SHORT TERM RE-IDENTIFICATION OF AUTOMATIC TELLER MACHINE (ATM) USERS VIA FACE AND BODY APPEARANCE FEATURES

*Ekberjan Derman, Güney Kayım**

Provus Information Technologies
Şişli - Istanbul, TURKEY
{ekberjanderman,guneykayim}@gmail.com

Albert Ali Salah

Boğaziçi Univ., Dept. Computer Engineering
Bebek - Istanbul, TURKEY
salah@boun.edu.tr

ABSTRACT

This paper describes a novel biometric scenario, where a person is authenticated at an ATM machine, and has to be re-identified from a camera within a very short time period, under very challenging illumination and pose conditions, and using data from a single session. The application scenario is the automatic retraction of forgotten card or cash at an ATM, which happens frequently, and causes inconvenience to users and loss of profits for the banks. We propose a multimodal authentication system that operates under the constraints imposed by this applications scenario, and implement face recognition and color based body appearance recognition to create a system that improves ATM behavior in case of forgotten card or cash by re-identifying the user from an embedded ATM camera. We focus on the scenario and the platform, and report tests with the proposed system under challenging conditions, obtained from ATMs placed in the field.

1. INTRODUCTION

Banks seek to reduce their infrastructure costs by shifting transactions of their customers to Automatic Teller Machines (ATMs) and Internet websites. Financial users especially prefer ATMs for physical transactions, like cash withdrawal or cash deposit. For these reasons, user experience at the ATM is a very important concern for the banks.

One of the issues that ATMs suffer from is card and/or cash forgetting (CCF), which is a surprisingly common situation. In CCF, the user forgets the card or cash after the transaction, and leaves the system. After a certain waiting period, these items will be swallowed by the ATM, and the user has to go through a tedious and costly process to retrieve the card/cash or have the card reissued. Moreover, the cash is stored in a separate container after CCF, and needs to be manually checked before it is returned to circulation. This means a lot of time and money is being wasted to cope with the consequences of CCF. According to statistical data available from MasterCard, a single ATM will have about 8-9 card retractions and 16-17 cash retractions per year. In a local banking sector, the number of ATMs can be in the order of tens of thousands, and the extra cost of CCF amounts to several millions of USD per year for a single country [4].

The CCF issues can be somewhat mitigated by allowing the ATM to re-identify the customer returning to the ATM to get back the forgotten items. This requires that the swallowing of the items should be delayed, and the next customer is monitored and checked to be a returning customer, or a novel customer. We describe in this paper how a camera-based system can monitor customers to improve ATM behavior.

In the proposed system, as soon as a customer inserts the card into the ATM and a session is initiated, the system starts face and body appearance detection using the camera located near the ATM and builds a temporary identity database for the customer. If the customer leaves the ATM without taking his/her card or cash, the ATM waits for the customer to be back, instead of retracting the forgotten item. If the system finds out there is a different customer approaching the ATM, the item is retracted instantly.

This scenario is fundamentally different than biometric authentication scenarios, in which a person's image is matched to a gallery image acquired, possibly, a long time before the matching and under different conditions. In this scenario, the matching image and the gallery image are separated by mere minutes at most. The main difficulties in our scenario are the uncontrolled and low quality real world images coming from the attached ATM camera (especially the facial images of the returning card holder or another approaching customer), and the extreme lighting conditions. Moreover, the processing and decision making should be carried out in a short time-frame to be of practical use.

The remainder of this paper is organized as follows. Section 2 provides an overview of the related work. Section 3 introduces the overall system design, Section 4 describes the methodology, including face and body appearance detection and recognition, Section 5 presents the experiments and eval-

*This work is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) with project number 3120779. The first two authors contributed equally to this work. Cite this paper as Derman, E., G. Kayım, A.A. Salah, "Short Term Re-Identification of Automatic Teller Machine (ATM) Users via Face and Body Appearance Features," Int. Workshop on Biometrics and Forensics, Limassol, March 2016.

