2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: 10.13031/aim.201700300
Paper Number: 1700300

Bird Detection, Tracking and Counting in Wine Grapes

Santosh Bhusal*, Shivam Goel**, Kapli Khanal*, Matthew Taylor **, Manoj Karkee*
****Department of Biological System Engineering, Center for Precision and Automated Agricultural Systems **School of Electrical Engineering and Computer Science***
Washington State University

Written for presentation at the
2017 ASABE Annual International Meeting
Sponsored by ASABE
Spokane, Washington
July 16-19, 2017

ABSTRACT. ***Bird damage in fruit crops is a critical problem in wine grapes, blueberries and other fruit crops especially during the weeks close to the harvesting period. Usually small birds such as Starlings, Robins and Finches feed extensively on wine grapes. Automated detection, localization, and tracking of these birds in the field will be necessary to identify best locations for installing bird scaring devices in the field as well as to use autonomous UAS operation to deter them. A section of wine grape plot (~30 m x 30 m) was constantly monitored using four GoPro cameras installed at the four corners of the plot. Videos were recorded at 1080p resolution with 30 frames per second. In this paper, Gaussian mixture-based Background/Foreground Segmentation Algorithm was used in detecting birds flying in and out of the wine grape plot. This algorithm can detect moving objects in a video irrespective of their shape, size and color. Detected birds were tracked over a period of time using Kalman filter. Then, a field boundary was defined to estimate the count of the birds flying in and out of the plot through the boundary. Two performance measures, precision and recall (sensitivity), were used to analyze the accuracy of the counting method. Precision refers to the usefulness of the system and recall measures its completeness. Results showed that the proposed method can achieve a precision of 85% in counting birds entering or leaving a crop field with a sensitivity of 87%. Such a system could have a wide range of applications when birds' presence is a problem such as in crop fields, airport and cattle farms.***

***Keywords.* Bird Detection, Bird Deterrence, Bird Tracking, Bird Counting, Vineyards**

Introduction

Bird damage in fruit crops such as wine grapes, blueberries, cherries and apples has been increasingly challenging issue for growers as new varieties have been more attractive to birds such as Robins, Starlings and Finches. In Washington State alone, fruit crop growers lost more than 80 million dollars to bird damage in 2013 (Anderson et al., 2013). Farmers producing a wide range of crops are in need of an effective and affordable system for deterring birds as the current techniques such as netting, chemical repellants, auditory and visual scaring devices, have been ineffective or costly for successful adoption in the fields. The long term goal of this this study is to develop a fully autonomous bird deterrence system for fruit crops such as wine grapes using Unmanned Aerial Systems (UASs). Automated bird detection and tracking is one of the major tasks towards the development of an autonomous bird deterrence system. In addition, automated bird detection and monitoring can be beneficial for numerous other applications such as understanding ecological factors/phenomenon, studying avian related diseases and understanding migratory behavior of birds**.** Bird detection has also been implemented to prevent accidents related to bird collision with wind turbines (Yoshihashi, Kawakami, Iida, & Naemura, 2015; J. Zhang, Xu, Cao, Yan, & Li, 2014).

The most commonly used features in bird detection are their shape features (Verstraeten et al., 2010) or motion features (Verstraeten et al., 2010; Shakeri, 2010). Other methods include template matching and model based filtering by incorporating the flight dynamics (Song & Xu, 2008). Large birds such as hawks and crows occupy reasonable space in an image (when acquired with industrial cameras available currently in the market from a closer distance) suitable for extracting complex features including appearance and shape using techniques such as Haar wavelets and Histogram of Oriented Gradient (HOG) descriptors. Those features can be used to train image classifiers such as Adaboost (Yoshihashi et al., 2015). In recent years, deep convolution neural networks (CNN) has been extensively used for object detections because of their robustness and accuracy in classifying pixels or objects into multiple groups. In a study conducted by (Takeki et al., 2016), a CNN outperformed the Haar and AdaBoost based classifiers in identification of small flying birds in an outdoor environment. Combining the CNN based deep residual convolution network with semantic segmentation using support vector machines, Takeki et al. (2016) achieved even better results in detecting birds with varying sizes. However, the sensitivity of their algorithm is relatively poor for tiny birds (bird size less than $\leq 15\times 15$ pixels).

