Dynamic fusion of human and machine decisions for efficient cost-sensitive biometric authentication

Johannes Coetzer
Department of Mathematical Sciences
Stellenbosch University
Stellenbosch, South Africa
jcoetzer@sun.ac.za

Jacques Swanepoel
Flightscope
Stellenbosch, South Africa
jpswanepoel@gmail.com

Robert Sabourin
École de Technologie Supérieure
University of Quebéc
Montréal, Canada
robert.sabourin@etsmtl.ca

***Abstract*—Despite growing interest in human-machine collaboration for enhanced decision-making, little work has been done on the optimal fusion of human and machine decisions for cost-sensitive biometric authentication. An elegant and robust protocol for achieving this objective is proposed. The merits of the protocol is illustrated by simulating a scenario where a workforce of human experts and a score-generating machine are available for the authentication of handwritten signatures on, for example, bank cheques. The authentication of each transaction is determined by its monetary value and the quality of the claimed author's signature. A database with 765 signatures is considered, and an experiment that involves 24 human volunteers and two different machines is conducted. When a reasonable number of experts are kept in the loop, the average expected cost associated with the workforce-machine hybrid is invariably lower than that of the unaided workforce and that of the unaided machine.**

***Index Terms*—human-machine collaboration, dynamic classifier fusion, cost-sensitive biometric authentication**

I. Introduction

Pattern recognition protocols that produce multiple candidate classifiers, and then combine their output, are popular and well-established [1]. Each candidate classifier can either be a continuous classifier that generates a score, or a discrete classifier that outputs a decision. Furthermore, by imposing different thresholds on a machine-generated score, different discrete classifiers can be obtained. Multiple candidate classifiers can be generated by extracting different features [2] or by utilising different modelling techniques [3].

By considering optimisation data, an optimal candidate classifier, or group of classifiers (e.g. a maximum attainable receiver operating characteristic (MAROC) curve) can be selected and then implemented on different data [4]. The selected classifiers are referred to as maximum attainable classifiers. When the cost associated with the misclassification of an instance (e.g. a handwritten signature), varies from one instance to another, a second optimisation stage is possible [5]. In these scenarios an optimal candidate classifier, or group of classifiers, that minimizes the expected cost, can be dynamically selected from the available maximum attainable classifiers during system implementation.

The candidate classifiers are traditionally machines (so-called hard sensors). More recently, researchers started to investigate the advantages of keeping humans (so-called soft sensors) in the loop [6], [7]. Humans are proficient at integrating information and incorporating context, while machines are adept at making fast, consistent and objective decisions. With the advent of the internet, collaboration among human experts is becoming increasingly viable.

In this paper the feasibility of using a score-generating machine and a workforce of human experts is investigated for the purpose of biometric authentication in a cost-sensitive environment. This investigation considers handwritten signatures on, for example, bank cheques, where authentication relies on both the transaction's monetary value and the quality of the client's signature. The protocol should lean towards rejecting large transactions with low-quality signatures, and accepting small transactions with high-quality signatures. Furthermore, the average expected cost associated with an expert-machine hybrid, should be lower than that of the unaided human workforce, *and* lower than that of the unaided machine.

Since the performance of humans are often comparable to that of machines in authenticating handwritten signature images [8], it is reasonable to investigate human-machine collaboration within this context. The empirical findings in [9] confirm that human team members are extremely sensitive to their workload in pressured, high-tempo situations and when supported by machines, perform better by maintaining team performance at acceptable levels.

It is concluded in [10] that, with *all* the human experts in the loop, the inclusion of an HMM-based machine simplifies cost-sensitive authentication and decreases the expected cost for *all* operating conditions. This self-contained paper improves on [10] in several ways: (1) a more detailed motivation, description and analysis of the dynamic classifier selection strategy is presented; (2) a *more proficient* classifier fusion protocol is proposed; and (3) the efficacy of the improved protocol is demonstrated (for two *different* machines) in scenarios where only a *subset* of the human workforce is available.

In Section II relevant ROC-based strategies for classifier fusion are introduced. This is followed by a discussion on ROC-based classification in a cost-sensitive environment (Section III). In Sections IV and V the proposed strategy for human-machine collaboration is introduced, followed by an analysis of the relevant handwritten signature data, the experimental protocol, and results.

