ_B = 2, between 10 and 40 observations of B are required.

How many video frames required to estimate μ_B?

M = # of video frames required
N = # of observations of B to estimate μ_B
$M = N/P_B$

If N = 40 and P_B = .1, then M = 40/.1 = 400

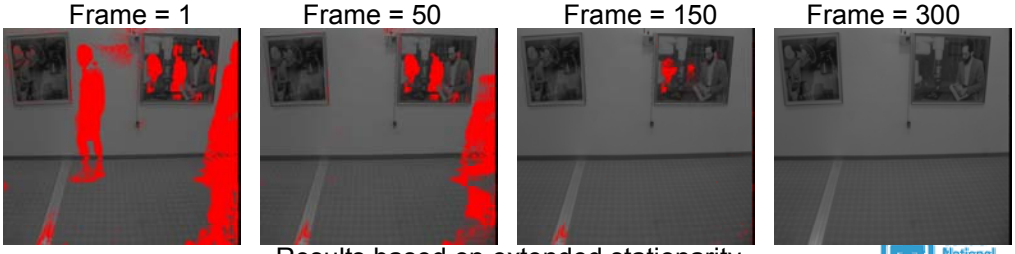
Estimation Speeds for Different Statistics



Red = sample mode
Blue = extended stationarity
Yellow = trimmed mean
Green = sample mode obtained by including spatially proximate samples in individual pixel histograms

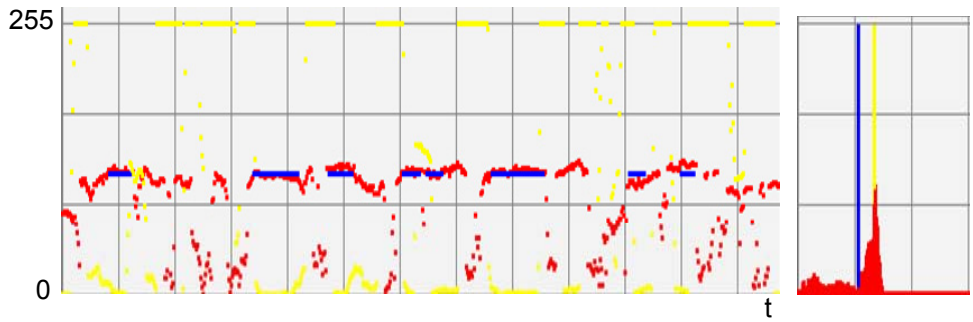
E = # of misclassifications in comparison of background estimate with "truth" image.

Frame = 1 Frame = 50 Frame = 150 Frame = 300



Results based on extended stationarity

Extended Stationarity Concept



Plot of pixel value x (red) as a function of time

Yellow = standard deviation of x values in a window of 15 consecutive samples

Blue = detected periods of extended stationarity

Histogram of pixel values (red) on right shows sample mean estimate (blue) and extended stationarity estimate (yellow) of μ_B.

Minimizing Nuisance Alarms

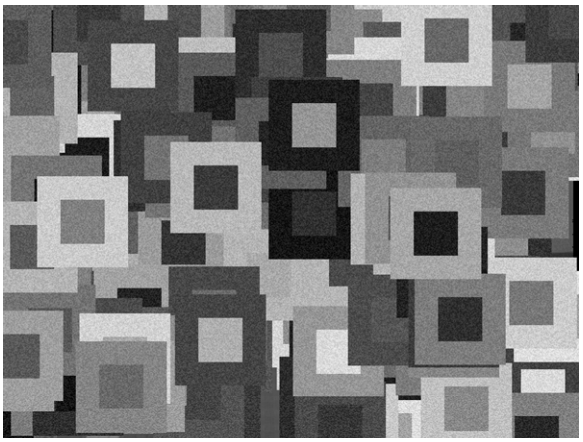
- **Require narrow peak to exist in pixel histogram – otherwise ignore pixel**
- **Require periods of extended stationarity to be detected – otherwise ignore pixel**
- **High-pass filter input images to reduce sensitivity to ambient lighting variations, shadows, etc.**

