Background subtraction by modeling pixel and neighborhood information

Shu-Jhen Fan Jiang[3], Kahlil Muchtar[1], Chih-Yang Lin[1], Li-Wei Kang[2] and Chia-Hung Yeh[3]

[1]Asia University, Taichung, Taiwan
E-mail: andrewlin@asia.edu.tw
[2]Academia Sinica, Taipei, Taiwan
E-mail: lwkang@iis.sinica.edu.tw
[3]National Sun Yat-sen University, Kaohsiung, Taiwan
E-mail: yeh@mail.ee.nsysu.edu.tw

Abstract— **In applications of the computer vision field, a vision system is usually composed of several low level and high level components, stacked on top of each other. A better design of the lower level components usually results in better accuracy of higher level functions, such as object tracking, face recognition, and surveillance. In this paper, we focus on the low level component design, background construction, which is one of the most basic elements for a surveillance system. The proposed method eases the problems that usually occur in background construction, including aperture problem, vacillating background, and shadow removal. In conventional background construction methods, only the history information (vertical direction) of pixels is usually considered. In contrast, the proposed scheme not only uses the vertical direction but also the neighborhood information (horizontal direction). Experimental results show that the proposed scheme can detect objects more delicate, alleviate the aperture problem, and identify shadow and discard it from detected objects.**

I. INTRODUCTION

In video surveillance systems, background subtraction is strongly needed in order to detect moving objects. There are numerous papers that have proposed different approaches in order to achieve the best result. They can be classified in following types; basic background modeling [1, 2, 3], statistical background modeling [4, 5, 6], fuzzy background modeling [7, 8, 9, 10, 11], background clustering [12, 13], neural network background modeling [14, 15], wavelet background modeling [16] and background estimation [17, 18, 19].

One of the promising papers, which employs a statistical approach and has inspired subsequent papers, is Stauffer and Grimson's method [5] that was proposed in 1999. Unlike a single Gaussian approach [4], they use a mixture of Gaussian models for modeling backgrounds. Such an approach can cope with a multimodal background and successfully detect moving objects in a vacillating background.

Nevertheless, since pixel processing was employed, shadows of moving objects will be classified as foreground and aperture problems are still unsolved. In addition moving objects will have a coarse texture in some cases when some border pixels of detected object will be regarded as background pixels or vice versa. Recently, Schindler and Wang [20] proposed a method, which considers not only the pixel itself but also its corresponding neighbors to acquire a smooth result. Furthermore, Liu, et al. [21] proposed a hierarchical manner using MRF (Markov Random Field) to cast shadows from the foreground. Instead of using a likelihood function for getting information from neighbors, our proposed method uses a modest way to model certain neighbor pixels by using Gaussian mixture models to determine the probability of that pixel being in the background.

The rest of paper is organized as follows. Firstly, we introduced the renowned Grimson and Stauffer's method that henceforth is called Gaussian Mixture Models (GMM). Secondly, our proposed method by merging pixel and neighborhood information will be unveiled in Section III. In Section IV, results and discussions will be depicted by providing quantitative and qualitative comparisons. In addition, conclusion will be provided in the last section.

II. THE GAUSSIAN MIXTURE MODEL

The Gaussian Mixture Model (GMM) was proposed by Grimson and Stauffer [5]. The authors proposed a time series method to model each background pixel into a certain number (K) of GMM. Typically K is a small number from 3 to 5. The weight associated with each Gaussian represents the portion of the data accounted for that Gaussian.

Formally, in the GMM model, each pixel in the scene is modeled by a mixture of K Gaussian distributions. The probability that a pixel is regarded as having a value X_t at a certain time is given as follows [5]:

$$P(X_t)=\sum_{j=1}^{K}\omega_{j,t}*\eta(X_t,m_{j,t},\sum{}_{j,t}), \tag{1}$$

where K is the number of Gaussian distributions, $\omega_{j,t}$ is the weight estimation of the j_{th} Gaussian in the mixture at time t, $m_{j,t}$ and $\sum_{j,t}$ are the mean value and covariance matrix respectively, of the j_{th} Gaussian in the mixture at time t, and η is a Gaussian *pdf* (probability density function), defined in Eq. (2).

$$\eta(X_t, m, \sum) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - m_t)^T \Sigma^{-1}(X_t - m_t)} \quad (2)$$

For computational efficiency, $\sum_{k,t}$ is defined as $\sigma_k^2 I$ to represent the covariance of the k^{th} model component.