II. Information fusion

The proposed protocol for human-machine collaboration is based upon performance evaluation in ROC space. In order to standardise the terminology and notation, key concepts in ROC analysis are first reviewed.

The true positive rate (TPR) for classifier C_A, i.e. t_A^+, approximates the probability that it will correctly classify a positive instance (authentic signature), while its false positive rate (FPR), i.e. f_A^+, approximates the probability that it will erroneously classify a negative instance (fraudulent signature).

In ROC space, the TPR and FPR represent the vertical and horizontal axes respectively. The performance of a discrete classifier (e.g. a human expert that provides a decision of either 'true' or 'false') can therefore be represented by a single point in ROC space. When two discrete classifiers are compared, the superior classifier's performance is represented by the more 'northwesterly' point in ROC space. The two machines considered in this paper are both examples of continuous classifiers, since they both output scores to which different decision thresholds can be applied to determine class membership. The performance of a continuous classifier is represented by an ROC curve. An ROC curve therefore consists of a number of FPR-TPR pairs, where each pair is associated with a specific threshold value and constitutes a discrete classifier. When two continuous classifiers are compared, the superior classifier has a larger area under its corresponding ROC curve (AUC).

The problem investigated in this paper is addressed by employing classifier combination through majority voting (MV) and iterative Boolean combination (IBC). MV is the most popular classifier combination strategy when a system has access to the output of three or more discrete classifiers, e.g. human experts, that make conditionally independent errors.

The IBC algorithm [11] combines the output of two continuous classifiers−their respective performances are represented by two ROC curves−by fusing the output of every threshold-specific discrete classifier associated with the one ROC curve with the output of every threshold-specific discrete classifier associated with the other ROC curve. Ten Boolean fusion functions are implemented for combining the output of any pair of discrete classifiers, C_A and C_B, where $\wedge$, $\vee$, $\neg$, and $\oplus$ denote conjunction, disjunction, negation, and the XOR operator, respectively: (1) $C_A \wedge C_B$, (2) $\neg C_A \wedge C_B$, (3) $C_A \wedge \neg C_B$, (4) $\neg(C_A \wedge C_B)$, (5) $C_A \vee C_B$, (6) $\neg C_A \vee C_B$, (7) $C_A \vee \neg C_B$, (8) $\neg(C_A \vee C_B)$, (9) $C_A \oplus C_B$, and (10) $\neg(C_A \oplus C_B)$. In this way a set of candidate hybrid classifiers are produced of which only the the optimal hybrids, represented by the MAROC curve, are selected.

III. Cost-sensitive classification

In order to select the specific hybrid classifier on a MAROC curve, which is associated with the lowest expected cost, the use of *iso-cost* lines with variable gradients is proposed.

It is reasonable to assume that the cost incurred by rejecting a negative instance and the cost incurred by accepting a positive instance both equals zero, i.e. $S(-|-) = S(+|+) = 0$. Given this assumption, the expected cost associated with a transaction that is authenticated by a classifier C_A can be expressed as follows [5],

$$E_A = S(-|+) \cdot (1 - t_A^+) \cdot P(+) + S(+|-) \cdot (f_A^+) \cdot P(-), \quad (1)$$

where $P(+)$ and $P(-)$ represent the prior probabilities of the questioned instance being positive and negative respectively. The cost incurred by rejecting a positive instance, and the cost incurred by accepting a negative instance, are denoted by $S(-|+)$ and $S(+|-)$ respectively.

It can be deduced from (1) that the iso-cost line in ROC space, that depicts the proficiency of all hypothetical classifiers associated with a *specific* expected cost E, is given by

$$t^+ = \left\{ \frac{S(+|-)P(-)}{S(-|+)P(+)} \right\} f^+ - \frac{E}{S(-|+)P(+)} + 1.$$

In Figure 1 (a) the horizontal, vertical and diagonal lines represent parallel iso-cost lines, for the scenarios where $S(+|-)P(-) = 0$, $S(-|+)P(+) = 0$, and $S(-|+)P(+) = S(+|-)P(-)$, respectively.