Pearson's Method of Moments

- Research partners at NMSU have been evaluating traditional methods (Pearson's method of moments) for estimating the parameters of a mixture density.
- These methods are generally much more computationally intensive than the methods based on the sample mode or the method based on extended stationarity.
- Can, however, estimate the background mean under more severe conditions

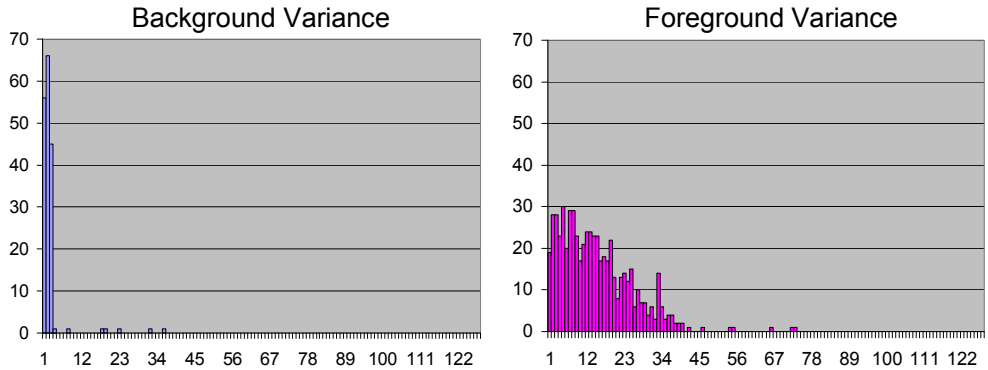
Sandia National Laboratories

Background Monitoring At 1% Background Visibility



Variance Histogram

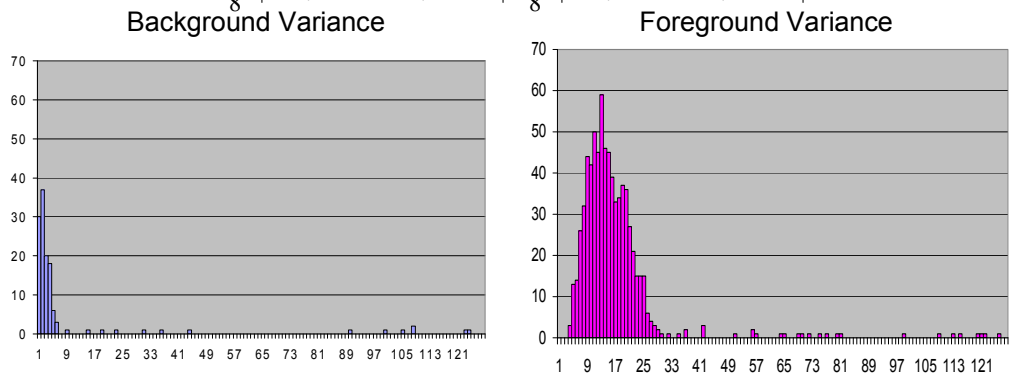
$$Variance = \left| p_{x,y}(t) - p_{x,y}(t-1) \right|$$



- Goal: Filter out pixels with *high variance*, because they are likely to be foreground pixels.

