# EXPLOITING SUBCLASS INFORMATION IN ONE-CLASS SUPPORT VECTOR MACHINE FOR VIDEO SUMMARIZATION

*Vasileios Mygdalis, Alexandros Iosifidis, Anastasios Tefas and Ioannis Pitas*

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
Email: {vmigdal,aiosif,tefas,pitas}@aiia.csd.auth.gr

## ABSTRACT

In this paper, we propose a method for video summarization based on human activity description. We formulate this problem as the one of automatic video segment selection based on a learning process that employs salient video segment paradigms. For this one-class classification problem, we introduce a novel variant of the One-Class Support Vector Machine (OC-SVM) classifier that exploits subclass information in the OC-SVM optimization problem, in order to jointly minimize the data dispersion within each subclass and determine the optimal decision function. We evaluate the proposed approach in three Hollywood movies, where the performance of the proposed SOC-SVM algorithm is compared with that of the OC-SVM. Experimental results denote that the proposed approach is able to outperform OC-SVM-based video segment selection.

***Index Terms***— One class classification, Subclass One-Class SVM, Supervised Video Summarization

## 1. INTRODUCTION

Video summarization techniques develop condensed versions of a video stream through the identification of the most important and pertinent content within the stream [1]. The derived video summaries can be subsequently exploited in various applications, like interactive browsing and searching systems, thereby offering the user the ability to efficiently manage and effectively assess video content [2, 3]. Different techniques vary in the type of content used, the performed analysis and the type of video summary representation. Regarding the type the exploited information, it may belong in either generic or domain specific type (e.g., sports, news, movies etc.) as well as information besides the video stream (external information, provided by a user). Objects, events, perceptions and features are extracted by analysing the available modalities (image, sound or text) for abstracting intuitive semantics from the video stream. The abstracted semantic content that needs to be included in the target summary, can be represented as a cue of still images (key frames), a video skim, or by employing graphical and textual descriptions [1].

Key frames are images extracted from the video stream, which represent a video segment. Key frame cues are sequences of key frames presented in temporal order [4]. Video skims are cues of extracted video segments concatenated in temporal order, i.e., a video of shorter version of the original stream, containing the informative content. This can be performed e.g., by detecting the silent regions in audio stream [5]. Graphical cues present an additional level of detail as supplement to other cues. For example, a two dimensional color coded block map of the video stream which distinguishes video segments corresponding to dialogs, explosions and on-screen text, is proposed in [6]. A textual cue representation detecting text presence within the video frames and extracting the corresponding video segments is proposed in [7]. The above described approaches can be applied for generic video types, while methods exploiting domain-specific information have also been proposed. In surveilance videos, motion detection techniques are used, in order to create summaries that contain sets of object actions, like pedestrian walking. Detected actions taking place in different direction and speed, are fused in a single scene to form a short length video or graphical cue containing as many actions as possible [8, 9].

In video summarization techniques with applications to movie postproduction, the state-of-the-art approach exploits content selection techniques and video skimming. Usually, long videos containing multiple shots are temporally segmented, either manually or automatically by applying shot detection algorithms. A user attention model is proposed in [10], where visual, audio and textual features are extracted by applying multimodal analysis. A saliency score for each frame is computed and the most salient frames are selected to be the key frames. Video segments around each key frame are concatenated using a fade-in fade-out technique in order to form the summarized video skim. A different approach is proposed in [11]. The video stream is segmented into shots, then face detection and tracking are performed on the segmented video clips. Clustering is performed on the extracted facial images, in order to determine which images belong to the same character. The extracted characters are selected to form a character community network, which forms a graph of interactions between the movie characters. Redundant

interactions are excluded from the video skim.

