DOI: 10.21917/ijivp.2013.0092

STRAY DOG DETECTION IN WIRED CAMERA NETWORK

C. Prashanth[1] and P. T. V. Bhuvaneswari[2]
Department of Electronics Engineering, Madras Institute of Technology, India
E-mail: [1]prash6392@gmail.com and [2]ptvb@annauniv.edu

Abstract

Existing surveillance systems impose high level of security on humans but lacks attention on animals. Stray dogs could be used as an alternative to humans to carry explosive material. It is therefore imperative to ensure the detection of stray dogs for necessary corrective action. In this paper, a novel composite approach to detect the presence of stray dogs is proposed. The captured frame from the surveillance camera is initially pre-processed using Gaussian filter to remove noise. The foreground object of interest is extracted utilizing ViBe algorithm. Histogram of Oriented Gradients (HOG) algorithm is used as the shape descriptor which derives the shape and size information of the extracted foreground object. Finally, stray dogs are classified from humans using a polynomial Support Vector Machine (SVM) of order 3. The proposed composite approach is simulated in MATLAB and OpenCV. Further it is validated with real time video feeds taken from an existing surveillance system. From the results obtained, it is found that a classification accuracy of about 96% is achieved. This encourages the utilization of the proposed composite algorithm in real time surveillance systems.

Keywords:
Surveillance System, Stray Dog Detection, Gaussian Filter, ViBe Algorithm, SVM Classifier

1. INTRODUCTION

Automatic surveillance is crucial in large and crowded environments such as airports, railway stations, bus terminals and markets. Typical surveillance systems consist of CCTV cameras and security cameras that are mounted at different locations throughout the monitored environment. The videos are manually checked continuously by human operators to detect the presence of abnormalities. Video processing techniques enable the automatic extraction of required information from the video feed directly.

Existing surveillance systems impose high level of security on humans but lacks attention on animals. Stray dogs could be used as an alternative to humans to carry explosive material. It is therefore imperative to ensure the detection of stray dogs for necessary corrective action.

In this paper, a novel composite algorithm is proposed to detect stray dogs in crowded environments. The proposed algorithm combines ViBe and Histogram of Oriented Gradients (HOG) to achieve the objective.

ViBe algorithm is a foreground extraction algorithm that extracts object of interest from the video feed [1]. HOG shape descriptor algorithm [2] is used to extract the features from the foreground object which is further classified as either human or stray dog using Support Vector Machine (SVM) classifier [10]. The proposed composite algorithm is fast and accurate as it combines the advantages of both ViBe and HOG.

The rest of the paper is organized as follows. In section 2, related works are discussed. Section 3 describes the proposed method, followed by section 4 which enumerates the results and finally section 5 concludes the paper.

2. RELATED WORKS

2.1 FOREGROUND EXTRACTION

Foreground object extraction from the input video is the first step in enabling the detection or classification of foreground objects. In surveillance systems, cameras are generally assumed to be static; implying that predominant change in frames obtained at different instants will be only due to foreground or extraneous objects that fall inside the Field Of View (FOV) of the camera. Further foreground extraction becomes significantly easier when camera motion is non-existent. Several techniques have been proposed to extract moving objects from a video. In this work, a foreground extraction called ViBe [1], proposed by Oliver Barnich and Marc Van Droogenbroeck is employed.

2.2 SHAPE DESCRIPTORS

Shape descriptors attempt to uniquely represent the shapes of objects present in the image. These shape descriptors can be organized into feature vectors which are then used for classifying object shapes. In this work, a shape descriptor known as the Histogram of Oriented Gradients (HOG) [2] is used.

Intensity gradients define edge directions vividly using which the shape of the object can be well described [2]. Initially, the input image is divided into several sub-windows spatially. These sub-windows are also referred to as cells. The local 1-D histograms of gradient directions or edge orientations are accumulated over all the pixels of each cell. The composite histograms form the final representation. To make the feature invariant to illumination, the local responses are contrast normalized. This normalized representation is known as the Histogram of Oriented Gradients (HOG). The HOG descriptor was initially proposed as pedestrian detection shape descriptor in [2]. Histogram of Oriented Gradients has also been used for several other applications.