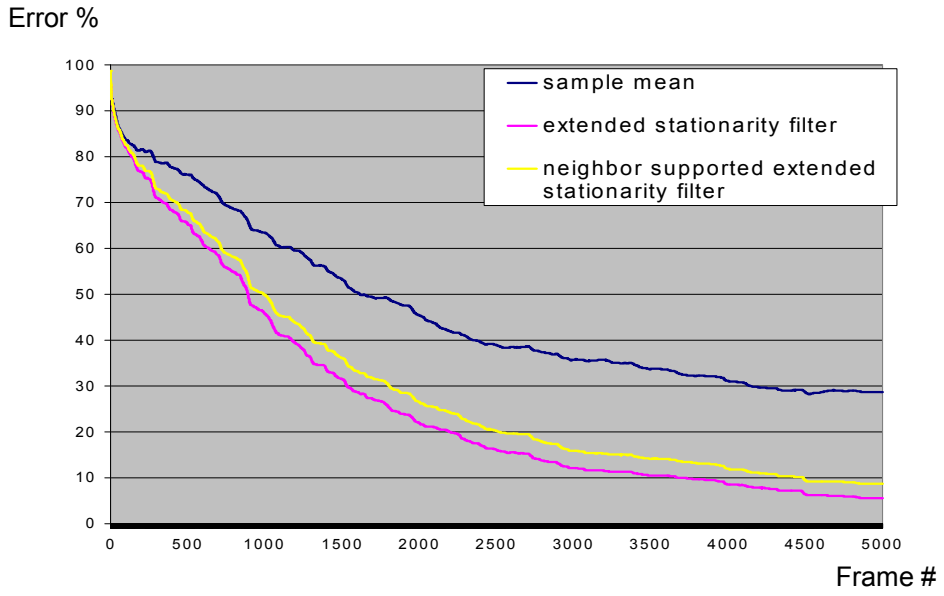
Neighbor Supported Variance Histogram

$$Variance = \frac{1}{2} * \left| p_{x,y}(t) - p_{x,y}(t-1) \right| +$$
$$\frac{1}{8} * \left| p_{x-1,y-1}(t) - p_{x-1,y-1}(t-1) \right| + \frac{1}{8} * \left| p_{x-1,y+1}(t) - p_{x-1,y+1}(t-1) \right| +$$
$$\frac{1}{8} * \left| p_{x+1,y+1}(t) - p_{x+1,y+1}(t-1) \right| + \frac{1}{8} * \left| p_{x+1,y+1}(t) - p_{x+1,y+1}(t-1) \right|$$

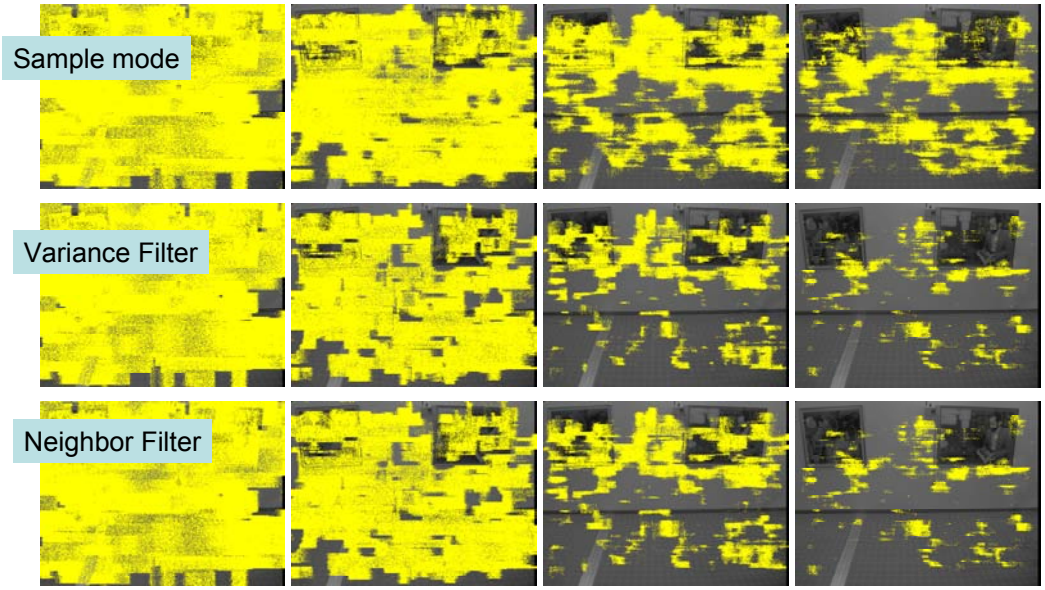


- Goal: Filter out pixels with high variance and *neighbors who have high variance*, because they are likely to be foreground pixels.