Since the probable background pixel values should stay longer and be more invariant than others, the corresponding Gaussians have the most supporting evidence and the least variance. This model was widely used in a real-time mode accompanied by an adaptive update process. When a new pixel comes in, it is checked against existing model components. The new pixel is said to match one of the weighted Gaussian distributions if its pixel value is within 2.5 standard deviations of the matched distribution. If any of the models are matched, the matched distribution will be updated. Otherwise (*i.e.*, none of the K Gaussian distributions can be matched to the current pixel value), the distribution that has the minimum weight is replaced with a distribution using the current value as its mean value, an initially high variance, and a low prior weight.

In the maintenance of the background model, the K distributions are sorted based upon the value ω/σ. The first B distributions are selected as the background model of a pixel for the scene and denoted as:

$$B = \arg\min_b \left(\sum_{k=1}^{b} \omega_k > T_B\right), \quad (3)$$

where T is a predefined threshold that represents the minimum quantity of the data that must be accounted for in the background model. T is usually set to about 90% in many applications.

After determining whether the incoming pixel will match one of the existing Gaussian distributions, the prior weights of K Gaussian distributions are updated as follows:

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t}), \quad (4)$$

where α is the learning rate and $M_{k,t}$ is 1 for the matched distribution and 0 for the unmatched distribution. Subsequently, weights for distributions are renormalized. If the incoming pixel matches a Gaussian distribution, the values of *mean* and *variance* of this distribution are updated as follows:

$$m_t = (1-\rho)m_{t-1} + \rho X_t, \quad (5)$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t - m_t)^T (X_t - m_t), \quad (6)$$

where

$$\rho = \alpha\eta(X_t \mid m_k, \sigma_k). \quad (7)$$

III. NEW BACKGROUND MODELING METHOD

In this section, we describe the proposed method and shadow removal.

From the previous section, conventional GMM only considers individual pixels to determine background and foreground value. Apart from taking pixel modeling into account, our proposed method also considers certain number of neighbors that are directly adjacent to the observed pixel. As [20] depicted, the spatial distribution has valuable information to support accurate subtraction. Unlike Schindler et al. [20] and Liu, et al [21] that exploit MRF (Markov Random Field) to achieve the smoothness value, we extend prominent GMM by involving certain neighborhood information.

As [22] explained, a pixel q at coordinates (x,y) has four direct neighbors, denoted by $N_4(q)$, that are horizontal and vertical neighbors with coordinates $(x+1, y)$, $(x-1, y)$, $(x, y+1)$, and $(x, y-1)$.

Besides horizontal and vertical neighbors, diagonal values can be considered as well. This type of neighbors is called 8-neighbors of q, denoted by $N_8(q)$, and is illustrated in Fig.1.

$x-1, y-1$	$x-1, y$	$x-1, y+1$
$x, y-1$	x,y	$x, y+1$
$x+1, y-1$	$x+1, y$	$x+1, y+1$

Fig. 1. Illustration of $N_8(q)$.

By considering those neighbors' information (n by n) for q and each block mean m_N measurement is depicted as follows,

$$m_N = \frac{1}{n \times n} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}, \quad (8)$$

Once we have m_N, the probability of q when comparing with a neighborhood model is shown as follows,

$$P(q) = \sum_{j=1}^{K} \omega_{j,t} * \eta(q, m_{N(j,t)}, \sum{}_{j,t}), \quad (9)$$

If q matches to one of the neighborhood Gaussians, the binary result of q is determined by the dominant binary value of N_8. If N_8 has a foreground dominant value, which means the number of 1 is more than that of 0 in N_8, the binary result of q will be regarded as foreground, or vice versa. With this approach, neighborhood information can give strong evidence to determine if the observed pixel is a background or foreground pixel. However, if none of the neighborhood Gaussian models match the observed pixel, the existing binary value of q from the pixel-based GMM is maintained.