Motion detection methods can be used to track moving objectives like birds, which generally consist of a suitable model for static objects in the videos called background objects. Background and consecutive frame subtraction (BGCFS) described in (Wei et al., 2014) modeled first 10 frame sequences with the immediate previous frame to form a model of the background, which is updated continuously over time. Gaussian mixture-based Background/Foreground Segmentation Algorithm (GMM) (Stauffer, 1999) has also commonly been used in detecting flying birds (Shakeri, 2010). This algorithm can detect moving objects in a video irrespective of their shape, size and color. Object detection in a video is independent of consecutive frames and tracking method maps the associated detections together (Bhusal, 2015). This method also provides a benefit of predicting future locations or estimating object locations in occluded environment. Most commonly used approach for tracking birds are probabilistic methods like Markov Chain Monte Carlo (MCMC) (Zhang et al., 2008) or Kalman filter as described for pedestrian detection in (Bhusal, 2015; Li et al.,2010)**.** MCMC involving large set of parameters, which is computationally expensive in the large frame size videos.

Most of the research reported in bird detection and tracking are related to open sky surveillance where sky is the predominant background in the images acquired. Targeted birds in the images have sharp shape and color features, even though shape varies with the flying actions of the birds. Berries in the vineyards are covered by the leaves and other parts of the canopies when looked at from the top, but are visible when looked at from bottom of the canopies. Canopy height of wine grape vines is generally less than or about 10 feet and therefore birds generally fly close to the ground to enter to a plot to eat fruits. To cover bird activities close to ground, the camera field of view needs to be directed towards the ground resulting in much complex background with the presence of vines, surrounding trees, bushes and the road around field boundaries (e.g. Figure 3). In the vineyard and other crop fields, frame rate to be used to record bird activities should be able to capture their high speed movement, resulting in a large amount of motion blurriness around bird region in the frames. Feature based detection described above will be challenging in such environment as the shape and size features are changing and shape and color features are heavily distorted by motion blurriness. To the best of our knowledge, automated video surveillance system for birds with agriculture application has not been reported till date. In this paper, a commonly used method for people and vehicle tracking is used for bird detection, tracking and counting in a vineyard environment with noisy background condition. In the following section we have discussed our approach for bird detection, tracking and counting.

Methods

As discussed before, detection and tracking birds and then estimating their number in the field is a crucial first step for developing autonomous or automated bird deterrence (Figure 1) systems. In the following sub-sections, these individual steps are discussed in details.

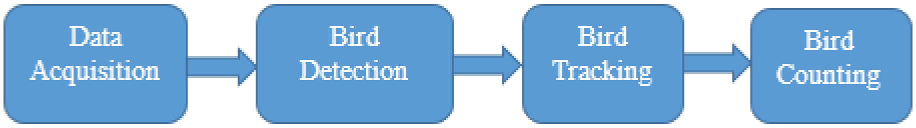

Figure 1: Flow chart of the proposed bird detection, tracking and counting system

Data Collection

The field experiment was carried out in a commercial vineyard in Prosser, WA (Olsen Bros Ranches Inc., Prosser, WA). For monitoring bird activities, a plot of approximately 30 m x 30 m was chosen in the vineyard close to the canyon where bird's activities were relatively high. Videos were recorded at1080p as the birds were relatively small and generally far away from the cameras. Four GoPro Hero 4 cameras were installed at the four corners of the field. Videos were recorded at 30 frames per second in the morning hours from 7:00 am to 9:00 am from September 20 to 25 in the year of 2016.

Figure 2: Google map view of the experimental plot selected in a commercial vineyard in Prosser, WA. Each side of the selected plot is roughly 30 m. Camera location is shown by the camera icons and the triangular regions represent tentative field of view for each camera. The cameras were used in wide angle view while recording the videos.

Bird Detection

Starlings, Robins and Finches are some of the birds most commonly feeding in wine grapes and other fruit crops. Depending on the size and location of these birds relative to the camera, flying birds appeared to have up to few hundreds of pixels in each image frame. Shape and size of these objects (birds) may vary substantially in successive frames as their shape and location change during their flight. Because these birds are quite fast and they change flight direction very quickly, their positions in the image can change widely between successive frames. As mentioned in introduction section, the shape features were heavily distorted by motion blurriness and therefore motion detection was used for detecting birds.