In this paper, we describe a method for video summarization that operates on a video segment (shot or take) level. Resulting video summaries can be described using the MPEG-7 AVDP descriptions [12, 13]. Since most of the available video content has been recorded in order to capture human activity, we employ a video segment description that has been shown to achieve state-of-the-art performance in a relating task, i.e., in human action recognition [14]. After obtaining a vectorial video segment representation, we formulate the video summarization problem as the one of automatic video segment selection based on a learning process that employs salient video segment paradigms. To the best of our knowledge, it is the first time that such an approach is exploited for content-based video summarization, since all the previously described methods exploit unsupervised, or rule-based, approaches. Since non-salient video segment description (negative class) is extremely difficult, we follow an one-class learning approach. We employ the One-Class Support Vector Machine (OC-SVM) classifier to this end. Movie trailers have been specially edited in order to catch user attention and, at the same time, to describe the movie plot. Thus, it can be considered that shots forming movie trailers are good examples of salient video segments. Therefore, in order to automatically learn what video segment properties are considered to be important for the summarization of movies, we train the OC-SVM classifier by using vectorial shot representations describing human activity information of multiple movies belonging to several categories, i.e., action, comedy, thriller, drama, etc. For the evaluation of the proposed approach, we have employed three Hollywood movies. Since video shots belonging to different movie types are expected to be quite different, in order to enhance performance we introduce a novel extension of the OC-SVM classifier that exploits subclass information in its optimization process.

The remainder of the paper is structured as follows. In Section 2 we describe the proposed video summarization method. The proposed SOC-SVM classifier is described in Section 3. Experimental results evaluating its performance on video summarization are described in Section 4. Finally, conclusions are drawn in Section 5.

## 2. VIDEO SUMMARIZATION

Let us denote by $\mathcal{U} = \{V_1, \ldots, V_M\}$ a video database consisting of $M$ video segments $V_i$. $V_i$ may be various takes obtained during movie production, or different video shots appearing in a larger video (e.g., a movie). In the latter case, we automatically segment long videos containing multiple shots in shorter ones, each corresponding to a video shot. We employ the method in [15] to this end. We would like to employ $V_i$, $i = 1, \ldots, N$ in order to create a summary $\mathcal{S}$ of $\mathcal{U}$, where $\mathcal{S} \subseteq \mathcal{V}$, i.e., a video formed by the most salient video segments $V_i$. This process is usually noted as video skimming [1].

Let us denote by $\mathcal{X} = \{X_1, \ldots, X_N\}$ another video database that contains $N$ salient video segments $X_i$. We would like to employ the video segments in $\mathcal{X}$ in order to train a classifier that can determine whether the video segments $V_i$ are salient or not. Since we focus on human activity-based video summarization, we would like to employ a video description that highlights video segment properties relating to human activities. To this end, we employ the Dense Trajectory-based video description [14] in order to describe the video segments in $\mathcal{X}$ and $\mathcal{U}$. This video description has been shown to provide state-of-the-art performance on a relating task, i.e., human action recognition in unconstrained videos.

Dense Trajectory-based video description calculates five descriptor types on the trajectory of densely-sampled video frame interest points that are tracked for a number of consecutive video frames. The five descriptor types are: Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF), Motion Boundary Histogram along direction $x$ (MBHx), Motion Boundary Histogram along direction $y$ (MBHy) and the normalized trajectory coordinates (Traj). We employ these video segment descriptions in order to obtain five video segment representations by using the Bag-of-Words model. That is, the descriptors calculated for the training video segments $X_i$, $i = 1, \ldots, N$ are clustered in order to determine five sets of descriptor prototypes (each for a descriptor type). Subsequently, each of the video segments $X_i$ and $V_i$ are represented by five vectors $\mathbf{x}_i^v$, $\mathbf{v}_i^v$, $v = 1, \ldots, 5$, respectively. In order to fuse the information appearing in different video representations, we combine the video segment representations with kernel methods, as in [14]. That is, we employ the RBF-$\chi^2$ kernel function, where different descriptor types are combined following a multi-channel approach [16]:

$$\mathbf{K}(\mathcal{X}_i, \mathcal{X}_j) = exp\Big( - \sum_v \frac{1}{4A^v} D\left(\mathbf{x}_i^v, \mathbf{x}_j^v\right) \Big), \qquad (1)$$

$D\left(\mathbf{x}_i^k, \mathbf{x}_j^k\right)$ is the $\chi^2$ distance between the BoW-based video representation of $\mathbf{x}_i$ and $\mathbf{x}_j$ with respect to the $k$-th channel. $A^k$ is the mean value of the $\chi^2$ distances between the training samples for the $k$-th channel.