The update process for the neighborhood model is shown as follows,

$$m_{N(t)} = (1-\rho)m_{N(t-1)} + (\rho)m_{Nt}, \quad (10)$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(q - m_{N(t)})^T (q - m_{N(t)}), \quad (11)$$

The remainder of equation would be the same as conventional GMM.

It should be noticed that for a border pixel q of the image, some neighbors of q will lie outside. In such a case, we only take pixel q that has complete neighbors into account; in other words, the border pixels will use only the pixel-based GMM to determine its binary results. The entire flowchart of our proposed method is shown in Fig. 2.

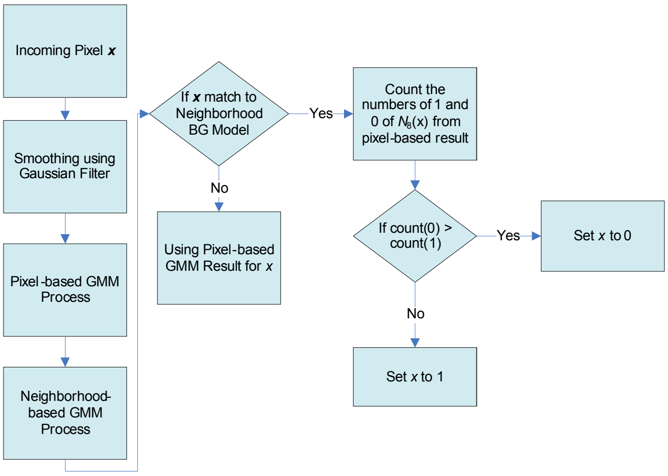


Fig. 2. Flowchart of proposed method.

One of the common problems in background modeling is incorrect labeling when a shadow is classified as foreground. This problem takes place since the *RGB* (Red, Green, Blue) color space contains chromaticity and luminance together [22]. Based on this reason, separation of chromaticity and luminance is imperative in order to model luminance independently. Furthermore, a shadow will have a certain portion of light and thus a certain range can be yielded by deciding a pre-defined threshold for luminance changing. Thereby, instead of using *RGB* color space, a new color space (r, g, I) will be applied to remove shadows. When the color and intensity have been separated from the (r, g, I) color space, the shadow will only alter intensity without changing r and g channel [20].

IV. Experimental Results

The performance of the proposed method is compared with Stauffer and Grimson's method [4], using several video sequences. The video sequences were acquired from outdoor environments and a public dataset of common surveillance cases. The simulated environment for the experiments was equipped with a 2.93 GHz Core 2 Duo Intel processor and 2 GB of memory. The image resolution was set to 320×240 pixels. All algorithms were implemented in C++.

For the sake of labeling and segmenting the foreground pixels, the connected components algorithm [24, 25] was applied to each background modeling method. The parameters used in the experiments are listed in Table 1, where α is the learning rate, TH_B is used in (3), K denotes the number of Gaussians, N represents neighbor size, β and γ are used to define the range of shadow intensity.

CAVIAR public dataset [26] is used to measure qualitative results. Furthermore, a similar dataset is used to calculate quantitative results that will be represented by these following criteria [27]

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{P} \tag{13}$$

$$\text{Similarity} = \frac{TP}{TP+FN+FP} \tag{14}$$

$$F\text{-measure} = \frac{2}{(1/\,precision + 1/\,recall)} \tag{15}$$

It should be noticed that TP, FP, FN, TN, P and N are True Positive, False Positive, False Negative, True Negative, Total of Positive and Total of Negative value respectively.

TABLE I
The parameter values used in the experiments

Parameters	α	TH_B	K	N	β	γ
Stauffer and Grimson's method	0.005	0.9	3	-	-	-
Proposed method	0.005	0.9	3	3 x 3	0.12	0.21

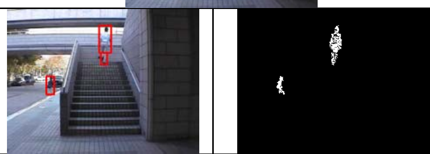
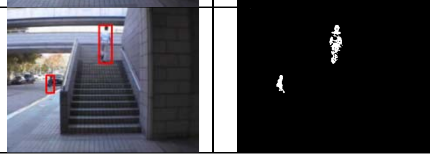
Type/Video set	Frame 74
Original Frame	
Original GMM	
GMM using neighborhood information and RGI color space	

Fig. 3. Comparison using outdoor scene dataset.