Algorithm Performance at 1% Background Visibility



Estimated Background



Frame: 50 500 2000 5000

Statistical Scene Model

Assumptions

For a large number of samples of any one pixel in a surveillance video the intensity histogram should be a mixture of two Gaussian functions:

- the background intensity – small variance
  for example, a wall.
- the foreground intensity – large variance
  for example, people walking.

Sandia National Laboratories

Physical Science Laboratory
New Mexico State University

Pearson's Method of Moments

Pearson's Method of Moments yields estimates of the means, variances and the relative weight of each distribution.

Using the Estimates

Assuming the background component is the distribution with smaller variance, the background mean can be observed and changes in the mean will imply changes in the background.

Application

- Even if the means of the distributions are not obvious visually in relative frequency histograms, this method can still extract them.
- Although more computationally intensive, this method requires relatively few frames (<100) and is therefore closer to a video "real-time" algorithm.

Sandia National Laboratories

Physical Science Laboratory
New Mexico State University

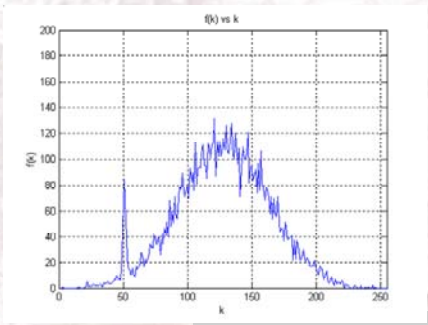
When is the Mode a Good Estimate?

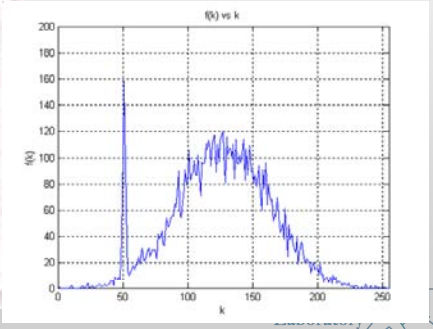
$$P_B < \frac{1}{1+\frac{\sigma_F}{\sigma_B}}$$

Mode ≅ μ_F

$$P_B > \frac{1}{1+\frac{\sigma_F}{\sigma_B}}$$

Mode ≅ μ_B

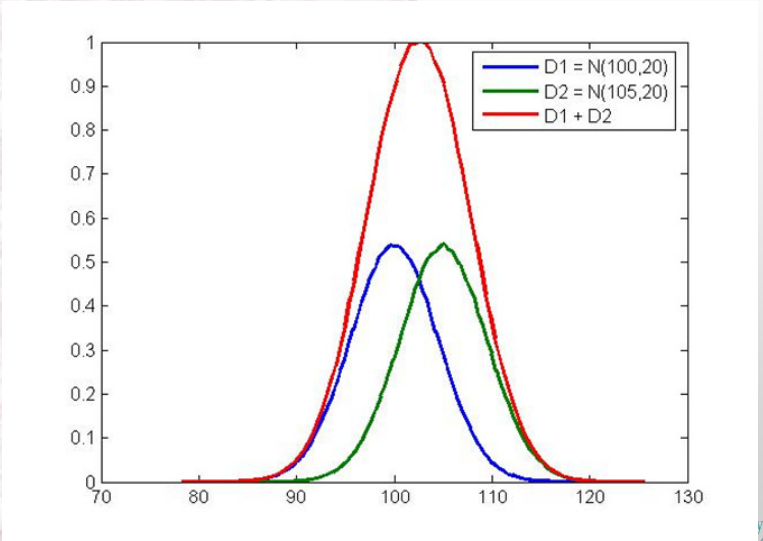




Sandia National Laboratories

New Mexico State University

Pearson's Method of Moments will remove completely buried components in a mixture such as the one below.



