Video Segmentation based on Multi-kernel Learning and Feature Relevance Analysis for Object Classification

S. Molina-Giraldo, J. Carvajal-González, A. M. Álvarez-Meza and G. Castellanos-Domínguez
Signal Processing and Recognition Group, Universidad Nacional de Colombia, Manizales, Colombia

Keywords: Background Subtraction, Multiple Kernel Learning, Relevance Analysis, Data Separability.

Abstract: A methodology to automatically detect moving objects in a scene using static cameras is proposed. Using Multiple Kernel Representations, we aim to incorporate multiple information sources in the process, and employing a relevance analysis, each source is automatically weighted. A tuned Kmeans technique is employed to group pixels as static or moving objects. Moreover, the proposed methodology is tested for the classification of abbandoned objects. Attained results over real-world datasets, show how our approach is stable using the same parameters for all experiments.

1 INTRODUCTION

A system that monitors an area by camera and detects moving people or objects is called a surveillance system. Intelligent video surveillance systems can achieve unsupervised results using video segmentation, with which the moving objects can be extracted from video sequences. Many segmentation algorithms have been proposed. Among them, algorithms with background subtraction usually show superior performance (Chen et al., 2007). Background subtraction is a typical and crucial process for a surveillance system to detect moving objects that may enter, leave, move or left unattended in the surveillance region. Unattended objects as bags or boxes in public premises such as airports, terminal bus and train stations are a threat for these places because they can be used as a mean of terrorist attacks, especially for bombs (González et al., 2012).

Image sequences with dynamic backgrounds often cause false classification of pixels, one common solution is to map alternate color spaces, however it has fallen to solve this problem and an enhanced solution is the use of image features, where the distributions at each pixel may be modeled in a parametric manner using a mixture of Gaussians (Klare and Sarkar, 2009) or using non-parametric kernel density estimation (Elgammal et al., 2002). The self organizing maps have been also explored as an alternative for the background subtraction task, because of their nature to learn by means of a self-organized manner local variations (Maddalena and Petrosino, 2008), however, these techniques have the drawback of manually setting a large number of parameters.

In this work, a methodology called Weighted Gaussian Kernel Video Segmentation (WGKVS) is proposed, which aims to construct a background model and then, incorporating multiple information sources by a MKL framework, performs a background subtraction enhancing thus the representation of each pixel. A relevance analysis for the automatic weight selection of the MKL approach is made. Furthermore, a tuned Kmeans technique is employed to group pixels as static or moving objects. The proposed WGKVS is tested in the surveillance task of the classification of abbandoned objects in the scene. In this regard, using the segmented frame, the objects detected as not belonging to the background model that are spatially splitted, are relabeled as new independent objects and then characterized with the methodology implemented in (González et al., 2012) for further classification.

The remainder of this work is organized as follows. In section 2, the proposed methodology is described. In section 3, the experiments and results are presented. Finally, in sections 4 and 5 we discuss and conclude about the attained results.

Molina-Giraldo S., Carvajal-González J., M. Álvarez-Meza A. and Castellanos-Domínguez G. (2013).
Video Segmentation based on Multi-kernel Learning and Feature Relevance Analysis for Object Classification.
In *Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods*, pages 396-401
DOI: 10.5220/0004269403960401

2 THEORETICAL BACKGROUND

2.1 Background Initialization

The first step of the proposed WGKVS is a background model initialization. Given an image sequence $\boldsymbol{H}$ with q frames, we propose to use a subsquence of frames $\boldsymbol{H}_{(t:k)}$ to initialize a background model based on the approach exposed in (Gutchess et al., 2001). This approach, using an optical flow algorithm is succesfully able to construct a statistical background model with the most likely static pixels during the subsequence for each RGB component, and it also computes its standard deviation. We also propose to compute a background model using the normalized RGB components (rgb) in order to supress the shadows casted by the moving objects as described in (Elgammal et al., 2002). Hence, a background model is stored in a matrix $\boldsymbol{Y}$.