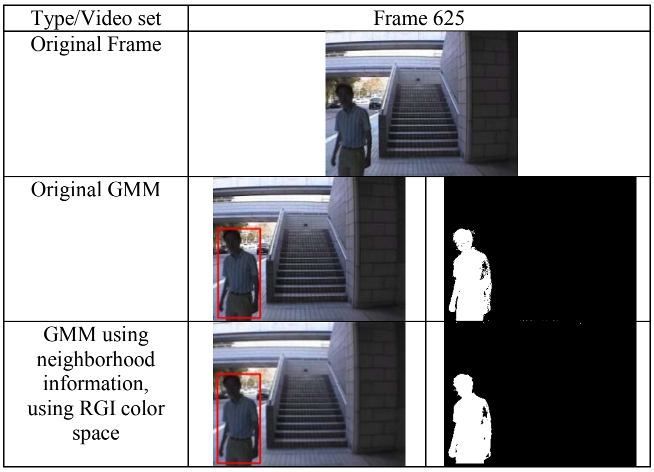


Fig. 4. Other comparison using outdoor scene dataset.

From the above outdoor case (Figs. 3 and 4), it can be seen that the proposed method can yield a more delicate foreground and at the same time alleviate the aperture problem. By combining pixel and neighborhood information, incorrect labeling can be relatively reduced.

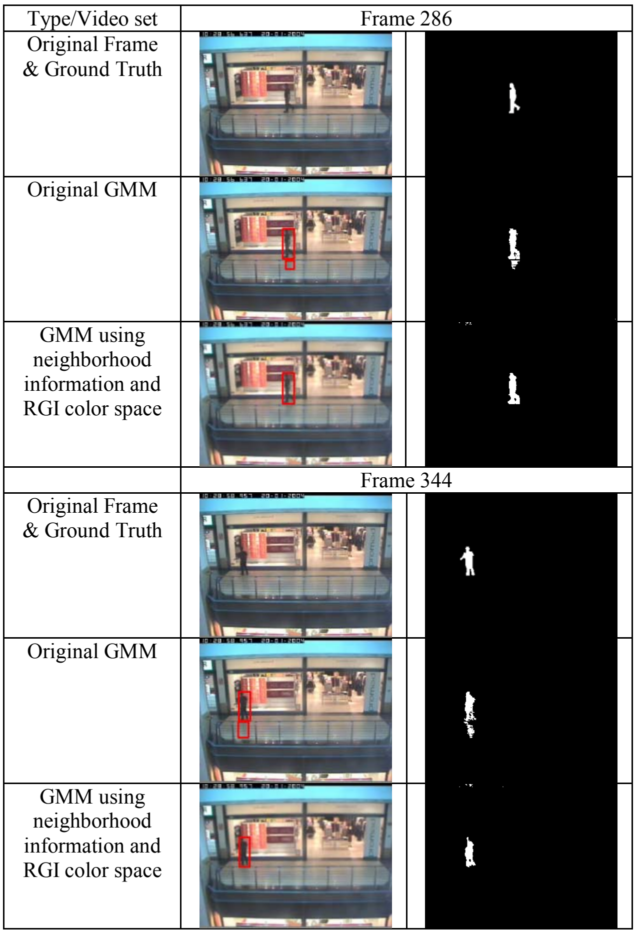


Fig.5. Qualitative result of CAVIAR public dataset; **OneLeaveShopReenter2front**.