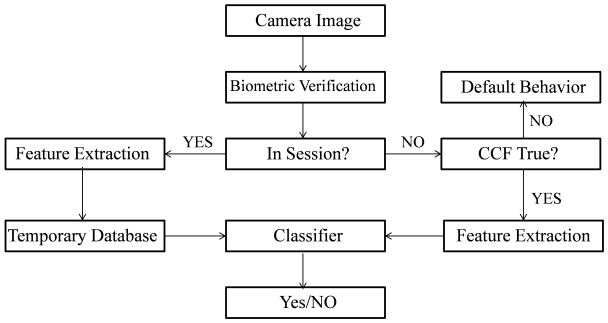

Fig. 1. General flowchart of the proposed system.

uates the system performance. Finally, Section 6 concludes the paper.

2. RELATED WORK

Since our focus is on the application related requirements of the authentication scenario, we do not describe related work in face authentication from video. While the face modality has been studied widely for biometric purposes, it is relatively less considered for ATM-based usage. The main difficulty is that ATMs are typically placed outdoor, and operate under widely changing illumination conditions.

Babaei et al. presented a concept of using face recognition for ATM users together with other biometric features like fingerprint, iris recognition and hand/finger geometry [1]. However, they did not provide any specific method for an actual ATM user scenario. Aru et al. proposed an ATM security model that would combine a physical access card, a PIN, and face recognition to increase the reliability of ATM transactions [2]. Peter et al. presented a face verification based method to improve ATM security [3]. In their proposed system, face verification is performed on the still images of the ATM user, and compared with images of a gallery to make a decision. However, this approach depends on a pre-generated gallery, which typically involves images acquired a long time before the actual use, and under different illumination conditions.

In [4], Derman et al. proposed an approach for actual CCF scenarios for frontal face images captured by an embedded ATM camera. Their work was the first framework to focus on the CCF scenario, but used only a single biometric, namely the face, for authentication. Due to difficult illumination conditions in the field, the facial image by itself produced poor results. In this work, we operate in a similar application scenario, but use profile facial images instead frontal faces, and we also propose fusing face and body appearance verification results. Since our scenario requires the initial acquisition and re-identification to happen within a very short period (within 1 minute at most), the illumination conditions are mostly stable, but not uniform or controlled.

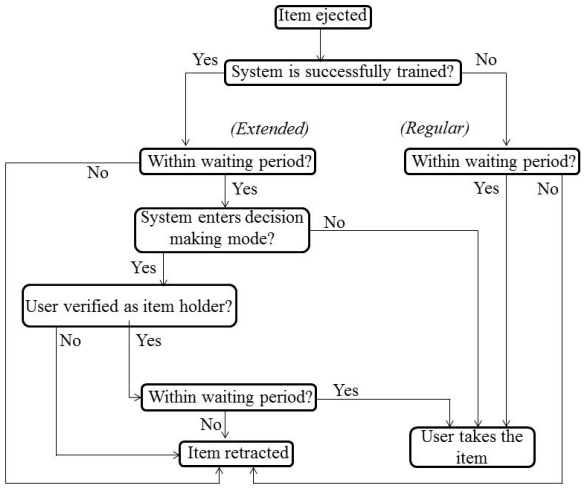

Fig. 2. All possible scenarios after an item is ejected by ATM.

3. SYSTEM DESIGN

The system model corresponding to the investigated ATM and CCF scenario is depicted in Figure1. In our work, when the user starts to interact with the ATM, we first collect user images from the mounted ATM camera and then perform face detection to crop the face region out of its surroundings. Following that step, a region of interest (ROI) is set for modeling body appearance. We use color based descriptors for this part. The ROI is determined relative to the face area. The system monitors whether the current status is "in session", that is, whether the user is still making transaction with the ATM. While "in session", the system performs feature extraction to build a temporary face and body appearance database for the user (i.e. the holder of the card) using the features of retrieved face and body appearance images. On this temporary database, a classifier is trained. The classifier enters decision mode only when it is successfully trained, that is, when it has at least a sufficient number of samples. When the "In Session" status is updated as "No", that means the user has left the ATM. If the transaction is completed normally, the temporary database is discarded. We consider here only cases when the user has forgotten to take card and/or cash after the transaction. In this case, the system processes the first person arriving to the system to match them with the temporary database. Feature extraction is performed based on these new "out-of-session" face and appearance images and then these images are delivered to the classifier. If the final result indicates that the "out-of-session" images belong to the card holder, then the ATM will wait instead of retracting the card or the cash. Otherwise, the system decides that a new user is to be served, and the ATM immediately retracts card/cash for protection. The current ATM system, when faced with a CCF event, waits for a preconfigured period (typically 20-30 seconds) to retract the forgotten item. In our system, we extend