2.2 Background Subtraction based on Multi-kernel Learning and Feature Representation

Recently, machine learning approaches have shown that the use of multiple kernels, instead of only one, can be useful to improve the interpretation of data (Rakotomamonjy et al., 2008). Given a frame $\boldsymbol{F}$ from the image sequence $\boldsymbol{H}$ and a background model $\boldsymbol{Y}$ obtained from the same sequence, using a set of p feature representations for each frame pixel $f_i = \{f_i^z : z = 1,\ldots,p\}$ and each pixel $y_i = \{y_i^z : z = 1,\ldots,p\}$ belonging to the background model, based on the Multi-Kernel Learning (MKL) methods (Gonen and Alpaydin, 2010), a background subtraction procedure can be computed via the function:

$$\kappa_\omega\left(f_i^z, y_j^z\right) = \omega_z \kappa(f_i^z, y_i^z), \qquad (1)$$

subject to $\omega_z \geq 0$, and $\sum_{i=1}^{p} \omega_z = 1$ $(\forall \omega_z \in \mathbb{R})$. Regarding to video segmentation procedures, each pixel of each frame $\boldsymbol{F}$ can be represented by a dissimilarity measure with a background model by using p different image features (e.g. Color components, textures), in order to enhance the performance of further segmentation stages. Specifically, we propose to use the RGB and the rgb components as features and as basis kernel $\kappa\{\cdot\}$, a gaussian kernel $\boldsymbol{G}$ defined as:

$$\boldsymbol{G}^z(f_i^z, y_i^z) = \exp\left(-\frac{1}{2}\left(\frac{|f_i^z - y_i^z|}{\sigma_i^z}\right)^2\right), \qquad (2)$$

where σ_i^z corresponds to a percentage of the standard deviation of pixel y_i in the feature z from the background model.

As it can be seen from (1), it is necessary to fix the ω_z free parameters, to take advantage, as well as possible of each feature representation. To complete the feature space, the row m and column position n are added as features, in order to keep the local relationships among pixels. Therefore, a feature space $\boldsymbol{X}_{((m\times n)\times 8)}$ is obtained.

2.3 MKL Weight Selection based on Feature Relevance Analysis

Using the feature space $\boldsymbol{X}$, we aim to select the weights values ω_z in MKL by means of a relevance analysis. This type of analysis is applied to find out a low-dimensional representations, searching for directions with greater variance to project the data, such as Principal Component Analysis (PCA), which is useful to quantify the relevance of the original features, providing weighting factors taking into consideration that the best representation from an explained variance point of view will be reached (Daza-Santacoloma et al., 2009). Given a set of features $(\boldsymbol{\eta}_z : z = 1,\ldots,p)$ corresponding to each column of the input data matrix $\mathbf{X} \in \mathbb{R}^{r\times p}$ (a set of p features describing a pixel image h_i), the relevance of $\boldsymbol{\eta}_z$ can be identified as ω_z, which is calculated as $\boldsymbol{\omega} = \sum_{j=1}^{d} |\lambda_j \mathbf{v}_j|$, with $\boldsymbol{\omega} \in \mathbb{R}^{p\times 1}$, and where λ_j and $\mathbf{v}_j$ are the eigenvalues and eigenvectors of the covariance matrix $\mathbf{V} = \mathbf{X}^\top\mathbf{X}$, respectively.

Therefore, the main assumption is that the largest values of ω_z lead to the best input attributes. The d value is fixed as the number of dimensions needed to conserve a percentage of the input data variability. Then using $\boldsymbol{\omega}$, a weighted feature space is computed as: $\hat{\boldsymbol{X}} = \boldsymbol{X} \times diag(\boldsymbol{\omega})$.

2.4 Video Segmentation based on Kmeans Clustering Algorithm

In order to segment the moving objects, a Kmeans clustering algorithm with a fixed number of clusters equal to 2 is employed over $\hat{\boldsymbol{X}}$, hence, the objects that do not belong to the background model (moving objects) are groupped in a cluster and the objects that belong to the background model (static objects) in the other one. Initially, the clusters are located at the coordinates given by the matrix $\boldsymbol{Q}$, which is obtained by the cluster initialization algorithm called *maxmin* described in (Cuesta-Frau et al., 2003), making the segmentation process more stable. Finally, with the attained labels $\boldsymbol{l}$, using a post-process stage, groups of pixels detected as moving objects that do not surpass a value u of minimum size for an object are deleted.

