



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Dynamic Motion Saliency For Theft Detection and Alert

Harsha Varthini.S¹, Mahalakshmi.T¹, Mahalaxmi. S¹, Malini.E¹, N.Radha²

B.E Student, Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy, India¹

Assistant Professor, Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy, India²

ABSTRACT: In image processing motion detection is a very important problem and using this concept a novel saliency-based video object extraction (VOE) framework is presented for theft detection. Using motion and visual saliency information extracted from the input video sequence, the background and foreground will be separated. A conditional Random Field (CRF) is applied for effective integration of the saliency induced features. A video camera is used to detect the motion and capture the image and send the captured image to the owner via email. Rather than using signals to show the light the infuse gas like chloroform is used to make the suspect unconscious. Then the frames continue to be saved and compared with the process iteration taking place once again.

KEYWORDS: Conditional random field, Motion saliency, Video object extraction, Visual saliency.

I. INTRODUCTION

The purpose is to develop a novel motion detection framework by extracting video foreground objects from video frames and sending alerts. The detection and segmentation of objects in the image sequences is very important to application areas such as human—computer interaction, content-based video coding and multi object tracking. To robustly differentiate independently moving objects in the given input video, the strategy to integrate temporal and spatial information from the video sequence is a key issue throughout the segmentation process.

If one needs to design an algorithm to automatically extract the foreground objects from a video, several tasks need to be addressed:

- 1) Unknown object category and unknown number of the object instances in a video
- 2) Complex or unexpected motion of foreground objects due to articulated parts or arbitrary poses
- 3) Ambiguous appearance between foreground and background regions due to similar color, low contrast, insufficient lighting, etc.

This paper presents a robust video object extraction (VOE) framework, which utilizes both visual and motion saliency information across video frame. The observed saliency information allows us to infer several visual and motion cues for learning foreground and background models, and a conditional random field (CRF) is applied to automatically determines the label (foreground or background) of each pixel based on the observed models. With the ability to preserve both spatial and temporal consistency, our VOE framework exhibits promising results on a variety of videos, and produces quantitatively and qualitatively satisfactory performance.

II. RELATED WORK

VOE problems can be addressed using two kinds of approaches namely, supervised and unsupervised approaches. Supervised methods require prior knowledge on the subject of interest and need to collect training data beforehand for designing the associated VOE algorithms. Another type of supervised method requires user interaction for annotating candidate foreground regions. A conditional random field (CRF) is applied for maximizing a joint



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

probability of color, motion, etc. models to predict the label of each image pixel. Although the color features can be automatically determined from the input video, these methods still need the user to train object detectors for extracting shape or motion features. Some researchers proposed to use some preliminary strokes to manually select the foreground and background regions, and they utilized such information to train local classifiers to detect the foreground objects. Even though these works produce promising results, it might not be practical for users to manually annotate a large amount of video data.

On the other hand, unsupervised approaches do not train any specific object detectors or classifiers in advance. For videos captured by a static camera, extraction of foreground objects can be treated as a background subtraction problem. In other words, foreground objects can be detected simply by subtracting the current frame from a video sequence.

The learning algorithm [6][9], which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers and a method for combining increasingly more complex classifiers in a “cascade” which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions were already implemented.

III. BACKGROUND SUBTRATION

In video surveillance systems, stationary cameras are typically used to monitor activities at outdoor or indoor sites. Since the cameras are stationary, the detection of moving objects can be achieved by comparing each new frame with a representation of the scene background. This process is called background subtraction [1][4][10] and the scene representation is called the background model. Results from background subtraction are used for further processing, such as tracking targets and understanding events.

Typically, in outdoor environments with moving trees and bushes, the background of the scene is not completely static. For example, one pixel can be the image of the sky at one frame, a tree leaf at another frame, a tree branch on a third frame and some mixture subsequently; in each situation the pixel will have a different intensity (color).

This research focuses on how to construct a statistical representation of the scene background that supports sensitive detection of moving objects in the scene. However in motion tracking one can use observed motion to learn patterns of activity in a site. Motion segmentation is based on an adaptive background subtraction method that models each pixel as a mixture of Gaussians [2][3][9] and uses an online approximation to update the model. The Gaussian distributions are then evaluated to determine which are most likely to result from a background process. The methods such as a mixture of Gaussians and kernel density estimation suffer from either the lack of flexibility, by fixing or limiting the number of Gaussian components in the mixture, or large memory requirement, by maintaining a non-parametric representation of the density. These problems are aggravated in real-time computer vision applications since density functions are required to be updated as new data becomes available.