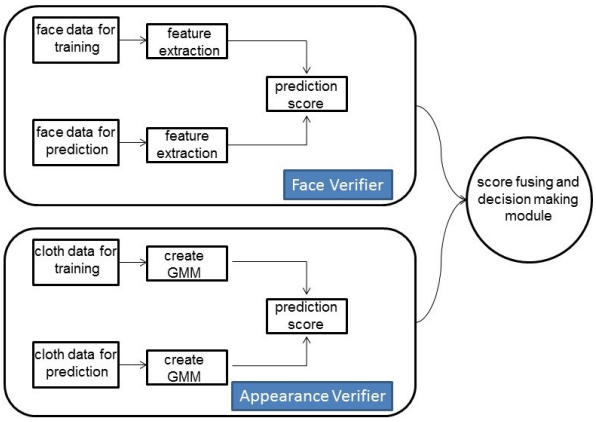

Fig. 3. Score level fusion of the two different modalities

this period to 60 seconds and after that let the ATM retract the item for security. However, if any person approaches the ATM within this period, then our system enters the decision making mode. In this case, the possible scenarios can be described as depicted in the Figure 2.

4. METHODOLOGY

4.1. Verification System

The verification system consists of *training* and *prediction* modes. It employs a face verification module, and an appearance-based module, whose outputs are fused at the matching score level (as shown in Figure 3). We report our system's performance in terms of *false acceptance rate* (FAR), *genuine acceptance rate* (GAR) and *receiver operating characteristic* (ROC) curves.

4.2. Face Detection

Face detection on the captured image is performed using OpenCV's built-in Haar cascade profile face detector[1]. Since the profile detector is trained for left-side profiles, the captured frame should be flipped according to the actual situation. Considering the fact that the ATM user's head position may vary (as shown in Figure 4), to maximize the face detection rate, we perform a rotation on the captured frame before training the classifier.

The best rotation angle for maximizing the face detection rate is obtained as:

$$C_{max} = \arg\max_{\theta_{best}} C(\theta_{best}), \quad (1)$$

where C is the number of faces detected under rotation angle θ, and possible values in 5^o increments are considered for this angle between $0^o - 40^o$.

[1] **http://www.opencv.org**

Fig. 4. Users with different head positions while using ATM

Fig. 5. Left: original image, Right: image flipped and rotated with best rotation angle. Red rectangle shows the detected face region

4.3. Preprocessing

We used several preprocessing operations. The captured RGB image is first converted to grayscale. The profile face detector operates with some margin and includes background, which is discarded with a fixed mask. The remaining image is further smoothed with a Gaussian filter, and resized into 64x64 pixels. Finally we perform an histogram equalization.

4.4. Face Verification

1) *Feature extraction*: After the preprocessing step, the Local Binary Pattern (LBP) operator is used as our face descriptor [6, 7, 8]. Here, we use uniform binary-based LBP with eight sampling points within one pixel neighborhood. Then, we represent the face region with m subregions R_0, R_1, ..., R_{m-1}, and compute a histogram independently within each of these regions. Finally, we combine the resulting m histograms to yield the spatially enhanced histogram vector.

2) *Classification*: We use Support Vector Machines (SVM) as our classifier [10]. Feature vectors of the ATM user's "in-session" images are used as positive samples, while feature vectors of face images obtained previously are used as negative samples for classifier training. The negative samples can be stored on the system. The trained sample size must meet the predefined minimum sample number to successfully train the classifier. Out-of-session images of the ATM user are used for the prediction step. The calculated signed distance value is transferred into a score between 0 and 1 using a sigmoid function [11]:

$$\alpha = D * A + B \quad (2)$$

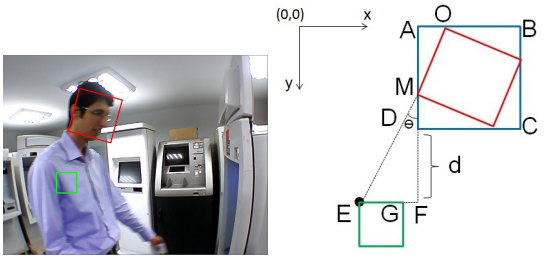

Fig. 6. Selection of the appearance region using parameters from face verification module

$$S = \frac{1}{1 + exp(-\alpha)} \tag{3}$$

where α represents the sigmoid factor, *D* represents the obtained signed distance by SVM after prediction, *A* and *B* are predefined parameters. We obtained values of *A* and *B* as -5.2 and -0.5 by using a non-linear regression method [12].