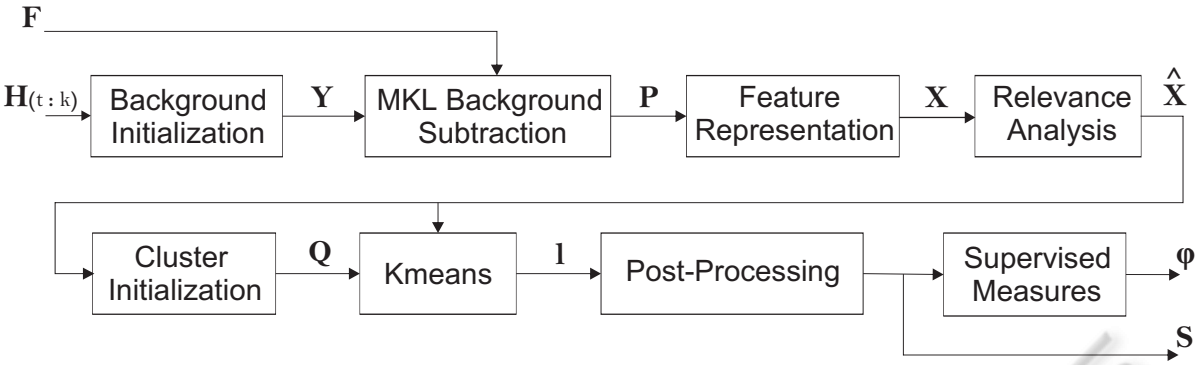

Figure 1: WGKVS Scheme.

The results are stored into a binary matrix $\boldsymbol{S}$. In Fig. 1 is ilustrated the general scheme for WGKVS.

2.5 Quantitative Measures

For measuring the accuracy of the proposed methodology for moving object segmentation, three different pixel-based measures have been adopted: *Recall* = $t_p/(t_p+f_n)$, *Precision* = $t_p/(t_p+f_p)$ and *Similarity* = $t_p/(t_p+f_n+f_p)$ (Maddalena and Petrosino, 2008), where t_p (true positives), f_p (false positives) and f_n (false negatives) are obtained while comparing against a hand-segmented ground truth.

2.6 Object Characterization and Classification

The WGKVS approach is applied into a real world surveillance task: the classification of abbandoned objects. Using the segmented frame $\boldsymbol{S}$, the groups detected as moving objects that are spatially splitted, are relabeled as new independent objects. With these new labels, each object is enclosed in a bounding box, and using the characterization process described in (González et al., 2012), each object is represented by 14 geometrical and 7 statistical features. A Knn classifier is trained using images belonging to the classes: people and baggage objects.

3 EXPERIMENTS

The proposed methodology is tested using three different Databases. Each Database includes image sequences that represent typical situations for testing video surveillance systems. Following, the Databases are described.

A-Star-Perception: This Databse is publicly available at http://perception.i2r.a-star.edu.sg. It contains 9 image sequences recorded in different scenes. Hand-segmented ground truths are available for each sequence, thus, supervised measures can be used. For concrete testing, the sequences: WaterSurface, Fountain, ShoppingMall and Hall are used. The first two sequences are recorded in outdoor scenarios which present high variations due to their nature, hence the segmentation process posess a considerable challenge. The other two sequences are recorded in public halls, in which are present many moving objects casting strong shadows and crossing each other, difficulting the segmentation task.

Left-Packages: Publicly available at http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1, this Database contains 5 different image sequences recorded at an interior scenario which has several ilumination changes. The main purpose of this database is the identification of abandoned objects (a box and a bag). For testing, hand-segmented ground truths from randomly selected frames are made.

MSA: This Database is publicly available at http://cvprlab.uniparthenope.it. It contains a single indoor sequence, with stable lighting conditions, nonetheless, strong shadows are casted by the moving objects. The purpose of this sequence is also the detection of abandoned objects, in this case a briefcase.

Three different experiments are performed, in all of them, the free parameter σ_i^z is heuristically set as 5 times the standard deviation of each pixel representation y_i^z. The minimum size of a detected moving object u is set as $0.005 \times (m \times n)$.