Sandia National Laboratories

New Mexico State University

More on the Method of Moments Estimation

The method of moments proves to be a useful tool in the separation of mixtures.

Consider the mixture of two univariate probability density functions with parameters: mean and variance:

$$f(x;\mu_1,\mu_2,\sigma_1,\sigma_2,p) = pg_1(x;\mu_1,\sigma_1) + (1-p)g_2(x;\mu_2,\sigma_2)$$

By setting the observed moments: $$V_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$$

equal to the theoretical moments: $$v_r = \int (x-\mu)^r f(x)dx$$

$$r = 0,1,\ldots,5$$

Sandia National Laboratories

Physical Science Laboratory
New Mexico State University

More on the Method of Moments Estimation

we obtain the following system of five nonlinear simultaneous equations:

$$p\delta_1 + (1-p)\delta_2 = 0$$
$$p(\sigma_1^2 + \delta_1^2) + (1-p)(\sigma_2^2 + \delta_2^2) = V_2$$
$$p(3\sigma_1^2\delta_1 + \delta_1^3) + (1-p)(3\sigma_2^2\delta_2 + \delta_2^3) = V_3$$
$$p(3\sigma_1^4 + 6\sigma_1^2\delta_1^2 + \delta_1^4) + (1-p)(3\sigma_2^4 + 6\sigma_2^2\delta_2^2 + \delta_2^4) = V_4$$
$$p(15\sigma_1^4\delta_1 + 10\sigma_1^2\delta_1^3 + \delta_1^5) + (1-p)(15\sigma_2^4\delta_2 + 10\sigma_2^2\delta_2^3 + \delta_2^5) = V_5$$
$$\delta_k = (\mu_k - \mu), \quad k = 1,2.$$

Pearson solved this system of equations to form a ninth degree polynomial in u.

Sandia National Laboratories

Physical Science Laboratory
New Mexico State University

More on the Method of Moments Estimation

Applying a real negative root of the ninth degree polynomial as $\hat{u}$ in the following equation and solving for δ_1, δ_2 as roots :

$$\delta^2 - \frac{\hat{w}}{\hat{u}}\delta + \hat{u} = 0$$ where $\hat{w}$ is a polynomial in $\hat{u}$

yields estimates of δ_1, δ_2 as roots.

The estimates of the five parameters of the mixture may be derived as follows:

$$\hat{\mu}_1 = \bar{x} + \hat{\delta}_1 \qquad \hat{\sigma}_1^2 = \frac{1}{3}\hat{\delta}_1\left(\frac{2\hat{w} - V_3}{\hat{u}}\right) + V_2 - \hat{\delta}_1^2$$

$$\hat{\mu}_2 = \bar{x} + \hat{\delta}_2 \qquad \hat{\sigma}_2^2 = \frac{1}{3}\hat{\delta}_2\left(\frac{2\hat{w} - V_3}{\hat{u}}\right) + V_2 - \hat{\delta}_2^2$$

$$p = \frac{\hat{\delta}_2}{\hat{\delta}_2 - \hat{\delta}_1}$$

Sandia National Laboratories

Physical Science Laboratory
New Mexico State University

Considerations to take when using the Method of Moments Estimation in Practice

- Occasionally, more than one real negative root will come from the ninth degree polynomial. In this case the theoretical sixth moment should be computed using the results from all sets of estimations and the root that agrees best with the observed sixth moment chosen.
- When the mean of the two distributions are equal the method of moments will continue to estimate the mean and variance within the same degree of error.
- However, when the variance of each distribution are equal the method of moments will accurately estimate the means but the variance may not be correct. This statement is made based on the manipulation of data created using a random number generator.