4.5. Appearance verification

In a conventional CCF incident, we expect to see the user returning the ATM with the same appearance as in the transaction step, since at most one minute is passed between the acquisition of the gallery image and the test image. In the training step, the system learns the background information of the appearance region (of size 30x30 pixels) by using a Gaussian Mixture Model (GMM). During the prediction step, the system decides whether the given appearance region belongs to the background or not. This result can further be transformed into a prediction score between zero and one by:

$$S = 1.0 - \frac{N_{foreground}}{N_{total}} \tag{4}$$

where S, $N_{foreground}$ and N_{total} refer to prediction score, foreground pixel number, and total pixel number, respectively.

Selection of a valid appearance region (both during training and during operation of the system) is carried out as follows (Figure 6). We first obtain the user's face region (red rectangle), its bounding box (blue rectangle) and best rotation angle θ. Then the coordinates (E_x, E_y) of the original point (point E) of the cloth region (green rectangle) are computed:

$$E_x = A_x - [(D_y - M_y) + d] \tan(\theta) \tag{5}$$

$$E_y = A_y + AD + d \tag{6}$$

Here, *d* is a pre-set parameter. Using validation set results on real data, we empirically set *d* as:

$$d = \begin{cases} 450 & \text{if } E_y \leq 450 \\ 450 - A_y - AD & \text{if } E_y > 450 \end{cases} \tag{7}$$

4.6. Fusion

The scores obtained by face and appearance verification are combined together by the Sum Rule [5]. Scores of different modalities are assigned a weight value between 0 and 1. This combination form can be expressed as follows:

$$S_{fused} = S_{FV} \times W_{FV} + S_{CV} \times W_{CV} \tag{8}$$

$$W_{FV} + W_{CV} = 1 \tag{9}$$

$$W_{FV}, W_{CV} \in \mathbb{R}^p, 0 \leq W_{FV}, W_{CV} \leq 1 \tag{10}$$

$$S_{FV}, S_{CV}, S_{fused} \in \mathbb{R}^p, 0 \leq S_{fv}, S_{CV}, S_{fused} \leq 1 \tag{11}$$

in which, S_{fused}, S_{FV}, S_{CV}, W_{FV} and W_{CV} indicate the final fused score, the face verification score, the cloth verification score, the weight of face verification, and the weight of cloth verification, respectively. The final fused score, S_{fused}, is compared with a predefined threshold value, *h*. If it is higher than *h*, classifier regards its corresponding user as matching identity, or else, as impostor.

5. EXPERIMENTAL VALIDATION

5.1. Dataset Collection

We constructed our own database, which consists of real ATM interaction sessions recorded from 57 subjects. These subjects were given a pre-specified scenario as follows: First, the user starts to interact with the ATM for a banking transaction involving cash withdrawal. Once the user withdraws the cash, he or she leaves the ATM without getting the card. Then after some short period, the user realizes that the card was forgotten, and returns back to the ATM to retrieve the card. The average time for the system to be able to collect the "out-of-session" images for one user is only about two seconds, since the user returns and gets the forgotten card quickly once entering the camera range. Also, the size and quality of the "out-of-session" images are very low, due to the fact that the distance between the user and the ATM camera is changing. This motion blur creates the biggest challenge for the face authentication task in our case, and it is one of the principal reasons why our scenario is significantly different compared to an ordinary face authentication scenario (as shown in Figure 6). Currently, we only consider cases with a single ATM user. New data are collected for the case of multiple users.