The first experiment aims to prove the effectiveness of the proposed WGKVS approach when incorporating more information sources into the segmentation process with an automatic weighting selection. To this end, the image sequences WaterSurface, Fountain, ShoppingMall, Hall, LeftBag and LeftBox are used. The WGKVS segmentation results are compared against GKVS (WGKVS with all equal weights), and traditional GKVS-RGB (GKVS using

only RGB components). In Fig. 2 are shown the different segmentation results for the frame 1523 of the sequence WaterSurface. The releavance weights are shown in Fig. 3. In Tables 1, 2 and 3 are exposed the attained results for each method.

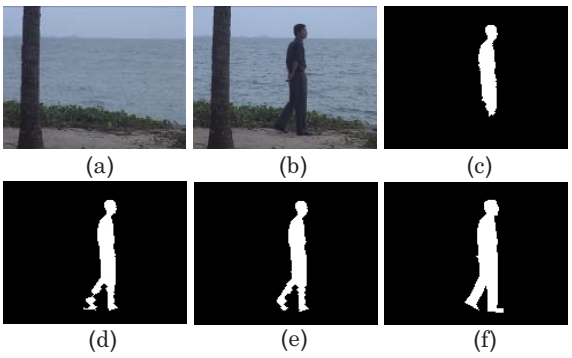

Figure 2: WaterSurface (Frame 1523). (a) Background Model. (b) Original Frame. (c) GKVS-RGB. (d) GKVS. (e) WGKVS. (f) Ground Truth.

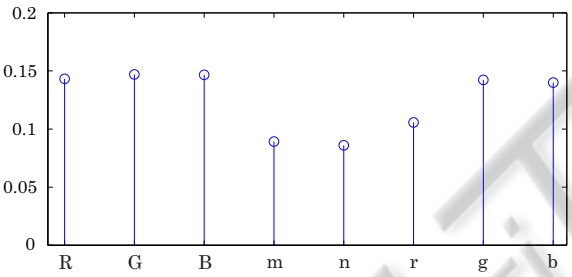

Figure 3: Relevance Weights for Sequence WaterSurface (Frame 1523).

Table 1: Segmentation Performance for GKVS-RGB.

Video	Recall	Precision	Similarity
WaterSurface	0.677	0.995	0.676
Fountain	0.509	0.897	0.480
ShoppingMall	0.436	0.385	0.302
Hall	0.489	0.809	0.434
LeftBag	0.610	0.839	0.555
LeftBox	0.697	0.906	0.647

Table 2: Segmentation Performance for GKVS.

Video	Recall	Precision	Similarity
WaterSurface	0.762	0.995	0.759
Fountain	0.559	0.909	0.528
ShoppingMall	0.571	0.680	0.442
Hall	0.518	0.829	0.462
LeftBag	0.614	0.842	0.560
LeftBox	0.699	0.910	0.651

Table 3: Segmentation Performance for WGKVS.

Video	Recall	Precision	Similarity
WaterSurface	0.770	0.994	0.767
Fountain	0.587	0.908	0.552
ShoppingMall	0.643	0.715	0.512
Hall	0.520	0.837	0.473
LeftBag	0.627	0.848	0.571
LeftBox	0.729	0.915	0.674

The second type of experiments are performed to compare the WGKVS algorithm against a traditional video segmentation algorithm named Self-Organizing Approach to Background Subtraction (SOBS), which builds a background model by learning in a self-organazing manner the scene variations, and detects moving object by using a background subtraction (Maddalena and Petrosino, 2008). The SOBS video segmentation approach has been used as a reference to compare video segmentation approaches and it has been also included in surveilance systems surveys (Raty, 2010). The software for the SOBS approach is publicly available at http://cvprlab.uniparthenope.it/index.php/download/92.html. For testing, the 10 parameters of the SOBS approach are left as default. In Figs. 4 and 5 are the segmentation results using WGKVS and SOBS for the frame 0996 of the sequence LeftBag and frame 1980 of the sequence ShoppingMall respectively. In table 4 are the segmentation results for the SOBS algorithm.