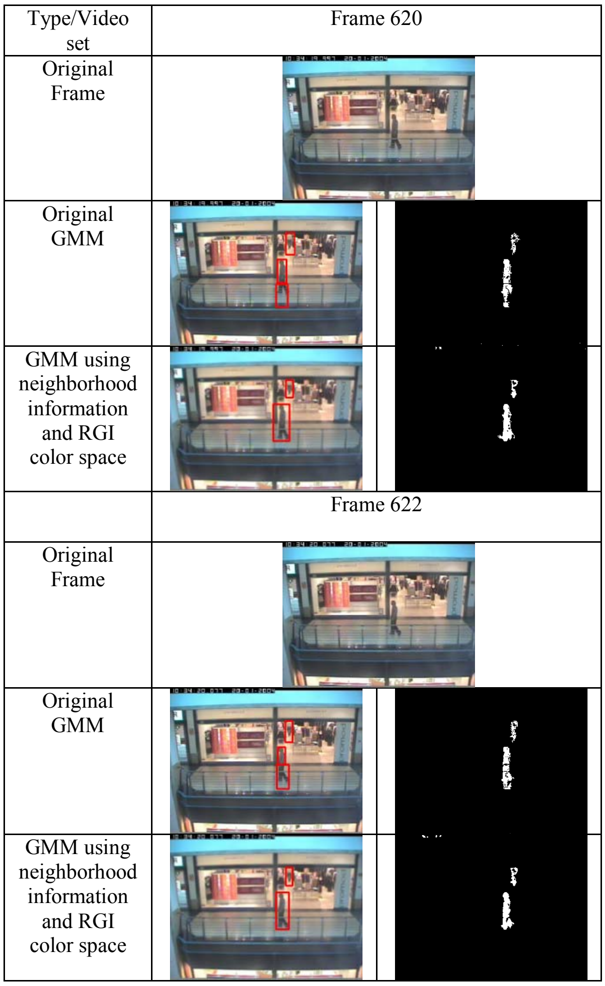


Fig.6. Qualitative result of CAVIAR public dataset; **OneShopOneWait2front**.

Two different types of CAVIAR public datasets (Figs.5 and 6) show how the proposed method can significantly discard the shadow and enhance the shape of detected objects. As mentioned in the previous section, the *r, g, I* color space treats color and intensity independently. Two pre-defined thresholds are used to range the common intensity of shadow. From a qualitative perspective, we can see that the proposed method can obtain results that are almost similar to the ground truth of the observed frame.

TABLE II
1st CAVIARDATA1 public dataset [26] quantitative result
Dataset type: One Leave Shop Reenter 2 front

Frame Number: 286				
Method/Measurement Unit	Recall (TP Rate)	Precision	Similarity	F-Measure
Original Stauffer & Grimson Method	0.971616	0.628531	0.617198	0.763293
Proposed Method using RGI color space	0.982533	0.696594	0.688073	0.815217

TABLE III
2nd CAVIARDATA1 public dataset [26] quantitative result
Dataset type: One Leave Shop Reenter 2 front

Frame Number: 344				
Method/Measurement Unit	Recall (TP Rate)	Precision	Similarity	F-Measure
Original Stauffer & Grimson Method	0.823105	0.631579	0.556098	0.714734
Proposed Method using RGI color space	0.794224	0.847784	0.695103	0.82013

Precision can be seen as a measure of exactness, whilst recall is a measure of completeness. The *F*-measure considers both the precision and the recall in computing the score, which can be interpreted as a harmonic mean of precision and recall. Tables 2 and 3 depict the outperformed results when comparing the proposed method with the conventional GMM.

V. Conclusions

A collaboration of pixel and neighborhood based model has been presented. Based on qualitative and quantitative results, we can deduce that the neighborhood information, even at low level vision, is imperative to enhance object shapes and to alleviate aperture problems. By using a new color space (*r, g, I*) of conventional GMM, the results demonstrate that the shadow problem has been successfully relieved.

Acknowledgment

This work was supported by National Science Council, Taiwan, under Grant NSC 101-2221-E-468-021.