After calculating the kernel matrices for the training and test video segments, we would like to calculate a vectorial representation for each of the video segments in $\mathcal{X}$ and $\mathcal{U}$. To this end, we apply a multi-level data grouping process that is inspired by relative work in deep learning [17]. At each level $l$ of this process, we determine $l$ groups of training video segments $\mathcal{X}_i$ by applying the kernel K-Means algorithm [18]. Subsequently, we calculate the distances between the training and test video segments representation in the kernel space from the $l$ cluster centers. Let us denote by $\mathbf{p}_i^l \in \mathbb{R}^l$, $i = 1, \ldots, N$ and $\mathbf{q}_i^l \in \mathbb{R}^l$, $i = 1, \ldots, M$ the distance vectors obtained for the training and test video seg-

ments, respectively. The distance-based representations of the training and test video segments are obtained by concatenating the distance vector obtained for all the $L$ processing steps, i.e.:

$$\mathbf{p}_i = [\mathbf{p}_i^{1T}, \ldots, \mathbf{p}_i^{LT}]^T \tag{2}$$

and

$$\mathbf{q}_i = [\mathbf{q}_i^{1T}, \ldots, \mathbf{q}_i^{LT}]^T. \tag{3}$$

Finally, $\mathbf{p}_i$, $\mathbf{q}_i$ are mapped to similarity-based representations $\mathbf{z}_i$, $\mathbf{g}_i$ having elements equal to $z_{ik} = \frac{p_{ik}}{\sum_n p_{in}}$ and $g_{ik} = \frac{q_{ik}}{\sum_n q_{in}}$.

After obtaining $\mathbf{x}_i \in \mathbb{R}^D$, $i = 1, ..., N$, we employ them in order to train the proposed Subclass One-Class SVM classifier, as will be described in the following subsection.

## 3. SUBCLASS ONE-CLASS SUPPORT VECTOR MACHINES

In this Section, we describe in detail the proposed SOC-SVM classifier that incorporates subclass information in the OC-SVM optimization process. Subclasses have been successfully used in dimensionality reduction techniques [19, 20], Single-hidden Feedforward Neural (SLFN) network training [21] and multi-class SVM training [22]. In all these cases, subclasses are determined by applying a clustering technique, e.g., the K-Means algorithm [23], on the samples belonging to each of the classes independently. In our case, since the entire training set belongs to one class only, we cluster all the training vectors $\mathbf{z}_i$ in order to determine $K$ subclasses, each represented by the corresponding mean cluster vector by $\mathbf{m}_k$, $k = 1, \ldots, K$.

We employ the mean cluster vectors $\mathbf{m}_k$ in order to calculate the within-subclass dispersion of the training vectors $\mathbf{z}_i$ by:

$$\mathbf{S} = \sum_{k=1}^{K} \sum_{\mathbf{z}_i \in \mathcal{M}_k} \frac{N_k}{N} (\mathbf{z}_i - \mathbf{m}_k)(\mathbf{z}_i - \mathbf{m}_k)^T, \tag{4}$$

where $\mathcal{M}_k$ denotes the $k$-th subclass formed by $N_k$ training vectors. Let us denote by $\mathbf{w}$ a vector that can be used in order to project the training data $\mathbf{z}_i$ to a one-dimensional feature space where the the within-subclass dispersion is minimized. It is straightforward to show that the training data dispersions in the resulted (one-dimensional) feature space is given by $\tilde{s} = \mathbf{w}^T\mathbf{S}\mathbf{w}$. By observing this, we propose to minimize $\tilde{s}$ subject to separability constraints given in OC-SVM [24], i.e.:

$$\min_{\mathbf{w}, \xi_i, \rho} \quad \frac{1}{2}\mathbf{w}^T\mathbf{S}\mathbf{w} + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \tag{5}$$