Lei Hu et al. employs HOG for constructing a texture based background model, from which the motion of the foreground object is detected [4]. HOG have also been used to detect the orientation of vehicles for assisting autonomous vehicle systems by Paul. E. Rybski et al. [5]. Ilias Kamal has proposed the composite use of multidimensional HOG and Linear Support Vector Machines SVM for car recognition [6].

Hsin-Chun Tsai et al. have proposed a novel multilayered HOG and built a computationally efficient awareness system for happiness expression [7]. An ear recognition using multi-scale HOG has been proposed in [8] for biometric application. HOGs have also been utilized in reducing the occurrence of false positives when Harr-Like classifiers are used for license plate detection [9].

The shape descriptor is incorporated in the proposed algorithm because dogs can be distinguished from humans by their shape. Dogs tend to have a greater number of horizontal edges, running across the lengths of their bodies than human beings. On the other hand, human beings are generally characterized by more number of vertical edges along their height. Also the key feature of the HOG descriptor is that it is sensitive to rotation.

Generally shape descriptors are required to be rotation invariant because they must produce the same response to a particular shape even under different orientations. However, this property of the HOG descriptor in being rotation variant is vital to the classification procedure. This is because when the silhouette of a dog is rotated by 90 degrees to the right or left, the number of vertical edges in the resulting image increases and the shape may become closer to a human silhouette. As a result of this observation, the HOG shape descriptor could aid in the detection, classification and segregation of stray dogs from human beings.

2.3 SUPPORT VECTOR MACHINES

The most popular Support Vector Machine (SVM) classifier is used to distinguish between stray dogs and humans in this work. SVM's have been initially designed to function as binary classifiers [10]. SVM classifiers fall under the supervised learning class of machine learning algorithms. There are two phases in building a SVM classifier namely the testing phase and the training phase. Given a set of points in an n-dimensional space, the SVM classifier constructs a separating hyperplane that splits the data points into 2 classes. The Basic SVM classifier requires that the data be linearly separable in the input feature space. If the data points are not linearly separable, then these data points can be mapped to a higher dimensional space where they become linearly separable with the help of a kernel function. SVM classifiers can also be used as multiclass classifiers by employing the One vs Many and Many vs Many approaches. In this work, the classification between stray dogs and humans beings is carried out using a binary SVM classifier.

3. METHODOLOGY

In this section, the working of the proposed composite algorithm is discussed in detail. Fig.1 represents a scenario considered in this work, which consists of humans, stray dogs and Surveillance cameras.

The video feed obtained from the surveillance camera is subject to plenty of random pixel variations with time. These pixel variations are manifestations of the noise caused by the camera itself and not due to the presence of foreground objects. Generally, in surveillance systems the random variation in the brightness intensities can be misinterpreted as moving objects. Hence to mitigate this problem, the frames read from the camera are first convolved with a Gaussian mask and then smoothened.

The process of smoothening helps in smudging the pixel intensities in the spatial domain which reduces the effect of these random variations. The amount of smoothening is determined by the standard deviation (σ) value of the Gaussian mask. In this work, a σ value of 5 is used which succeeded in lowering the effect of pixel jitter.

Once the frame is smoothened, the ViBe foreground extraction algorithm is used to detect the presence of foreground objects. According to the authors in [1], let the value of a pixel in the Euclidean color space be given by $v(x)$, where x denotes the location of the pixel. Let v_i denote a background sample value of index i. The ViBe algorithm maintains a model $M(x)$, for every pixel belonging to the background. The background model contains N background samples at location x obtained from previous frames.