An adaptive background subtraction (Stauffer, 1999) method was used to detect those birds. This algorithm can detect moving objects in a video irrespective of their shape, size and color. Each pixel in a frame was modeled using number of Gaussians to be a part of stationary background objects. Pixels that don't fit within any of the background models were classified as the foreground moving objects. These moving pixels form a blob in the foreground binary image. We posed a size constraint to remove the noisy detections. Detected blobs with a size of 100 to 2000 pixels were considered as the potential areas of birds in the images.

Bird Tracking and Counting

Kalman filter with constant velocity (acceleration=0) model was used for tracking birds in the videos acquired from the vineyard plot. Six different parameters (position, width and height from the detector, and horizontal and vertical velocities of the birds) $[x \quad y \quad v_x \quad v_y \quad w \quad h]^T$ were used as the state variable for the Kalman filter while the information obtained from the detector bounding box $[x \quad y \quad w \quad h]^T$ was used to update the state variables over time. Detected birds were tracked using Kalman filter in a similar way as discussed (Li et al., 2010) to maintain their spatial relation with detection. Since, there can be more than one bird in each frame, a multi-target tracking approach was needed A tracking class was defined to maintain the state of flying birds (The MathWorks, Inc. 2017) for multi target tracking. Each objects of the tracking class or the tracks represented a flying bird in the video. When multiple detection occurred, a vector array of bird tracks was created within this class. New tracks were initiated each time when a new detection occurred and were updated in their successive detections. A birds may lose its detection because of occlusion, imposed size constraint, being permanently crossing the field of view (FOV) of the camera or hiding inside the canopies. To overcome the first two issues, any tracks that are no longer detected in three consecutive frames were permanently removed from the vector array list.

The cost of associating j^{th} detection with i^{th} track was computed using equation (1) c. As the occurrence of heavily size differing detection and tracks were less frequent more weight was given to distance measure ($\alpha = 0.8$). By incorporating the area difference between the detected and estimated (using tracking) bounding boxes representing two different birds, the chances that a detected bird (say x) being associating to the track of another bird (say y) can be minimized assuming there was a reasonable difference in their bounding box size. In additional to this, a positional constraint was introduced by limiting the distance between the detected location and estimated location (tracked location) to be within 100 pixels to avoid the association of distant detection and an existing track. An imaginary boundary was defined on the experimental plot by a line close to the center of the camera field of view. The tracking class of a bird was updated when a bird (its trajectory) crossed the boundary from one side to another, to determine its status as incoming bird or outgoing bird.

$$cost(i,j) = \frac{\alpha D(i,j)}{D_{i_{min}}} + \frac{(1-\alpha)A(i,j)}{A_{i_{min}}} \quad (1)$$

Where, α is the weight, $D(i,j)$ is Euclidian distance measure between the centroid of j^{th} detection with i^{th} track, $A(i,j)$ is difference in area of the bounding boxes between those two detection or tracks. $D_{i_{min}}$ is the perpendicular distance between the i^{th} track (or trajectory) and a detected location. $A_{i_{min}}$ is the minimum difference in area for the i^{th} track while comparing with all detections. The main idea of dividing the associating metrics by the minimum value was to avoid the need of thresholding parameter for cost estimation because of large variation in bird movements in successive frames. With this normalization, the cost function will be 1 for identical association. Figure 3 shows a single frame of the view as an example of how the data was analyzed. Complete process of bird detection, tracking and counting was implemented in C++ environment. OpenCV-3.1.0 library was used for implementing various computer vision algorithms in a hardware system consisting of an Intel Core i7 CPU running at 3.40GHz clock speed.