Reference: Everitt, B.S., and Hand, D.J., *Finite Mixture Distributions*, London: Chapman and Hall, 1981

Sandia National Laboratories

Physical Science Laboratory
New Mexico State University

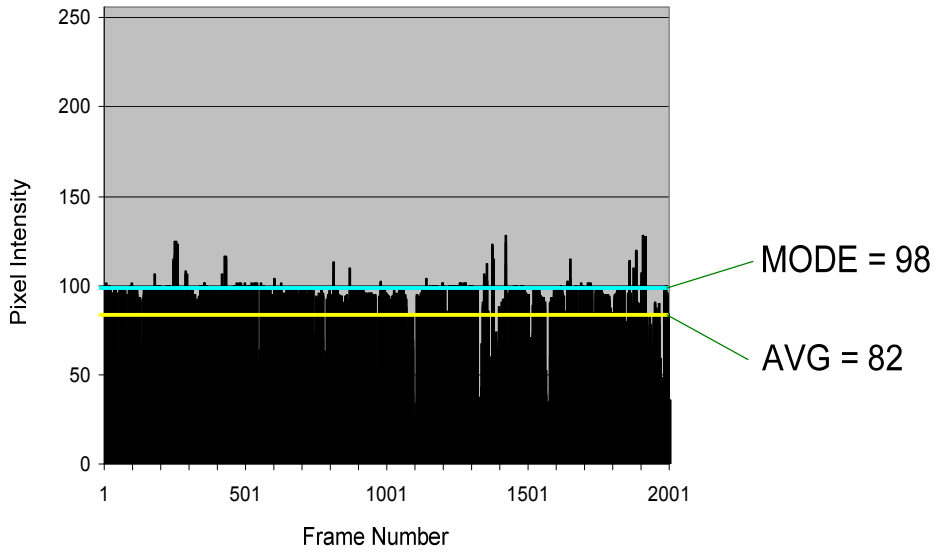
Summary of Methods and Request for Questions

- The modal method of estimating the intensity value of background can take as many as 1000 frames, approximately 30 seconds of average video time. However, this method is much less effort computationally than MOM which is very important when considering 300,000 pixels per frame.
- When the background component of the mixture is buried within the foreground component, the baseline modal method will fail.
- The method of moments will closely estimate the means of those more troubling situations described above and can take as few as 100 frames or 3 seconds of video, however, this method is very processor and memory intensive involving solving several non-linear simultaneous equations for each pixel.
- A combination of the two methods is being considered. Research is in progress to improve the computational efficiency of using the method of moments for single mean estimation.

Sandia National Laboratories

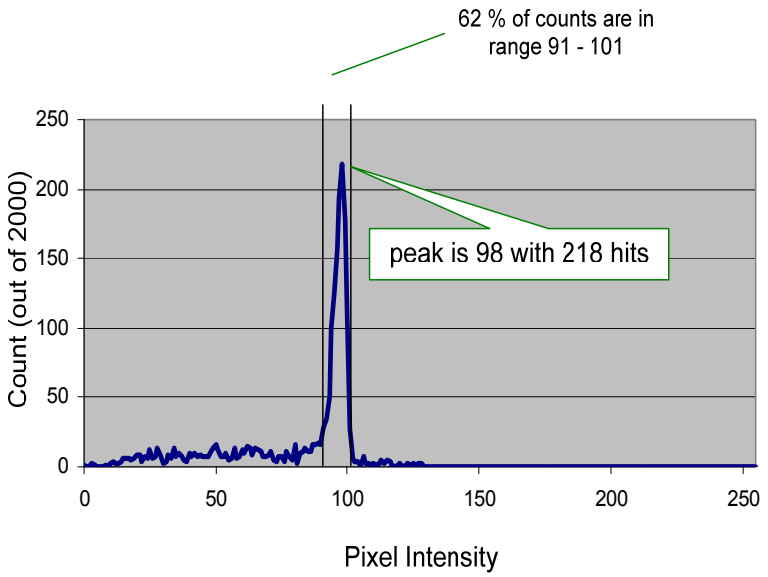
Physical Science Laboratory
New Mexico State University