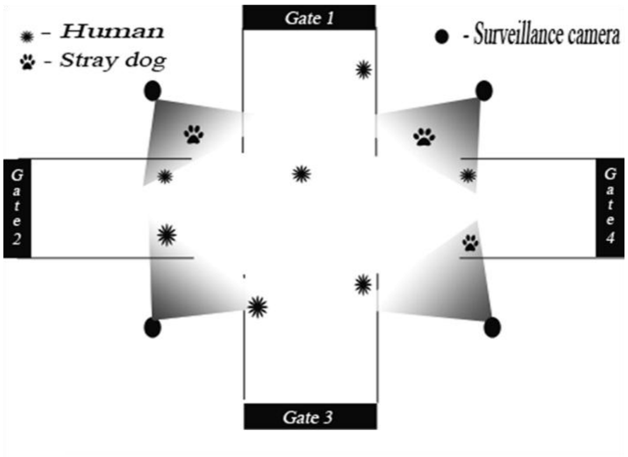


Fig.1. Scenario consisting of humans, dogs and surveillance cameras

The foreground object extraction is viewed as the problem of classifying a pixel as belonging to the foreground or to the background. For every pixel $v(x)$ a sphere $S_R(v(x))$ of radius R is drawn in the Euclidean color space. The number of samples in the background model of $v(x)$ that fall inside the sphere is calculated. If the number is greater than a particular threshold denoted by #, then the pixel is classified as belonging to the background. This is represented diagrammatically by the authors in [1] as shown in Fig.2. The entire algorithm with description of selection of parameters is discussed in detail in [1].

The ViBe algorithm outperforms the state of the art foreground extraction algorithms and is also fast. This makes it an ideal algorithm for security and surveillance applications where both accuracy and agility are of utmost importance. From the above discussion, it is evident that the radius 'R' of the sphere determines the amount of change required in the pixel intensity for a new pixel to be considered as a foreground object.

Surveillance operations have to be performed with equal accuracy during all periods, irrespective of the illumination condition of the monitored environment. Illumination conditions vary as the day progresses. During times of evening and night, the overall brightness level in the environment decreases. As a result, even when a foreground object is moving through the frames read by the camera, the resulting variation in pixel intensities may not be as high as the variations resulting from the same foreground object moving in the afternoon, when brightness levels are high. Therefore, the ViBe algorithm has to adapt to the changing environment. The radius R of the sphere used in the ViBe algorithm has to be smaller when the lighting conditions are bad and larger when the environmental conditions are bright.

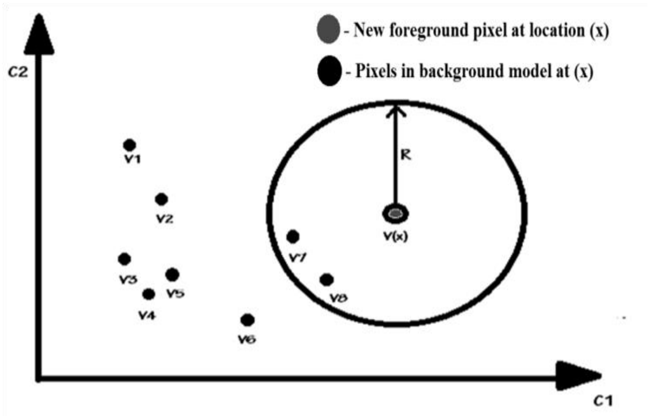

- New foreground pixel at location (x)
- Pixels in background model at (x)
C2
C1
R
v(x)
V1
V2
V3
V4
V5
V6
V7
V8

Fig.2. Comparison of a new pixel value v(x) with background model using a sphere $S_R\,(v(x))$ in Euclidian color space (C1, C2)

The first moment of the image or the 2-dimensional mean (μ) of the image is a measure of the overall brightness of the image. Hence, the radius of the sphere used in the ViBe algorithm is kept proportional to the mean brightness intensity in the image. This is achieved by setting a threshold on the mean value (μ). The threshold will ensure that proper surveillance is carried out during all periods of the day. In this work, a threshold value of 60 for μ is experimentally determined to function satisfactorily. The radius of the sphere used in the ViBe algorithm is set as per Eq.(1).

$$R(\mu)=\begin{cases}20 & if\mu \geq 60\\ 10 & if\mu < 60\end{cases} \tag{1}$$

The number of background pixel samples retained for every location in the frame is set to 20. As the foreground object enters or leaves the FOV of the camera, only partial shapes of the object are captured. Extracting features from partial shapes might result in improper classification. Further partial objects shapes have sizes which are smaller than the actual object. To overcome this problem, a window of size lesser than the frame size is used. An object can be considered for classification only if it falls completely inside the window. The exact dimensions of this frame window are discussed in the next section.

Sudden illumination changes in the input video can be misinterpreted as foreground objects. When large illumination changes occur, the foreground objects returned are extremely large or in some cases the entire frame itself. Additionally, small objects moving in the video, such as the rustling of leaves in the background are also not to be considered as foreground objects. To address these issues, two size thresholds are used in the proposed work. The upper size threshold eliminates all the extremely large foreground objects resulting from drastic illumination changes. The lower size threshold removes the minor objects and other noise artifacts.

