Cognitively Motivated Habituation for Novelty Detection in Video

Vishal S. Vaingankar, Vineet S. Chaoji, Roger S. Gaborski, Ankur M. Teredesai

Laboratory for Applied Computing,
Rochester Institute of Technology,
102 Lomb Memorial Drive, Rochester, NY 14623
{vsv8846, vsc2002, rsg, amt}@cs.rit.edu

Extended Abstract

In this paper we describe the learning component of a framework for novelty detection in video data. The motivation for this learning component is derived from biologically inspired theories such as habituation. We refer to this novelty detection framework as VENUS (Video Exploitation and Novelty Understanding in Streams). Habituation is well known as an effect by which a system ceases to respond after repeated presentations of the same stimulus (Siddle, Kuiack and Kroese 1983). Work by Paul Crook (2000) implements a neural network to habituate to the surrounding elements and classifies stimulus patterns as familiar or novel. Computational modeling of habituation has been proposed by Wang (1995) and recently applied in mobile robots by Marsland, Nehmzow and Shapiro (1999). It is generally perceived that we possess a sense of distinction between what is normal and abnormal about the environment based on our prior experience.

Any aspect of the scene that does not fit into this definition of normalcy tends to be classified as a novel event. Previous work in this area models habituation as an exponential function that describes the short-term and long-term memory aspects of learning. The novelty detection in the proposed VENUS framework differs from the past work in this area as follows: first, the system uses intensity contrast, edge, color and motion information as the low-level features motivated by the human focus of attention theory instead of the commonly used high-level knowledge of the objects. As opposed to systems that store codebook information (Stauffer and Grimson 2000) about the tracked objects, this system learns patterns of activity in the scene from the extracted low-level features. This makes VENUS environment independent. Second, we use the habituation theory for modeling learning of repeated events and utilize it to distinguish between truly novel and normally seen events. Finally, we use a relaxed version of k-means for clustering events to detect novelty. The system classifies any activity in the videos as an event, such as people walking, cars entering and leaving an area. These events are classified as novel events if they have not been witnessed before in the scene. The paper presents results on natural video streams.

The video sequences are processed in still and motion saliency channels. The still saliency channel processes every frame individually and generates topographical saliency maps. Consider the situation where someone places an object in an empty corridor and leaves. The still saliency channel detects this object as a salient item. Since this object was not part of the original scene, the introduction of the object fires a novel event, which is a feature of the still novelty detection module. The motion saliency channel detects the salient moving objects of the scene, such as the motion of the person who brought the object in the corridor. Further details about the low level feature saliency can be found in the work by Gaborski et. al.[annie]. These low level features are used as inputs to the novelty learning and detection module. The foundation of novelty detection derives its source from outlier detection, an area of active research within the statistical learning community. A comprehensive survey for novelty detection using statistical methods is provided by Markou and Singh (2003). With the increasing emphasis on surveillance and monitoring applications, there is a considerable interest in novelty detection within both the computer vision and machine learning community. Yamanishi and Takeuchi (2002) discuss novelty detection for non-stationary (adaptive) data while incorporating the effect of forgetting previous data. Oh, Lee and Kote (2003) discuss an algorithm for detecting motion on segments of frames and clustering these segments to identify normal events. Accordingly, anything not within these clusters forms an abnormal event. Their system lacks a comprehensive learning component for novelty detection, which is the focus of our work.

Novelty Detection using learning

The novelty detection framework in VENUS has two modules: the still novelty detection module and the motion novelty detection module. On presentation of a never-seen-before feature value, the learning and novelty detection module classifies it as being novel. If the same feature value is observed repeatedly over time, the system habituates to it and stops flagging it as a novel event. An event can be novel by virtue of any of its low-level features or a combination of them. After a certain amount of time without additional occurrences of the same event, the system recovers its original sensitivity for the feature, i.e. the habituation effect decreases. This concept is based on Kohonen's theory of novelty detection filters with a forgetting effect (1988). The theory states that the system can only memorize patterns when it is frequently exposed to them. The memorized pattern tends to be forgotten if it is not reinforced repeatedly over time. The forgetting term is similar to the dis-habituation effect described by Wang (1995). Novelty detection and learning in this system is region based where a region is an 8-by-8 pixel area of the image. Each frame of the video is thus divided into several such regions. The regions that detect motion get excited.

The working of the system could be explained with the following example: Consider a video sequence in which people are walking from right to left at a speed of 5 mph. When a person passes over a region (during a group of contiguous frames), the left directional motion map gets invoked. The excited regions of the map provide motion values that correspond to the speed of the person walking. The value obtained over time, for each region, from the directional maps, can be represented by a Gaussian mixture model. Each excited region forms a single cluster that represents a single Gaussian distribution. Novelty detection in our case involves identifying clusters having different distribution for their motion values. Every previously unseen distribution forms a novel cluster that is inserted into a pool of clusters for that region. A recently created cluster is compared with all the existing clusters in the pool to see if it can be merged with it. Clusters with similar distribution are merged indicating a similar event has occurred. The similarity measure between two clusters is a function of theirs means and standard deviations. For example, when multiple people walk at 5mph over a region, clusters representing this speed are merged. This indicates that people walking is not a novel event any more. Now, when a person runs at 15 mph from right to left, the same excited regions that had previously formed a cluster representing 5 mph would now form a new cluster for 15 mph. Since the cluster corresponding to 15 mph is not similar to an existing cluster it starts a new cluster. This represents occurrence of a novel event. Similarly the above phenomenon will be observed if a person walks from left to right, thereby firing an event in the right directional map. This algorithm is incremental in nature, in that the clusters for a region are updated as events occur in the scene. The algorithm does not limit the number of clusters per region since the number of novel event cannot be predicted ahead of time.