The number of "in-session" images for the 57 subjects of our database varies between 1 and 1000, depending on the time they spent for their transactions and their head pose, while that of "out-of-session" images varies between 1 and 29, depending on their speed of return to the ATM, and the time they spend for getting the forgotten card. Some users may immediately take back their cards, while others may spend several seconds. Both males and females are included in our test. The height of our test subjects are between 160

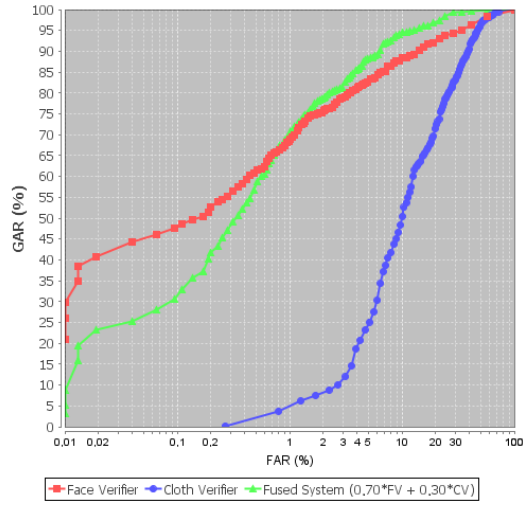

Fig. 7. ROC curves of the proposed system

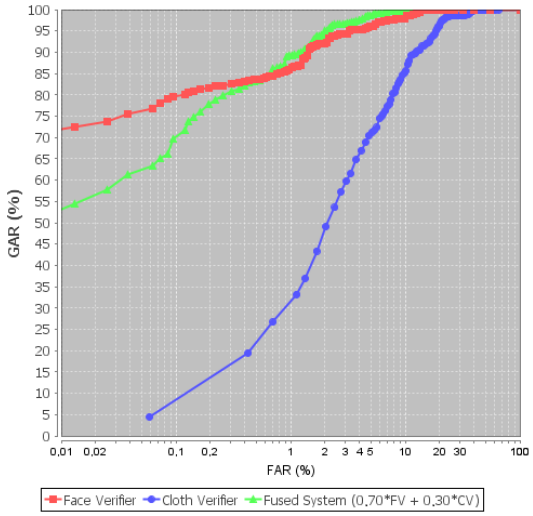

Fig. 8. ROC curves using "in-session" images for prediction

cm and 190 cm, so that users with different heights are considered. This is important, because the camera is static, and very tall and very short subjects can provide inadvertently cropped face images.

5.2. Evaluation Measures

Genuine and impostor scores are created for both face and appearance verification modules. Later, these scores are fused to get final decision. We decide different threshold values for getting FAR and GAR values to obtain ROC curves, which indicate the system's performance limitation, to evaluate each module's independent and fused performance.

5.3. Experiments and Test Scenarios

1) *Two class classification*: We divide the "out-of-session" images into two classes, as positives (matching identities) and negatives (impostors). We tested our proposed system on our database of 57 subjects. Module weight W_{FV} is selected as 0.7, the GMM to detect the background uses 10 clusters, cloth region size is set to 50x50, the minimum number of images for positive and negative classes is 100 for training, and the maximum number of images are 300 for the positive class (obtained from camera during session), and 600 for the negative class (stored offline).

While providing negative samples for each user, we choose a total of 26 other people's "in-session" face images for training, and the remaining 26 other people's "out-of-session" face images are used for impostor accesses. The resulting ROC curves of GAR and FAR values are as shown in Figure 7.

2) *"In-session" vs. "Out-of-session"*: The quality of the "out-of-session" images are very low compared to those of "in-session" images. We also tested the algorithm for "in-session" images that are not used in the training process to see the effect of image quality on algorithm performance. The corresponding ROC curves are shown in Figure 8.

The fused score threshold h is determined by experimental results and used to accept or reject the predicted new image. If the system only employs face verification and operates at a FAR value of 5%, we get a GAR value of 81.5%. If the system employs both face and appearance verification with an expected FAR value of 5%, we get a GAR value of 91%. The interesting aspect of the CFF scenario is that there can hardly be any intentional attacks on this system, as the frequency of CFF is very low. It is not feasible for an impostor to wait until someone forgets card or cash at an ATM. For this reason, a much higher FAR than an ordinary biometric scenario is acceptable in this application.

From these results, we can see that fused system gives better performance than any single module of face or appearance verification. For our final usage, depending on our expectation of FAR and GAR, we can select the actual threshold value from these results. In our case, the operation range permits a FAR around 5% for a GAR around 90%. Meanwhile, by comparing the results of Figure 7 with that of Figure 8, we observe that the increase in image quality results in an increase in the system performance. The performance results for good quality "in-session" images indicate that the proposed system gives acceptable results when the test images have sufficiently high quality.