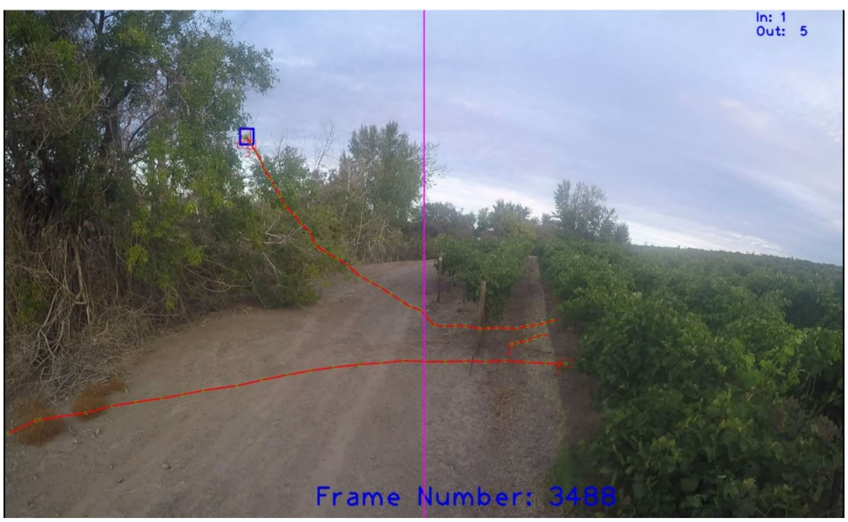

Figure 3: Bird Detection, tracking and counting in a video acquired from camera 2 Field of View (FOV). The straight line in the middle of the image was used as the boundary line to determine if a given bird was incoming into or outgoing from the field. A bird was assumed to be incoming if it was detected at the left side of the line and kept moving towards the right side. The end of the bird path (tracking curve lines in the figure) with a rectangle (either blue or red) is the location of the bird in this frame and the other end was the location where it was detected the first. All the green dots in the middle were the estimated or detected location of the same bird. The blue rectangle was used to represent that the birds detected in the current frame while red rectangle was used to show if the detection was lost and tracking was able to estimate the position of the bird in the current frame. On the top right corner of the image shows the incoming and outgoing bird count from the beginning of the algorithm. The frame shows the tracks of 3 different birds.

Accuracy Assessment

As mentioned in the methods section, birds flying in and out of the vineyard plot covered few pixels to few hundred pixels in the video frames. These pixels did not have any contrasting color with the background objects. The motion detection algorithm that was used is very sensitive to change in intensity at pixel level and can detect birds generally more accurately than human vision. Some of the locations in video frames where the proposed algorithm detected birds were challenging for human vision to verify. However, continuous observation over a few consecutive frames helped human vision to actually detect birds. To assess the accuracy of the bird counting method in terms of incoming and outgoing birds, videos were observed over a number of frames manually to verify if there were any bird activities in the given section of the videos.

Five video sections were used for evaluating the performance of the bird tracking and counting algorithm based on the population density (no of birds per second) of birds. High density section have more than 1 birds/sec (considering the true count), the medium density level was classified as the section having at least 1 bird within 10 sec and other sparsely populated regions are classified as low density section. Nearly 65% of the birds in study were from high density section, 25% from medium density section and rest from low density section. Four sections of the videos were collected from one camera viewing angle and one video section was collected from a different camera viewing angle to cover varying level of bird activities (Table 1). More than 100 different incoming and outgoing bird activities were verified manually. More than 8,700 frames were manually monitored to assess the performance of the algorithm. When both the algorithm and human annotator can verify the boundary line crossing, it was counted as a true count (true positive). If a certain bird lost its tracking information while crossing the boundary it was classified as missed count (false negative) whereas if a non-bird (false) track crosses the boundary line, it was marked as false count (false positive). Some of the detections that cannot be verified by human annotator even after looking into 6 frames were counted as false detection (might be a butterfly or an insect).

Table 1: Data Used for Count Assessment

Video no	Date of video capture	Camera View	Bird population density	No of frames	System incoming bird count	System outgoing bird count
1	9/20/16	2	High	900	31	16
2	9/23/16	2	Medium	1500	12	3
3	9/23/16	3	Low	3600	5	9
4	9/21/16	2	High	900	31	17
5	9/25/16	2	Medium	1800	10	1
		Total		8700	89	46

Two other measures were calculated from the results obtained to evaluate the performance of the counting system. Both incoming and outgoing incidents were summed together to compute those measures. Precision is the measure of relevancy of the system (Equation 2) while recall measures the sensitivity of the system (Equation 3). Higher value of precision and recall means that the system is counting majority of the birds with higher accuracy.