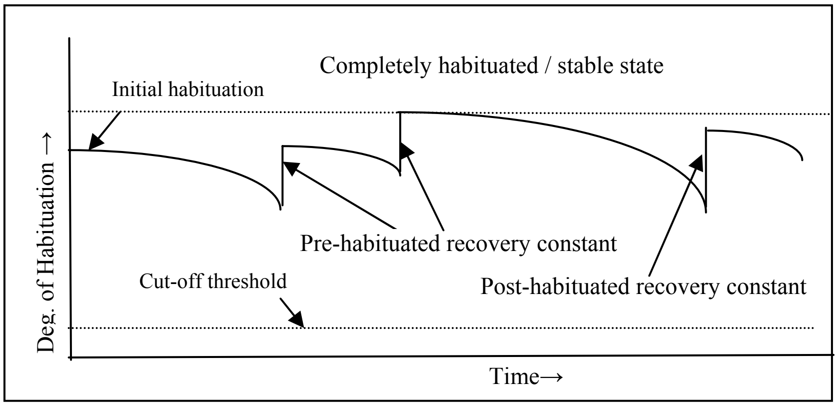

Figure: Degree of Habituation curve of a cluster versus time

New clusters added to the pool are assigned an initial habituation value and an initial decay rate that determine its temporal characteristics. The decay rate symbolizes the forgetting term described by Kohonen (1988). The slower the decay rate the longer is the retention period for the event. Each cluster follows a sigmoidal-like habituation curve as shown in the figure. The habituation function for a cluster is given by:
$H(t) = 1 - [1/(1 + e^{-at})]$ where *H(t)* is the habituation value *t* frames after the creation of the cluster and *a* is the current decay rate of the cluster. When clusters are merged we update the decay rate for the older cluster. This indicates that the learnt event was reinforced resulting in increased retention. A cluster with habituation value below the cut-off threshold is considered completely decayed and is discarded from the pool of clusters. Effectively, the system has forgotten the event that the discarded cluster represented. Hence the forgotten event becomes novel once again. This models the concept of forgetting in habituation theory. The initial decay rate is set to zero which can go up to one. Value of zero indicates no decay (longer retention) while one indicates maximum decay (shorter retention). The decay rate for a cluster is adjusted as follows:
$a_t = 1 - [e/f]$ where a_t is the decay rate *t* frames after its creation, *f* is the number of frames passed since the creation of the cluster and *e* is the number of times the cluster merged with similar clusters. The *e/f* term indicates the reinforcement (cluster merging) rate. Higher the reinforcement rate, closer the new decay rate to zero. Smaller the reinforcement rate, closer the new decay rate will be to one. As per habituation theory, an event is not instantaneously learnt. It takes some number of occurrences before a system gets completely habituated. The recovery in degree of habituation prior to the system reaching complete habituation (also known as stable state) is lesser than the recovery after reaching complete habituation as seen in the figure. Novelty is inversely related to the degree of habituation the cluster has attained. Higher the habituation value, the lower is its features novelty and vice versa. Experimental results on natural scenes validate our learning approach. Due to lack of space the results are not included. The result files can be downloaded from http://www.cs.rit.edu/~amt/vision/

References

Crook, P. 2000. Spotting Novelty: A Neural Network Model for Familiarity Discrimination.

Gaborski, R.; Vaingankar V.S.; and Canosa, R.L. 2003. Goal Directed Visual Search Based on Color Cues: Co-operative Effects of Top-down & Bottom-up Visual Attention. In *Proceedings of the Artificial Neural Networks in Engineering, Rolla, Missouri.*

Haering, N. C., Qian, R.J., and Sezan, M. I. 1999. A Semantic Event Detection Approach and Its Application to Detecting Hunts in Wildlife Video. In *IEEE Transactions on Circuits and Systems for Video technology*, 10:857—868.

Kohonen, T. eds. 1988. Self-Organization and Associative Memory. New York: Springer-Verlag.

Markou, M.; and Singh, S. 2003. Novelty Detection: A Review, Part I: Statistical Approaches, *Signal Processing*. Forthcoming.

Marsland, S.; Nehmzow, U.; and Shapiro, J. 1999. A model of habituation applied to mobile robots. In *Proceedings of Towards Intelligent Mobile Robots*. Bristol, UK.

Medioni,G.; Cohen, I.; Brmond, F.; Hongeng, S.; and Nevatia R. 2001. Event Detection and Analysis from Video Streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23: 8, 873-889.

Oh, J; Lee, J.; and Kote S. 2003. Real Time Video Data Mining for Surveillance Video Streams. In *Proceedings of the Seventh Pacific-Asia Conference on Knowledge Discovery and Data Mining,* 222-233. Seoul, Korea: Springer-Verlag

Siddle, D. A. T.; Kuiack, M.; and Kroese, S. B. 1983. The Orienting reflex. (pp. 149-170). In *Physiological Correlates of Human Behaviour.* Edited by Gale A. and Edwards, J. Academic Press: London.

Stauffer, C.; and Grimson, E. 2000. Learning Patterns of Activity Using Real-Time Tracking. In *IEEE Transactions* on *Pattern Analysis and Machine Intelligence.* 22(8):747-757.

Wang D.L. 1995. Habituation. Arbib M.A. (ed.), *The Handbook of Brain Theory and Neural Networks.* 441-444, MIT Press.

Yamanishi, K., and Takeuchi, J. 2002. A unifying framework for detecting outliers and change points from non-stationary time series data. In *Proc. of the Eighth ACM SIGKDD, ACM Press.* 676-681.