A novel kernel density approximation technique [3] based on the mean-shift mode finding algorithm, and describe an efficient method to sequentially propagate the density modes over time. It is memory efficient and it inherits the flexibility of non-parametric methods by allowing the number of components to be variable. It is applied to on-line target appearance modeling for visual tracking, and its performance is demonstrated on a variety of videos. Further it is possible to determine whether a foreground region contains multiple people and can segment the region into its constituent people and track them [5] and it is also possible to determine whether people are carrying objects, and can segment objects from their silhouettes, and construct appearance models for them so they can be identified in subsequent frames.

Codebook (CB) background subtraction algorithm was intended to sample values over long times, without making parametric assumptions. Mixed backgrounds can be modeled by multiple codewords [10]. The speed of this algorithm is to relocate the most recently updated codeword to the front of the codebook list. Most of the time, the matched codeword was the first codeword thus relocated, making the matching step efficient



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 3, March 2017

IV. AUTOMATIC OBJECT EXTRACTION

The existing unsupervised VOE approaches assume the foreground objects as outliers in terms of the observed motion information, so that the induced appearance, color, etc. features are utilized for distinguishing between foreground and background regions. In this work, a saliency-based VOE framework learns saliency information in both spatial (visual) and temporal (motion) domains. By advancing conditional random fields (CRF), the integration of the resulting features can automatically identify the foreground object without the need to treat either foreground or background as outliers.

A. Determination Of Visual Saliency:

To extract visual saliency of each frame, image segmentation[7] is performed on each video frame and color and contrast information are extracted[8]. The resulting image segments (superpixels) are applied to perform saliency detection. Turbopixels are used to produce edge preserving superpixels with similar sizes, which would achieve improved

visual saliency results. For the k th superpixel r_k , calculate its saliency score $S(r_k)$ under the condition $r_k \neq r_i$ as follows:

$$S(r_k) = \sum \exp(D_s(r_k, r_i) / \sigma_s^2) \omega(r_i) D_r(r_k, r_i), \quad (1)$$

where D_s is the Euclidean distance between the centroid of r_k and that of its surrounding superpixels r_i , while σ_s controls the width of the kernel. The parameter $\omega(r_i)$ is the weight of the neighbor superpixel r_i , which is proportional to the number of pixels in r_i . Compared to, $\omega(r_i)$ can be treated as a constant for all superpixels due to the use of Turbopixels (with similar sizes). The last term $D_r(r_k, r_i)$ measures the color difference between r_k and r_i , which is also in terms of Euclidean distance.

Consider the pixel i as a salient point, if its saliency score satisfies $S(i) > 0.8 * \max(S)$, and the collection of the resulting salient pixels will be considered as a salient point set. Since image pixels which are closer to this salient point set should be visually more significant than those which are farther away, further the saliency $S'(i)$ for each pixel i can be determined as follows:

$$S'(i) = S(i) * (1 - \text{dist}(i) / \text{distmax}), \quad (2)$$

B. Determination Of Motion Saliency

Our proposed framework aims at extracting motion salient regions based on the retrieved optical flow information. A moving pixel q_t at frame t is determined by:

$$q_t = q'_{t,t-1} \cup q'_{t,t+1}, \quad (3)$$

where q' denotes the pixel pair detected by forward or backward optical flow propagation. Do not ignore the frames which result in a large number of moving pixels at this stage as [13, 14] did, and thus our setting is more practical for realworld videos captured by freely-moving cameras.

V. SYSTEM OVERVIEW

A. System Architecture:

Kernel Density is used and the background is subtracted. Optimization is done in One Dimension only. The SVM is trained using background probability vectors and binary labels annotated from validation frames. A background probability vector for each pixel is computed based on the modeled backgrounds, and the trained SVM is applied to the vector for classification.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