To evaluate the computational requirements of our scheme in practical conditions during our experiments, we utilized a conventional system for ATMs operating in field. Specifically, the ATM we used for our test has an Intel Core 2 Duo 3.00GHz CPU and 2GB RAM, using Windows XP Professional Version 2002 Service Pack 3 as operating system.

Fig. 9. Failure case: face region covered by physical appearance (left) and by garment (right)

The calculation speed of our proposed system on this ATM is 15 frames per second (fps), which meets the processing requirement for timely response.

5.4. Analysis of Failure Cases

Face detection is vital for our proposed system, and the system fails when we cannot successfully obtain face images for related sessions. Some potential causes of such cases may include face region covered by user hair or clothes (as shown in Figure 9). In these situations, our proposed system is bypassed in case of any CCF incidents and resumes a fall-back scenario of conventional time-out based retraction, which is how the incumbent ATMs act currently. This means that when the system fails, the default behavior is exhibited. Any successful delivery of the card or cash to the user (real or impostor) can be logged.

6. CONCLUSION

In this work, we proposed a computer vision based ATM user identification framework using face and appearance verification to reduce card and cash retraction. We evaluated the proposed system under various conditions, and with our own database, based on a real scenario. The experimental results reveal that our proposed system is promising for mitigating the card/cash forgetting issue.

Our "out-of-session" test condition is the one that is closest to the expected real world application. It is more important in this application to keep a high true positive rate (convenience) as the impostors cannot be expected to mimick the actual users in CFF. Any improvement in the true positive rate directly translates to money saved for the banking institution, and the worst case (zero true positive rate) corresponds to what the current ATM systems have. For future work, we plan to improve the image resolution and quality to decrease the negative effects of motion blur while keeping within the computational limits imposed by the ATM system. Moreover, other challenging cases such as multiple faces appearing in the scene are potential research directions. Fusing information from the ATM camera itself is potentially useful, but in practice, ATM brands have different camera placements.

7. REFERENCES

[1] H. R. Babaei, O. Molalapata and A. A. Pandor, *Face Recognition Application for Automatic Teller Machines (ATM)*, in ICIKM, 3rd ed. vol.45, pp.211-216, 2012.

[2] Aru, O. Eze and I. Gozie, *Facial Verification Technology for Use in ATM Transactions*, in American Journal of Engineering Research (AJER), [Online] 2013, pp. 188-193, Available: http://www.ajer.org/papers/v2(5)/Y02501880193.pdf

[3] K. J. Peter, G. Nagarajan, G. G. S. Glory, V. V. S. Devi, S. Arguman and K. S. Kannan, *Improving ATM Security via Face Recognition*, in ICECT, Kanyakumari, 2011, vol.6, pp.373-376.

[4] E. Derman, Y. K. Geçici and A. A. Salah, *Short Term Face Recognition for Automatic Teller Machine (ATM) Users*, in ICECCO 2013, Istanbul, Turkey, pp.111-114

[5] A. Ross and A. Jain, *Information Fusion in Biometrics*, in Pattern Recognition Letters, vol.24, pp.2115-2125, 2003.

[6] T. Ahonen, B. Hadid and M. Pietikainen, *Face Description with Local Binary Patterns: Application to Face Recognition*, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 2037-2041, Dec. 2006.

[7] T. Ojala, M. Pietikainen and D. Harwood, *A Comparative Study of Texture Measures with Classification Based on Featured Distributions*, in Pattern Recognition, vol. 29, pp. 51-59, Jan. 1996.

[8] T. Ojala, M. Pietikainen and T. Maenpaa, *Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns*, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 971-987, Aug. 2002.

[9] T. F. Pereira, M. A. Angeloni, F. O. Simoes and J. E. C. Silva, *Video-based Face Verification with Local Binary Patterns and SVM Using GMM Supervectors*, in ICCSA 2012, Part I, LNCS 7333, pp.240-252.

[10] C. C. Chang and C. J. Lin, *LIBSVM: A Library for Support Vector Machines*, in ACM Transactions on Intelligent Systems, 2:27:1-27:27, 2011 Available: http://www.csie.ntu.edu.tw/ cjlin/papers/libsvm.pdf

[11] B. Zadrozny and C. Elkan, *Transforming Classifier Scores Into Accurate Multiclass Probability Estiamtes*, in ACM SIGKDD 2002, pp. 694-699.

[12] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, Hoboken, NJ:Wiley-Interscience, 2003.