After these two thresholds, a rectangular structuring element of size 2 × 4 is used to dilate the foreground object. This dilation helps in bolstering shape of the foreground object. The dilated frame is then labeled to determine the number of objects present in the input frame. Labeling enables the algorithm to handle multiple objects in the input video. The process of labeling assigns an integer label to all the groups of pixels in the input frame. As a result, all pixels belonging to a particular object or group are assigned the same label. 8-point connectivity is considered while labeling the pixels. The feature extraction process and classification is carried out individually for all the objects in the input frame.

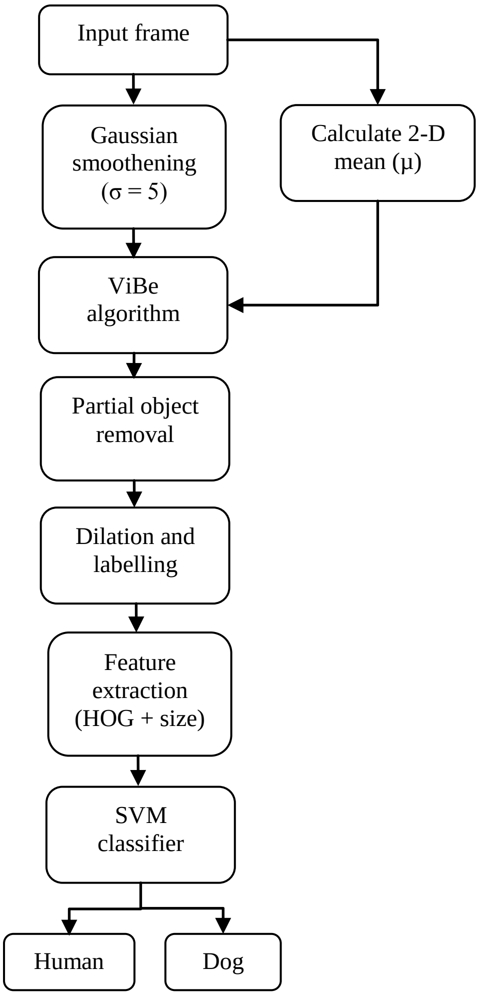


Fig.3. Flow diagram of proposed algorithm

Once the complete foreground object is extracted from the video feed, the next step is to compute the HOG shape descriptor for classification. The overlapping sub-windows of size 3 × 3 are found to work acceptably. The number of bins for the histogram is chosen as 9. Consequently, a HOG feature vector of length 81 is obtained. Along with the HOG descriptor, the size of the foreground object is also used as a feature. Stray dogs have sizes that are considerably smaller than humans and therefore the size feature could be used along with the HOG descriptor in the classification process. The final feature vector is of length 82. The overall flow diagram of the proposed work is shown in Fig.3.

An SVM binary classifier classifies the input video based on this feature vector into humans and stray dogs. A polynomial kernel function of order 3 is used to build the classifier. The SVM classifier has been trained with 19 videos which consist of 9 stray dog videos and 10 human videos.

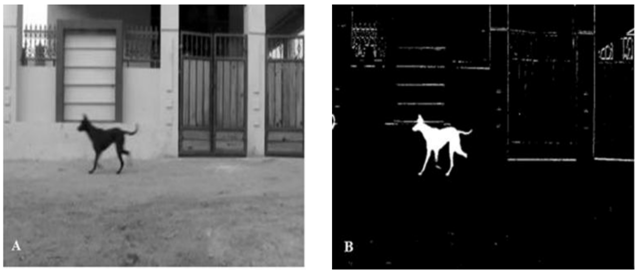
4. RESULTS AND DISCUSSIONS

A database of 50 surveillance videos is treated. The database created consists of videos of humans and stray dogs recorded from surveillance cameras in different environments under different periods of the day. The database also contains videos with multiple humans, multiple dogs and a combination of both. The different environments in the database include locations with trees and vegetation and the entrance of buildings for backgrounds. The videos in the database that were captured from actual surveillance cameras have duration of about 10-15 minutes. The videos of humans that were obtained from the KTH dataset were of considerably lower duration of less than a minute. These videos from the KTH dataset were primarily used to train the classifier with the human shapes in different orientations. Since human motion with different velocities such as running, walking and jogging also affects the captured shape of the human object, the classifier was trained with these cases too, so as to increase its classification accuracy.