Pixel Intensity Over 2000 Frames

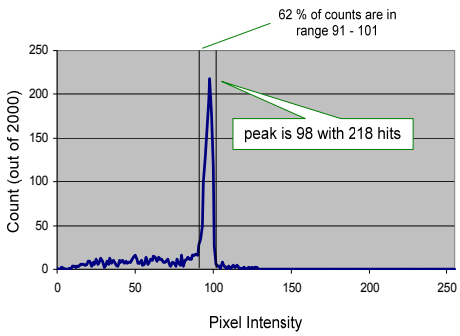


Sandia National Laboratories

Pixel Intensity Histogram



Pixel Intensities
Normal vs Extremely Dynamic



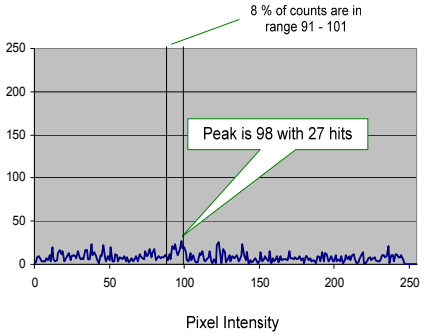
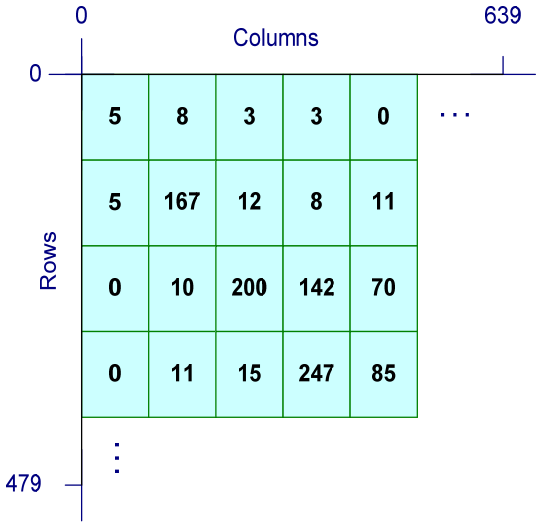
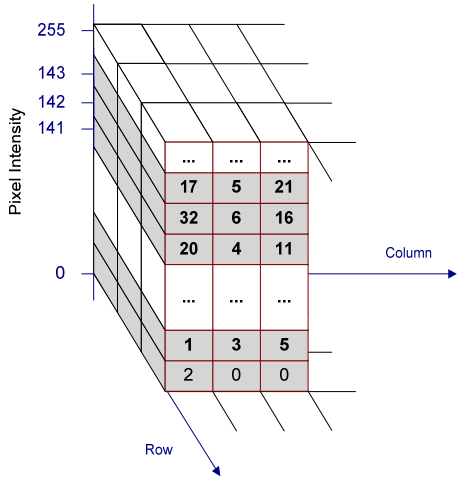


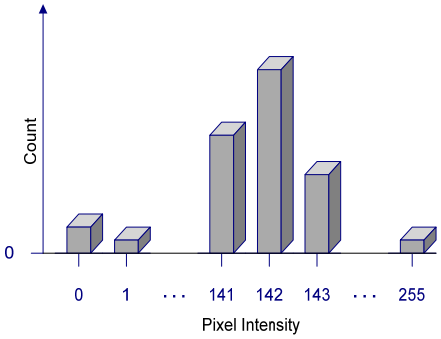
Image Data



480 rows X 640 columns = 307,200 pixels

Pixel Intensity Histogram





References

1. **Van Trees**, H., Detection, Estimation, and Modulation Theory Part I (J. Wiley and Sons, New York, 1971).

2. Athanasios **Papoullis**, "Probability, Random Variables, and Stochastic Processes", Third Edition, McGraw Hill.

Distribution

1	MS 0123	D. L. Chavez 01011
1	MS 1009	J. J. Carlson 06474
2	MS 9018	Central Technical Files, 8944
2	MS 0899	Technical Library, 4536