Since the overwhelming majority of questioned signatures on bank cheques are authentic, the pragmatic strategy will be to set the prior probabilities equal to $P(+) \approx 1$ and $P(-) \approx 0$. For this scenario, an almost optimal expected cost ($E \ll 1$) can be attained by accepting all questioned signatures as demonstrated in Figure 1 (b). Said strategy will however cause any manual or automated authentication protocol to be redundant. It is more sensible to embark from the assumption that the prior probabilities are equal, i.e. $P(+) = P(-) = 0.5$. All human experts are therefore directed to be as unbiased as possible. The strategy for selecting an optimal hybrid classifier from a set of candidates is also based on this assumption. As a result (1) simplifies as follows,

$$E_A = 0.5 \left[S(-|+) \cdot (1 - t_A^+) + S(+|-) \cdot (f_A^+) \right]. \quad (2)$$

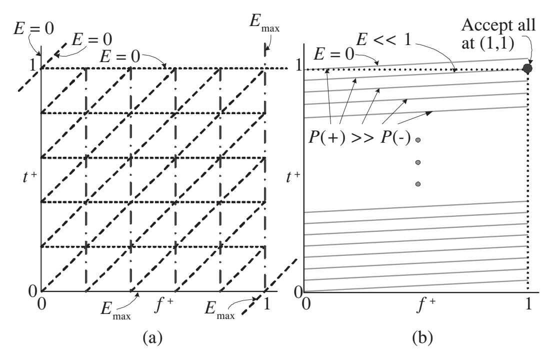

Fig. 1. (a) The horizontal, vertical and diagonal lines represent *iso-cost lines* for the scenarios where $S(+|-)P(-) = 0$, $S(-|+)P(+) = 0$, and $S(-|+)P(+) = S(+|-)P(-)$, respectively. (b) The iso-cost lines for *'pragmatic'* prior probabilities for questioned signatures on bank cheques ($P(+) \approx 1$ and $P(-) \approx 0$) are depicted by parallel solid grey lines, each with a small positive gradient.

When the error costs ($S(+|-)$ and $S(-|+)$) are kept constant, the line in ROC space represented by $t^+ = Mf^+ + N(E)$ depicts the proficiency of all hypothetical classifiers that correspond to the specific expected cost E, where $M =$

$S(+|-)/S(-|+)$ and $N(E) = 1-(2E)/S(-|+)$. By considering different values of E, different parallel iso-cost lines can be obtained for a specific value of M. Note that M therefore denotes both a specific cost ratio, *and* the cost gradient of the corresponding iso-cost lines. The aforementioned two terms are henceforth used interchangeably. After a cost-ratio is specified, only one iso-cost line (with gradient M) intersects a linearly interpolated version of a MAROC curve at a *single* point (see Figure 2 (b)). The aforementioned point is optimal in the sense that it represents the performance of the hybrid classifier that corresponds to the lowest expected cost$-$this classifier is therefore selected.

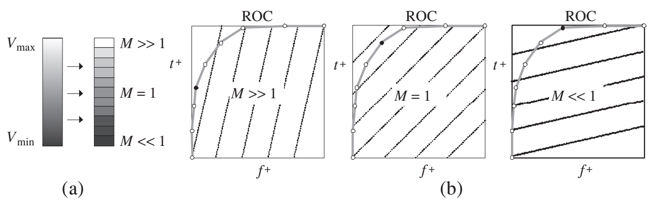

Fig. 2. The geometric interpretation of the proposed classifier selection protocol. (a) A hypothetical transaction value V is mapped to a specific cost gradient M. (b) A hypothetical MAROC curve, associated with a specific human expert, and three different cost scenarios with corresponding iso-cost lines. For each cost scenario the performance of the optimal hybrid with the lowest expected cost is denoted by a solid marker $-$ this classifier is selected.

IV. System design

It is assumed that, during an initial enrolment phase, a financial institution requires each new client to produce a number of authentic samples of his/her handwritten signature. The utilisation of these samples for signature modelling is explained in Section V. The proposed protocol is partitioned into an optimisation and implementation stage.