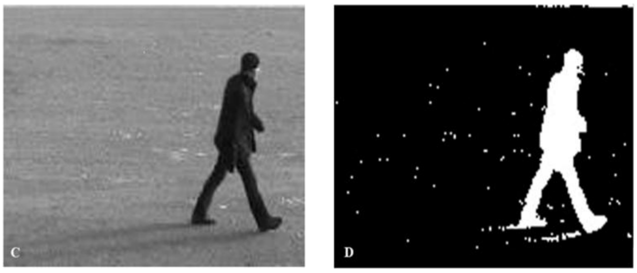
Also video of humans walking and running were taken from the KTH human action dataset [3]. The resolution of the frames captured by the surveillance camera is an important consideration for security applications since it decides the quality of the video. The higher the resolution of the camera, the higher will be its cost. The proposed composite algorithm worked efficiently even under very low resolutions. The input frames from the video were downsized continuously till the performance of the classification technique deteriorated significantly.

Initially, the frames captured from the surveillance camera are of size 800 × 600. The proposed composite algorithm has worked satisfactorily achieving a classification accuracy of 96% even when the frames were reduced to resolutions of 160 × 120. Fig.4(a) to Fig.4(d) shows the silhouettes of a stray dog and a human being obtained after the ViBe algorithm.

The inner frame used to remove partial objects consists of 10 percent lesser number of rows and columns. For instance, if the size of the input frame is 800 × 600, the size of the inner frame would be 720 × 540. Therefore, the foreground object will be considered for further processing only if all the pixels connected to the object fall inside this inner window. From Fig.4, the ViBe algorithm extracts the moving foreground object effectively. The frame containing the dog depicts a shot captured during day time, whereas the frame containing the person represents a shot captured when illumination in the environment is less. The ViBe algorithm's radius is effectively adjusted depending on the 2-Dimensional mean value (μ) of the input frame, enabling it to function effectively in different lighting conditions.



(a) (b)



(c) (d)

Fig.4. Result of Foreground objects segmentation. 4(a) and 4(c) show the input frame of a dog and man respectively. 4(b) and 4(d) are the corresponding outputs of the ViBe algorithm

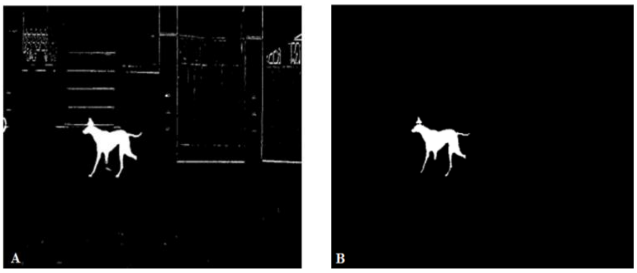
The small extraneous artifacts that are present along with the segmented dog and person in Fig.4(b) and Fig.4(d) are removed by the post processing. The post processing procedures include partial object removal, removal of tiny artifacts, dilation and labeling. Fig.5(a-d) depicts the results obtained after post processing the frames shown in Fig.4.

In the next stage, the objects are labeled and the HOG feature is extracted from the object shape silhouettes. The HOG vector is of length 81 since 3 × 3 windows and 9 histogram bins have been used for calculation. The sizes of the foreground objects are available from the previous stage. Now, it is appended with the HOG feature. Thus the length of the final feature vector is 82.

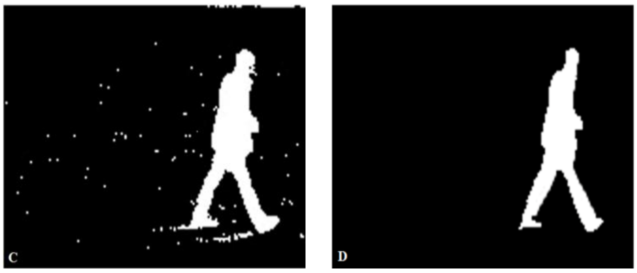
The SVM classifier has been trained using 19 videos out of which 10 videos contained humans and 9 contained stray dogs. The hybrid algorithm is tested using 50 surveillance videos taken under different lighting conditions and environments. The accuracy of the proposed hybrid algorithm along with the confusion matrix is given in Table.1 and Table.2 respectively.