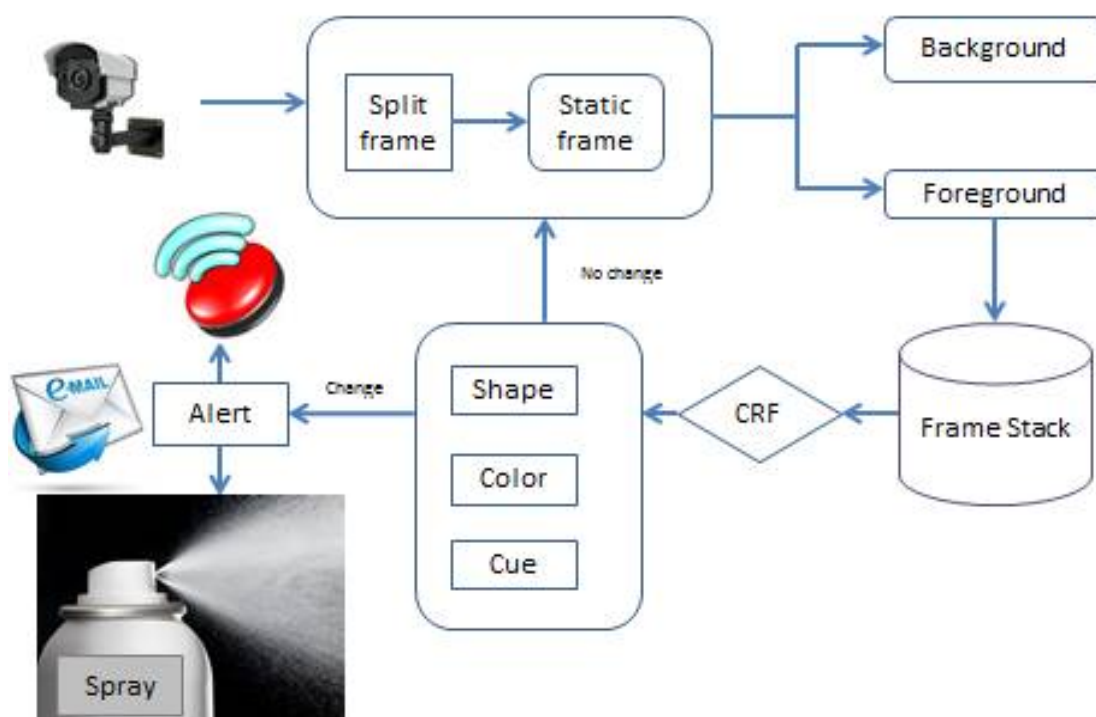


Fig 1. System architecture

The work of the system is distinguished by three key contributions:

- 1) The introduction of a new image representation called the "integral image" which allows the features used by our detector to be computed very quickly.
- 2) A learning algorithm, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers.
- 3) A method for combining increasingly more complex classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

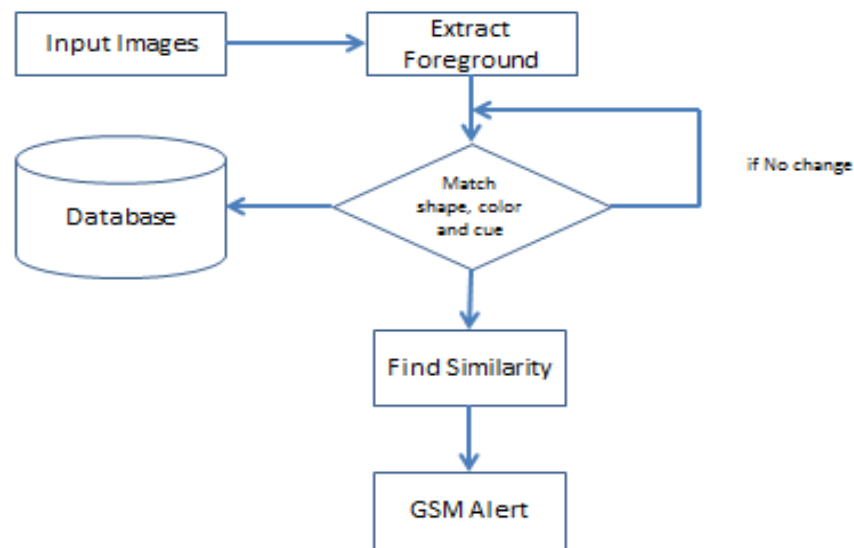


Fig 2. FlowChart

VI.SIMULATION AND RESULTS

The Simulation studies involve initiating the camera and it is set in motion detection mode. Various video sequences under different environment had been given as the input to the system and it is tested. Whenever any motion is detected the image will be captured and an alert message is sent via email to the corresponding end user. If something seems to be suspicious then infuse gas such as chloroform will be sprayed to make the suspect unconscious. In order to initiate the spray a motor will be used which will start and stop its operation when positive or negative polarity is given to the relay. When positive polarity is given motor will rotate in forward manner and when negative polarity is given then it will rotate in reverse manner. The stepdown transformer has been used to reduce the voltage from 230V to 12V. Thus the simulation has been performed.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