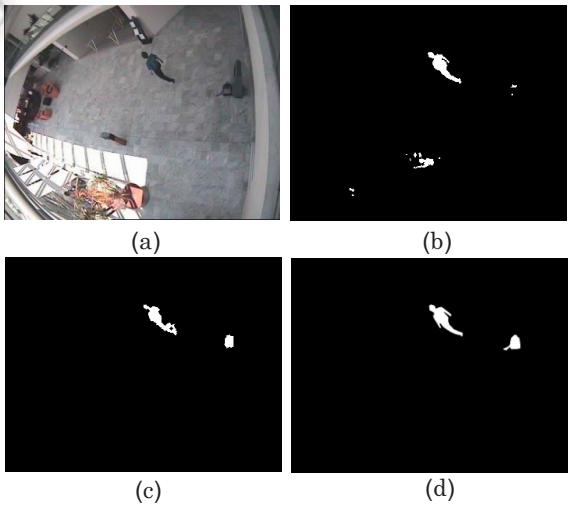

Figure 4: LeftBag (Frame 0996). (a) Original Frame. (b) SOBS. (c) WGKVS. (d) Ground Truth.

Finally, the third type of experiment is made in order to test the proposed WGKVS for the classfication of abbandoned objects. In this sense, the process de-

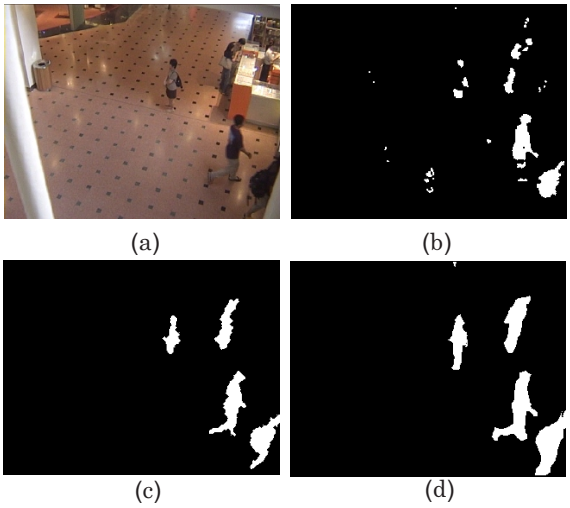

Figure 5: ShoppingMall (Frame 1980). (a) Original Frame. (b) SOBS. (c) WGKVS. (d) Ground Truth.

Table 4: Segmentation Performance for SOBS.

Video	Recall	Precision	Similarity
WaterSurface	0.709	0.998	0.708
Fountain	0.349	0.971	0.346
ShoppingMall	0.522	0.861	0.482
Hall	0.708	0.888	0.648
LeftBag	0.472	0.642	0.373
LeftBox	0.746	0.806	0.634

scribed in section 2.6 is employed. For testing, the sequences: LeftBag, LeftBox and MSA are used. The aim is to classify objects as: people or baggage objects (e.g. briefcases, boxes, backpacks, suitcases). A knn classifier is trained using a dataset of 70 images of people and 82 images of baggage objects, and as validation, we use the objects segmented by the WGKVS. It is important to remark, that the objects from the dataset used for training are characterized by the same process. In Fig. 6, are shown some resulting bounded objects. In total, 38 objects are used in the validation database, 11 belong to the baggage objects class and 27 to the people class. In Fig. 7, are shown two samples of the characterization process for a person and a bag. The classification results are exposed in table 5.

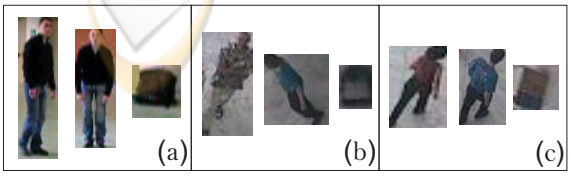

Figure 6: Segmented Object Samples using WGKVS. (a) MSA. (b) LeftBag. (c) LeftBox.

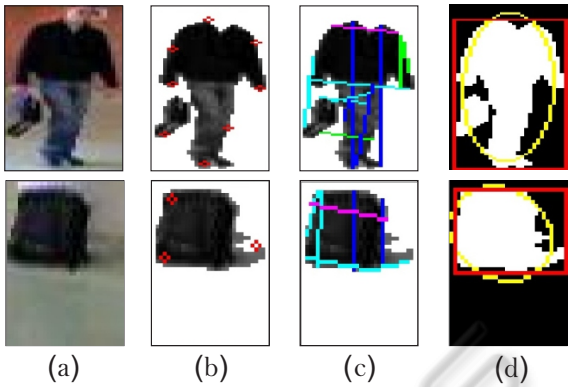

Figure 7: Geometrical Features Examples. (a) Original Object. (b) Corners. (c) Lines. (d) Fitting Shapes.