$$\textbf{s.t.:} \quad \mathbf{w}^T\mathbf{z}_i \geq \rho - \xi_i, \;\; i = 1, ..., N, \tag{6}$$

$$\xi_i \geq 0, \tag{7}$$

where $\xi_i$ are the slack variables and $\nu$ denotes the trade-off between minimizing the two terms. An additional constraint $\mathbf{w}^T\mathbf{S}\mathbf{w} > 0$ is also imposed in the above optimization problem denoting the positive-definiteness of $\mathbf{S}$.

Based on the Karush-Kuhn-Tucker (KKT) theorem [25], the above described optimization problem can be solved by finding the saddle point of the Lagranzian:

$$\begin{aligned}\mathcal{L}(\mathbf{w}, \xi_i, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{1}{2}\mathbf{w}^T\mathbf{S}\mathbf{w} + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \\ &- \sum_{i=1}^{N}\alpha_i\left(\mathbf{w}^T\mathbf{z}_i - \rho + \xi_i\right) - \sum_{i=1}^{N}\beta_i\xi_i,\end{aligned}$$

leading to the following optimality conditions:

$$\frac{\vartheta\mathcal{L}}{\vartheta\mathbf{w}} = 0 \;\Rightarrow\; \mathbf{S}\mathbf{w} = \sum_{i=1}^{N}\alpha_i\mathbf{z}_i, \tag{8}$$

$$\frac{\vartheta\mathcal{L}}{\vartheta\xi_i} = 0 \;\Rightarrow\; \beta_i = \frac{1}{\nu N} - \alpha_i, \tag{9}$$

$$\frac{\vartheta\mathcal{L}}{\vartheta\rho} = 0 \;\Rightarrow\; \sum_{i=1}^{N}\alpha_i = 1. \tag{10}$$

Thus, given that $\mathbf{S}$ is non-singular, the optimal vector $\mathbf{w}$ is given by:

$$\mathbf{w} = \mathbf{S}^{-1}\sum_{i=1}^{N}\alpha_i\mathbf{z}_i. \tag{11}$$

Replacing (8)(9)(10) in $\mathcal{L}(\mathbf{w}, \xi_i, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and using the KKT conditions, the optimization problem in (5) can be reformulated to its dual form:

$$\max_{\boldsymbol{\alpha}} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbf{z}_i^T\mathbf{S}^{-1}\mathbf{z}_j, \tag{12}$$

subject to $0 \geq \alpha_i \geq \frac{1}{\nu N}$ and $\sum_i \alpha_i = 1$. After solving (12), a test vector $\mathbf{v}_i$ can be introduced to the classifier. Its response is given by:

$$f(\mathbf{g}_i) = \left(\mathbf{w}^T\mathbf{S}^{-1}\mathbf{g}_i - \rho\right). \tag{13}$$

By observing (12) we can see that the solution of the proposed SOC-SVM classifier is similar to that of OC-SVM. In order to exploit standard OC-SVM implementations [26], we can use an approach similar to the one proposed in [27]. We can apply eigenanalysis on the matrix $\mathbf{S}$ in order to decompose it to $\mathbf{S} = \mathbf{V}\mathbf{L}\mathbf{V}^T$, where $\mathbf{V}$ is an orthonormal matrix that contains the eigenvectors of $\mathbf{S}$ and $\mathbf{L}$ is a diagonal matrix containing the eigenvalues of $\mathbf{S}$. Then, we can employ the matrix $\mathbf{P} = \mathbf{V}\mathbf{L}^{-\frac{1}{2}}$ in order to map the original training vectors $\mathbf{z}_i$, $i = 1, \ldots, N$ to vectors $\mathbf{y}_i$ by:

$$\mathbf{y}_i = \mathbf{P}^T\mathbf{z}_i. \tag{14}$$

It can be shown that:

$$\mathbf{y}_i^T\mathbf{y}_j = \mathbf{z}_i^T\mathbf{P}\mathbf{P}^T\mathbf{z}_j = \mathbf{z}_i\mathbf{S}^{-1}\mathbf{z}_j. \tag{15}$$

Thus, by applying OC-SVM on the vectors $\mathbf{y}_i$ corresponds to applying the proposed SOC-SVM with the optimization problem given in (5). In the case where $\mathbf{S}$ is singular, one can choose to keep fewer eigenvectors of $\mathbf{S}$ (the ones corresponding to the non-zero eigenvalues) for $\mathbf{y}_i$ calculation.