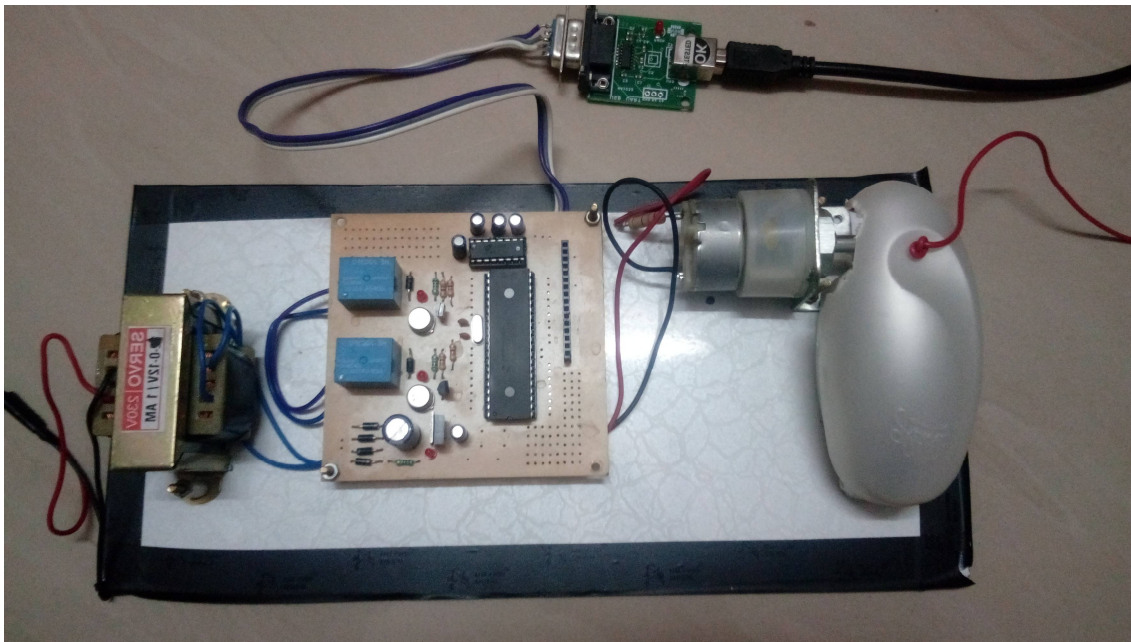


Fig 3. Breadboard connection

VII. CONCLUSION AND FUTURE WORK

In this paper, propose a multiple feature integration algorithm for background modeling and subtraction, where the background is modeled with a generative method and background and foreground are classified by a discriminative technique. KDA is used to represent a probability density function of the background for RGB, gradient, and haar like features in each pixel, where 1D independent density functions are used for simplicity. For classification, an SVM based on the probability vectors for the given feature set is employed. Our algorithm demonstrates better performance than other density-based techniques such as GMM and KDE, and the performance is tested quantitatively and qualitatively using a variety of indoor and outdoor videos. In future web services may be used for the implementation of the project which will enhance the usage in IP address like scenarios where videos over the internet can be matched for the motion. This will prevent further load on the system and cause few overheads.

REFERENCES

- [1] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," Proc. European Conf. on Computer Vision, Dublin, Vol.2, pp. 751-767, June 2000.
- [2] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. Pattern Anal. Machine Intell., Vol.22, Issue 8, pp. 747-757, 2000.
- [3] B. Han, D. Comaniciu, and L. Davis, "Sequential kernel density approximation through mode propagation: Applications to background modeling," Asian Conference on Computer Vision, Jeju Island, Korea, 2004.
- [4] Y. Ma and G. Qian, Eds., "Intelligent Video Surveillance Systems and Technology" chapter 4: Adaptive background modeling and subtraction: a density-based approach with multiple features, CRC Press, pp. 79-103, 2010.
- [5] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Real-time surveillance of people and their activities," IEEE Trans. Pattern Anal. Machine Intell., Vol. 22, Issue 8, pp. 809-830, 2000.
- [6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, HI, pp. 511-518, 2001.
- [7] https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Criminisi_eccv2008.pdf
- [8] <https://www.codeproject.com/Articles/33838/Image-Processing-using-C>
- [9] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for task of background subtraction," Pattern Recognition Letters, vol. 27, Issue 7, pp. 773-780, 2006.
- [10] K. Kim, T.H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," Real-Time Imaging, vol. 11, Issue 3, pp. 172-185, 2005.