References

[1] Lee B., Hedley M. Background Estimation for Video Surveillance, IVCNZ 2002, pages 315-320, 2002.
[2] McFarlane N., Schofield C. Segmentation and tracking of piglets in images, BMVA 1995, pages 187-193, 1995.
[3] Zheng J., Wang Y., Nihan N., Hallenbeck, E. Extracting Roadway Background Image: A mode based approach, Journal of Transportation Research Report, No 1944, pages 82-88, March 2006.
[4] Wren C., Azarbayejani A., Darrell T., Pentland A. Pfinder: Real-Time Tracking of the Human Body, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, No. 7, pages 780-785, July 1997.
[5] Stauffer C., Grimson W. Adaptive background mixture models for real-time tracking, CVPR 1999, pages 246-252, 1999.
[6] Elgammal A., Harwood D., Davis L. Non-parametric Model for Background Subtraction, ECCV 2000, pages 751-767, Dublin, Ireland, June 2000.
[7] Sigari M., Mozayani N., Pourreza H. Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application, International Journal of Computer Science and Network Security, Volume 8, No. 2, pages 138-143, 2008.
[8] El Baf F., Bouwmans T., Vachon B., Type-2 fuzzy mixture of Gaussians model: Application to background modeling, ISVC 2008, pages 772-781, Las Vegas, USA, December 2008.
[9] Zhang H., Xu D Fusing Color and Texture Features for Background Model, International Conference on Fuzzy Systems and Knowledge Discovery, Volume 4223, No. 7, pages 887-893, September 2006.
[10] El Baf F., Bouwmans T., Vachon B., “Fuzzy Integral for Moving Object Detection”, FUZZ-IEEE 2008, pages 1729-1736, Hong-Kong, China, June 2008.
[11] Sivabalakrishnan M., Manjula D., Adaptive Background subtraction in Dynamic Environments Using Fuzzy Logic, International Journal on Computer Science and Engineering, Volume 02, No. 2, pages 270-273, 2010.
[12] Butler D., Sridharan S. Real-Time Adaptive Background Segmentation, ICASSP 2003, 2003.
[13] Kim K., Chalidabhongse T., Harwood D., Davis L. Real-time Foreground-Background Segmentation using Codebook Model, Real-Time Imaging, 2005.
[14] Culbrik D., Marques O., Socek D., Kalva H., Furht B. Neural network approach to background modeling for video object segmentation”, IEEE Transaction on Neural Networks, Volume 18, No. 6, pages 1614–1627, 2007.
[15] Maddalena L., Petrosino A. A self organizing approach to background subtraction for visual surveillance applications, IEEE Transactions on Image Processing, Volume 17, No. 7, pages 1729–1736, 2008.
[16] Biswas S., Sil J., Sengupta N. Background Modeling and Implementation using Discrete Wavelet Transform: a Review, JICGST-GVIP, Volume 11, Issue 1, pages 29-42, March 2011.
[17] Toyama K., Krumm J. Brumitt B., Meyers B. Wallflower: Principles and Practice of Background Maintenance, International Conference on Computer Vision, pages 255-261, Corfu, Greece, September 1999.
[18] Messelodi S., Modena C., Segata N., Zanin M. A Kalman filter based background updating algorithm robust to sharp illumination changes, ICIAP 2005, Volume 3617, pages 163-170, Cagliari, Italy, September 2005.
[19] Chang R., Ghandi T., Trivedi M., Vision modules for a multi sensory bridge monitoring approach, ITSC 2004, pages 971-976, October 2004.
[20] Schindler K, Wang H. Smooth Foreground-Background Segmentation for Video Processing, ACCV 2006, Hyderabad, India, Volume 3852, pages 581-590, January 2006.
[21] Liu Z, Huang K, Tan T. Cast Shadow Removal in Hierarchical Manner Using MRF, IEEE Transactions on Circuits and Systems for Video Technology, Volume 22, No.1, January 2012.
[22] Digital Image Processing, Gonzalez and Woods, 2nd Edition, 2002.
[23] Ramesh Jain, Rangachar Kasturi and Brain G. Schunck, “Machine Vision”, McGraw-Hill Book Co., Chapter 10, 1995.
[24] F. Chang, C. J. Chen, and C. J. Lu, “A Linear-Time Component-Labeling Algorithm Using Contour Tracing Technique,” Computer Vision and Image. Understanding, vol. 93, no. 2, pp. 206-220, 2004.
[25] R. Lumia, L. Shapiro, and O. Zuniga, “A New Connected Components Algorithm for Virtual Memory Computers,” Computer Vision, Graphics, and Image Pmcessing, vol. 22, no. 2, pp. 287-300, 1983.
[26] EC Funded CAVIAR project/IST 2001 37540, found at URL: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
[27] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol.27, pp. 861-874, 2006.