The two scenarios that resulted in an improper classification consisted of the foreground objects; namely a human and a dog, present at a considerable distance from the camera and were moving away from the camera in a low, in a light environment. Since the objects kept moving further away from the surveillance camera, their sizes kept getting smaller and smaller. Consequently, their size of the foreground object became too small for the combination of the HOG descriptor and the SVM classifier to distinguish between them. In cases where the foreground objects were of appreciable size, the HOG descriptor showed enough variation that enabled the polynomial SVM classifier to successfully distinguish a human from a dog.

The simulation of the proposed composite algorithm is carried out in MATLAB 2010a and in OpenCV. The average overall classification time recorded is found to be 0.54 seconds.



(a) (b)



(c) (d)

Fig.5. Result of Post processing - Extraneous pixels other than the foreground object are removed as shown in 5(b) and 5(d) and only the predominant foreground object is retained.

5. CONCLUSION AND FUTURE WORK

In this paper, a novel composite algorithm for the detection and classification of stray dogs is proposed. The ViBe algorithm is chosen to extract the foreground object from the video obtained from the security camera. The algorithm is made adaptive depending upon the lighting conditions of the environment and is found to work effectively during all periods of the day. The feature vector containing the HOG descriptor along with the size of the foreground object is taken for classification. An SVM classifier with a polynomial kernel of order 3 achieved an accuracy of 96%. This algorithm shows promise due to its ability to work with a high accuracy; even in dynamic environments, at very low resolutions camera, at high speeds.

Table.1. Accuracy of the proposed composite algorithm

Total Number of Videos	**Correctly classified**	**Wrongly Classified**	**Accuracy %**
50	48	2	96

Table.2. Confusion Matrix

Class	**Human**	**Dog**
Human	25	1
Dog	1	23

ACKNOWLEDGEMENTS

The authors would like to thank Mr. M. Venkatesan, DGM (Electronics), Airport Authority of India (AAI), Chennai International Airport, Chennai for suggesting the problem of detection of stray dog in airport premises in connection with project submitted to Department of Information Technology (DIT), Government of India.

REFERENCES

[1] O. Barnich and M. Van Droogenbroeck, “ViBe: A Universal Background Subtraction Algorithm for Video Sequences”, *IEEE Transactions on Image Processing*, Vol. 20, No. 6, pp. 1709 - 1724, 2011.

[2] Navneet Dalal and Bill Triggs, “Histograms of oriented gradients for human detection”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886 - 893, 2005.

[3] C. Schuldt, I. Laptev and B. Caputo, “Recognizing human actions: a local SVM approach”, *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 3, pp. 32 - 36, 2004.

[4] Lei hu, Weibin Liu, Bo Li and Weiwei Xig, “Robust motion detection using Histogram of oriented gradients for illumination variations”, *Second International Conference on Industrial Mechatronics and Automation*, Vol. 2, pp. 443 - 447, 2010.

[5] P. E. Rybski, D. Huber, Daniel D. Morris and R. Hoffman, “Visual classification of coarse vehicle orientation using Histogram of Oriented Gradients features”, *IEEE Intelligent Vehicles Symposium*, pp. 921 - 928, 2010.

[6] I. Kamal, “Car recognition for multiple data sets based on histogram of oriented gradients and support vector machines”, *International Conference on Multimedia Computing and Systems*, pp. 328 - 332, 2012.

[7] Hsin–Chun Tsai et al., “A real-time awareness system for happiness expression based on the multilayer histogram of oriented gradients”, *Fourth International Conference on Awareness Science and Technology*, pp. 289 - 293, 2012.

[8] N. Damer and B. Fuhrer, “Ear Recognition using Multi scale Histogram of Oriented Gradients”, *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 21 - 24, 2012.

[9] Kuan Zheng, YuanXing Zhao, Jing Gu and QingMao Hu, “License plate detection using Haar-like features and histograms of oriented gradients”, *IEEE international Symposium on Industrial Electronics*, pp. 1502 - 1505, 2012.

[10] K. Muller et al., “An Introduction to Kernel-based learning algorithms”, *IEEE Transaction on Neural Networks*, Vol. 12, No. 2, pp. 181-201, 2001.