## 4. EXPERIMENTAL RESULTS

In this Section, we present experiments conducted in order to evaluate the proposed video summarization method and the performance of the proposed SOC-SVM classifier. We have employed three Hollywood movies belonging to action, adventure and drama categories, respectively. In order to train the proposed SOC-SVM classifier we have employed eighteen Hollywood movie trailers belonging to action, comedy, thriller and drama categories. It should be noted that the trailers of the three (test) movies are not included in the training set.

Here we should note that usually, video summarization techniques are evaluated based on qualitative criteria [10, 11], e.g., by calculating criteria like the 'informativeness', or the 'enjoyability' based on the ratings provided by users for the entire video summary. However, such criteria are too subjective. In order to perform quantitative evaluation of the performance of each classifier in video summarization, we employ the trailers of the three (test) movies and manually create ground truth labels denoting whether a video segment (shot) of each movie has been employed in order to form the trailer, or not. We introduce the test vectors $\mathbf{v}_i,\ i = 1, \ldots, M$ to the SOC-SVM classifier trained on the video shots of the training movie trailers and obtain their responses. Subsequently, we keep the video segments corresponding to the $L = pN$ maximal function values, where $0 < p < 1$, and calculate the percentage of the movie trailer belonging to the created video summary.

We set the dimensionality of the BoW-based video segment representations $\mathbf{x}_i^v$ equal to $4000$, which is a value that has been shown to provide satisfactory performance in a wide range of activity classification problems [14]. In order to obtain distance-based video segment representations we test the values $L = 2, 5, 10, 20$. We employ $\mathbf{z}_i$ in order to train the proposed SOC-SVM classifier. For comparison reasons, we also train the standard OC-SVM classifier [24] and a variant of OC-SVM exploiting the variance of the training data [28], which is noted as Minimum Variance One-Class SVM (MVOC-SVM) classifier hereafter.

The mean performance obtained by applying the three algorithms for different values of $p$ are illustrated in Table 1. As can be seen in this Table, the incorporation of the training data variance in the optimization process of the OC-SVM enhances performance since MVOC-SVM outperforms OC-SVM in most cases. The proposed SOC-SVM classifier, by exploiting subclass information (the results reported in the Tables correspond to a value of $K = 2$), further enhances performance, as the proposed SOC-SVM classifier clearly outperforms the remaining two algorithms in all the cases.

**Table 1**. Mean performance of the three algorithms.

| $p$ | OC-SVM | MVOC-SVM | **SOC-SVM** |
|---|---|---|---|
| 0.1 | 11.68% | 11.68% | **21.19%** |
| 0.2 | 22.39% | 22.69% | **29.11%** |
| 0.3 | 33.53% | 34.07% | **38.87%** |
| 0.4 | 41.36% | 45.09% | **54.62%** |
| 0.5 | 53.96% | 59.64% | **62.68%** |

In an attempt to explain the results obtained in the above described experiments, we have created the summaries of the three movies by exploiting the order of the video segments in $\mathcal{U}$. We have observed that video segments forming the obtained video summaries are quite similar to each other in terms of video saliency. This means that the proposed approach can be employed in order to assign saliency scores to the various video segments in terms of saliency and produce video segment suggestions that can be used in order to accelerate video post-processing, or to provide a good summarization of a video in terms of saliency.