Table 5: Confusion Matrix using the Knn classifier.

	People	Baggage Objects
People	21	6
Baggage Objects	1	10

4 DISCUSSION

From the attained results of experiment one, it can be seen that when working only with the RGB components, the method does not perform very good, lacking of extra information that could enhance the clustering process (see Fig. 2(c) and Table 1). When the rgb components and the spatial information are incorporated, the performance improves by a 9.95% of the similarity measure (see Fig. 2(d) and Table 2). Using the proposed WGKVS methodology, the best results are achieved improving the similarity measure by 4.32% over the GKVS (Fig. 2(e) and Table 3). The results for the second experiment, expose that the proposed WGKVS methodology clearly surpass the attained results of the SOBS algorithm using its default parameters, and as can be seen in Figs. 4 and 5, our approach achieves more reliable results for further stages like the classification of objects. The obtained segmented objects by the WGKVS for the third experiment (see Fig. 6), are accurate for an adequate characterization process (see Fig. 7). The latter can be corroborated with a classification performance of 84.21%. The missclassified samples belonging to the people class, are objects where the complete body of the person is not in the scene.

5 CONCLUSIONS

We have proposed a methodology called WGKVS, which using image sequences recorded by stationary cameras, segments the moving objects from the scene. The aim of the proposed WGKVS is to construct a background model based on an optical flow methodology, and using a MKL background subtraction approach, incorporates different information sources, each source is weighted using a relevance analysis and a tuned Kmeans algorithm is used to segment the resulting weighted feature space. Experiments showed that the weighted incorporation of the spatial and rgb features enhances the data separability for further clustering procedures. Moreover, the attained results expose that the proposed WGKVS has stable results using the same parameters for all the experiments, and that it is suitable for supporting real surveillance applications like the classification of abbandoned objects. As future work, the inclussion of other features which could enhance the process and a methodology for the automatic actualization of the background model are to be studied. Furthermore, the proposed WGKVS is to be implemented as a real time application.

ACKNOWLEDGEMENTS

This research was carried out under grants provided by a MSc. and a PhD. scholarship provided by Universidad Nacional de Colombia, and the project 15795, funded by Universidad Nacional de Colombia.

REFERENCES

Chen, T.-W., Hsu, S.-C., and Chien, S.-Y. (2007). Robust video object segmentation based on k-means background clustering and watershed in ill-conditioned surveillance systems. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 787 –790.

Cuesta-Frau, D., Pérez-Cortés, J., and Andreu-Garcıa, G. (2003). Clustering of electrocardiograph signals in computer-aided holter analysis. *Computer methods and programs in Biomedicine*, 72(3):179–196.

Daza-Santacoloma, G., Arias-Londoo, J. D., Godino-Llorente, J. I., Senz-Lechn, N., Osma-Ruz, V., and Castellanos-Domnguez, G. (2009). Dynamic feature extraction: An application to voice pathology detection. *Intelligent Automation and Soft Computing*.

Elgammal, A., Duraiswami, R., Harwood, D., and Davis, L. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163.

Gonen, M. and Alpaydin, E. (2010). Localized multiple kernel regression. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*.

González, J. C., Álvarez-Meza, A., and Castellanos-Domínguez, G. (2012). Feature selection by relevance analysis for abandoned object classification. In *CIARP*, pages 837–844.

Gutchess, D., Trajkovics, M., Cohen-Solal, E., Lyons, D., and Jain, A. (2001). A background model initialization algorithm for video surveillance. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 733–740. IEEE.

Klare, B. and Sarkar, S. (2009). Background subtraction in varying illuminations using an ensemble based on an enlarged feature set. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 66 –73.

Maddalena, L. and Petrosino, A. (2008). A self-organizing approach to background subtraction for visual surveillance applications. *Image Processing, IEEE Transactions on*, 17(7):1168–1177.

Rakotomamonjy, A., Bach, F. R., Canu, S., and Grandvalet, Y. (2008). SimpleMKL. *Journal of Machine Learning Research*, 9:2491–2521.

Raty, T. (2010). Survey on contemporary remote surveillance systems for public safety. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(5):493–515.