$$precision = \frac{Overall\ true\ count}{overall\ true\ count + overall\ false\ count} \tag{2}$$

$$recall = \frac{overall\ true\ count}{overall\ true\ count + overall\ missed\ count} \tag{3}$$

Results and Discussion

The performance of the bird counting system is shown in the bar diagram in Figure 4. The bar diagram shows the incoming and outgoing bird activities for all the videos used in this experiment. Performance of the bird detection method was assessed in terms of true detection, false detection and missed detection. Most of the videos were from camera 2 FOV. Camera 2 was located next to the tall trees which provided shelter for birds (Figure 3). As more birds were noticed in this side than any other sides, videos from this camera view were classified as high or medium bird population density (Table 1). The incoming and outgoing bird count achieved by the purposed method was also higher for the video collected from this FOV (Table 1). Camera 3 FOV was relatively in open area, thus limiting rapid to and fro movement of birds between the trees and canopies.

Table 2: Bird Counting Performance Measure Based on Varying Bird Population Density

Measure	High Density	Medium Density	Low Density
Precision (%)	81	94	93
Recall (%)	89	80	100

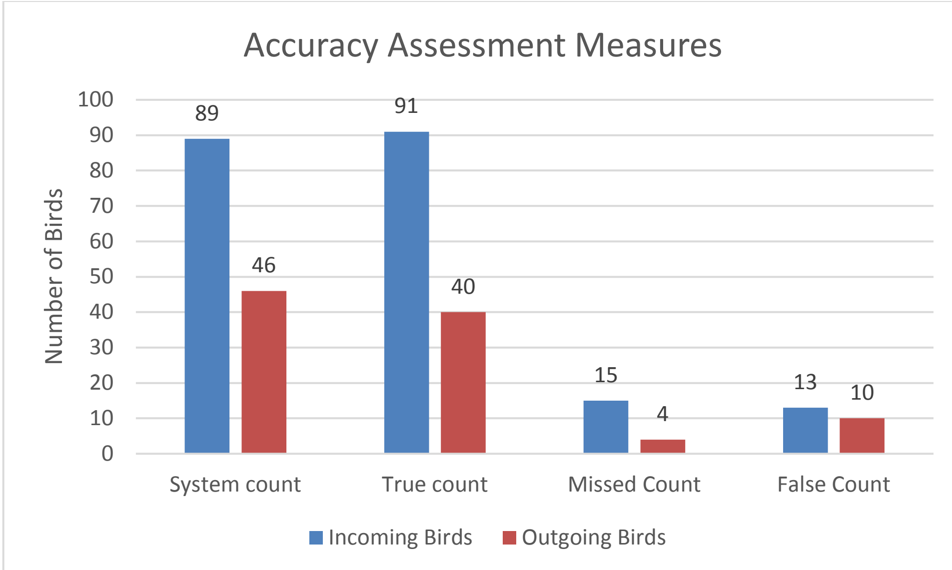

Figure 4: Bar diagram representation of birds counts estimated by the proposed method ("System count") and human annotator ("True count").

As mentioned in the methods section, precision and recall were used to measure the performance of the counting system. Overall, the results showed that the proposed method can achieve a precision of 85% in counting birds entering or leaving a crop field with a sensitivity of 87%. Table 2 shows the performance measure of the system at three level of bird population density; high, medium and low. The main reason for missed count was the size and position constraint introduced in the detection and tracking phase. When a bird is flying relatively close to the camera, its size and the position varied rapidly in successive frames. Even though the same bird was repeatedly detected for a number of times, the tracking class will initiate it as the new detection in such cases. Some of the birds were occluded in the boundary region because of the plant canopies and thus contributing to the missed count. Many birds, while flying from the nearby trees into the field, flew with their wings closed and thus resulting in high speed as well as smaller area in the frame. Such smaller areas were rejected as the false detection by the imposed size constraint. Sensitivity does not entirely depend on the detector performance (as seen by 80% recall value in medium density level) as removing the size threshold may not be enough to detect all the occluded birds.

False detection and counting occurred mostly in the highly cluttered frames. When multiple incoming and outgoing birds of similar sizes appear at the boundary region and some of them were lost by the detection algorithm, cross association could happen resulting in a false count. This problem, partly caused by the use of Euclidean distance for data association, cost with a low precision (81%) in the high density video sections. Bird tracks were very short (< 5 sec) and there were very few birds (1 bird in 10 sec) in medium and low density sections, which reduced the cross association and thus contributing in the higher precision.