## 5. CONCLUSION

In this paper, we described a method for video summarization exploiting an activity-based video segment description. We have formulated the problem as the one of automatic video segment selection based on an one-class learning process exploiting salient video segment paradigms. For this one-class classification problem, we have proposed a novel extension of the OC-SVM classifier that incorporates subclass information in the OC-SVM optimization problem. Experimental results denote that the proposed SOC-SVM classifier is able to outperform the standard OC-SVM approach, as well as another variant of the OC-SVM classifier exploiting the variance of the training data in the OC-SVM optimization problem.


# Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART).


## 6. REFERENCES


[1] A. G. Money and H. Agius, "Video summarisation: A conceptual framework and survey of the state of the art," *Journal of Visual Communication & Image Representation*, vol. 19, pp. 121–143, 2008.

[2] Y. Li, S. Lee, C. Yeh, and C. Kuo, "Semantic retrieval of multimedia," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 79–89, 2006.

[3] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: state of the art and challenges," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 1–19, 2006.

[4] C. Gianluigi and S. Raimondo, "An innovative algorithm for key frame extraction in video summarization," *Journal of Real-Time Image Processing*, vol. 1, no. 1, pp. 69–88, 2006.

[5] M. Furini and V. Ghini, "An audio-video summarisation scheme based on audio and video analysis," *IEEE Consumer Communications and Networking Conference*, 2006.

[6] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video abstracting," *Communications of the ACM*, vol. 40, no. 12, pp. 54–62, 1997.

[7] B. Luo, X. Tang, J. Liu, and H. Zhang, "Video caption detection and extraction using temporal information," *IEEE International Conference on Image Processing*, 2003.

[8] W. Fu, J. Wang, L. Gui, H. Lu, and S. Ma, "Online video synopsis of structured motion," *Neurocomputing*, vol. 135, pp. 155–162, 2014.

[9] K. Streib and J. Davis, "Summarizing high-level scene behavior," *Machine Vision and Applications*, vol. 25, no. 1, pp. 229–244, 2014.

[10] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas, and Y. Avrithis, "Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1553–1568, 2013.

[11] C. M. Tsai, L. W. Kang, C. W. Lin, and W. Lin, "Scene-based movie summarization via role-community networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1927–1940, 2013.

[12] Bangalore S Manjunath, Philippe Salembier, and Thomas Sikora, *Introduction to MPEG-7: multimedia content description interface*, vol. 1, John Wiley & Sons, 2002.

[13] ISO/IEC TR 15938-11:2005/Amd 1:2012, "Audiovisual description profile (avdp) schema," 2012.

[14] H. Wang, A. Klaser, C. Schmid, and C. L. Liu, "Dense trajectories and motion boundary descriptors for action recognition," *International Journal of Computer Vision*, vol. 103, no. 1, pp. 60–79, 2012.

[15] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 82–91, 2006.

[16] J. Zhang, M. Marszalek, M. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.

[17] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *arXiv : 1206.5538v3*, 2014.

[18] J.S. Taylor and N. Cristianini, "Kernel methods for pattern analysis," 2004, Cambridge University Press.

[19] X. Chen and T. Huang, "Facial expression recognition: A clustering-based approach," *Pattern Recognition Letters*, vol. 24, pp. 1295–1302, 2003.

[20] M. Zhu and A. M. Martinez, "Subclass discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1274–1286, 2006.

[21] A. Iosifidis, A. Tefas, and I. Pitas, "Minimum class variance extreme learning machine for human action recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1968–1979, 2013.

[22] G. Orfanidis and A. Tefas, "Exploiting subclass information in support vector machines," *IEEE International Conference on Pattern Recognition*, 2012.

[23] S. Theodoridis and K. Koutroumbas, "Pattern recognition," *Academic Press*, November 2008.

[24] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, , and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, pp. 1443–1471, 2001.

[25] R. Fletcher, "Practical methods of optimization," *Wiley*, 1987.

[26] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[27] T. S. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," *Advances in Neural Information Processing*, 1999.

[28] S. Zafeiriou and N. Laskaris, "On the improvement of support vector techniques for clustering by means of whitening transform," *IEEE Signal Processing Letters*, vol. 15, pp. 198–201, 2008.