Conclusions

A machine vision system was developed and evaluated for bird detection, tracking and counting in a vineyard environment. Gaussian mixture method for foreground and background segmentation method was used for bird detection as shape, and color features were heavily distorted due to higher speed of birds. Kalman filter-based multi target tracking system was integrated over the detected objects (birds) to track the trajectory of bird flights in successive frames. An imaginary boundary line was defined to identify if the trajectories of specific birds are incoming into the field or outgoing way from the field. Noisy detections of birds were filtered out by the tracking system and noisy trajectories were filtered out by the counting algorithm. The true detection 131, which led to the precision of the counting system to be 85% with a sensitivity of 87% even when most of the data used for assessment was from highly cluttered environment.

In the future, occlusion handling problems in data association will be resolved for improving the detection and tracking accuracy. Current system is relatively slower for real time application. Future efforts will also be focused on optimizing the computational speed. Graphical Processing Unit (GPU)-based parallel computing will be considered for improving the speed. Such an automated bird detection and tracking system can be used in various other applications including bird deterrence in airport and garbage disposal sites, as well is in studying and understanding bird behaviors.

References

Anderson, A., Lindell, C., Moxcey, K. M., Siemer, W., Linz, G. M., Curtis, P., . . . Steensma, K. (2013). Bird damage to select fruit crops: the cost of damage and the benefits of control in five states. *Crop Protection, 52*, 103–109.

Bhusal, S. (2015). Object Detection and Tracking in Wide Area Surveillance Using Thermal Imagery. (MS Thesis), Las Vegas NV, University of Nevada Las Vegas, Department of Electrical and Computer Engineering.

Li, X., Wang, K., Wang, W., & Li, Y. (2010). A multiple object tracking method using Kalman filter. *Information and Automation (ICIA), 2010 IEEE International Conference on*, (pp. 1862–1866). IEEE.

Shakeri, M. & Zhang H. (2010). Real-time bird detection based on background subtraction. *Intelligent Control and Automation (WCICA), 2012 10th World Congress on*, (pp. 4507–4510). IEEE.

Song, D., & Xu, Y. (2008). Monocular Vision-based Detection of a Flying Bird. *Texas A & M University*. College Station TX 77843, USA

Stauffer, C., & Grimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. (Vol. 2, pp. 246–252). IEEE.

Takeki, A., Trinh, T. T., Yoshihashi, R., Kawakami, R., Iida, M., & Naemura, T. (2016). Detection of small birds in large images by combining a deep detector with semantic segmentation. *Image Processing (ICIP), 2016 IEEE International Conference on*. (pp. 3977–3981). IEEE.

The MathWorks, Inc. 2017. Motion-Based Multiple Object Tracking. (2017). Retrieved from https://www.mathworks.com/help/vision/examples/motion-based-multiple-object-tracking.html

Verstraeten, W. W., Vermeulen, B., Stuckens, J., Lhermitte, S., Van der Zande, D., Van Ranst, M., & Coppin, P. (2010). Webcams for bird detection and monitoring: A demonstration study. *Sensors, 10*(4), 3480-3503.

Wei, L., Mirzaei, G., Majid, M. W., Jamali, M. M., Ross, J., Gorsevski, P. V., & Bingman, V. P. (2014). Birds/bats movement tracking with IR camera for wind farm applications. *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. (pp. 341–344). IEEE.

Yoshihashi, R., Kawakami, R., Iida, M., & Naemura, T. (2015). Construction of a bird image dataset for ecological investigations. *Image Processing (ICIP), 2015 IEEE International Conference on*. (pp. 4248–4252). IEEE.

Zhang, J., Xu, Q., Cao, X., Yan, P., & Li, X. (2014). Hierarchical incorporation of shape and shape dynamics for flying bird detection. *Neurocomputing, 131*, 179–190.

Zhang, X. Y., Wu, X. J., Zhou, X., Wang, X. G., & Zhang, Y. Y. (2008). Automatic detection and tracking of maneuverable birds in videos. *Computational Intelligence and Security, 2008. CIS'08. International Conference on*. (Vol. 1, pp. 185–189). IEEE.