Optimisation. During the optimisation stage both a machine and a workforce of human experts are required to authenticate a compilation of labelled authentic *and* fraudulent signature samples, in order to assess their proficiency. These samples are produced by representative writers in a controlled setting. The protocol for compiling and presenting these signatures is discussed in detail in Section V. By considering the representative signatures, an estimate of the combined performance of a specific human expert and each threshold-specific machine-generated classifier is obtained by considering each of the ten Boolean fusion functions defined in Section II. A set of candidate human-machine hybrids is therefore generated for said expert, after which only the MAROC curve is retained. This process is repeated for every expert so that a set of expert-specific MAROC curves is obtained.

The consultant and the client (financial institution) agree on a mapping between the transaction value $V \in [V_{\min}, V_{\max}]$ (in monetary terms) and a finite set of discrete cost gradients $\{M_i\},\ i = 1, 2, \ldots, K$, i.e. $V \mapsto \{M_i\}$ (see Figure 2 (a)).

For a specific cost gradient, the optimal expert-machine hybrid, i.e. the hybrid that corresponds to the lowest expected cost, is selected for each expert-specific MAROC curve. This process is repeated for every specified cost gradient. After the conclusion of the optimisation stage, only the optimal expert-specific hybrids for every specified cost gradient are stored and the optimisation process is only conducted once. When new experts are added to the human workforce, or when existing experts undergo updated proficiency tests, the optimal hybrid for only the aforementioned experts need to be re-calculated. Figure 2 (b) illustrates three different cost scenarios with the corresponding iso-cost lines, as well as a hypothetical expert-specific MAROC curve. For each of these scenarios, the performance of the optimal hybrid is denoted by a solid marker.

Implementation. During the implementation stage the optimal expert-specific hybrids are efficiently and dynamically selected. When investigating an unlabelled signature, claimed to belong to a specific writer and associated with a certain transaction value, the transaction value is first mapped to the appropriate cost gradient. Subsequently, for the cost-gradient in question, the output from all the optimal expert-specific hybrids are combined through MV. In this way all the available experts are included in the decision process.

The protocol presented here *differs* from the one proposed in [10]. According to the protocol adopted in [10] the output from all of the available human experts (and *not* the decisions of expert-specific hybrids) is *first* fused through MV$-$the majority-vote decision of the entire available workforce is *then* combined with every threshold-specific machine-generated decision. It is shown in Section V-C that the human-machine collaboration protocol proposed in this paper is *superior to* and *more robust than* the one proposed in [10], as long as a significant number of human experts are kept in the loop.

V. Experiments

The experimental data is split into an optimisation set (OS) and evaluation set (ES). The OS contains signatures produced by representative writers in a controlled setting, while the ES contains signatures from other writers emulating banking clients. It is reasonable to presume that positive signatures are obtainable for each writer in both the OS and ES. These signatures may be used to train writer-specific HMMs, act as reference signatures for writer-independent LDF-based classifiers, or serve as reference for human experts.

Labelled positive and negative signatures are associated with writers in the OS only. These signatures may be used for estimating the proficiency of the human experts and the machine in question, as well as for selecting the optimal expert-specific hybrids. Unlabelled positive and negative signatures, that belong to writers in the ES, are used to estimate the generalisation capability of the proposed protocol.

A. Data

The efficacy of the proposed protocol is illustrated by considering a selected subset of signatures within a larger database originally captured online [12]. This dataset contains dynamic signatures from 51 different writers. In order to emulate signatures extracted from bank cheques, the dynamic data is transformed into static images by applying a morphological dilation operator to the pixels positioned at the captured pen tip coordinates [13]. Only skilled forgeries are used for experimentation. A skilled forgery is produced by an individual who had ample time to study a set of known (labelled) authentic signatures at his/her leisure. Akin to the terminology in [12], this dataset contains 15 authentic 'training' signatures and 75 'test' signatures for each writer. The 75 'test' signatures consist of 15 authentic samples and 60 skilled forgeries.

B. Experimental protocol

For each writer in the dataset, all of the 15 authentic training signatures are selected for signature modelling. A reduced test set, that consists of only 15 signatures, is now constructed. This new test set is employed during the optimisation and implementation stages of experimentation and contains a randomly selected number (between 0 and 15) of skilled forgeries. The rest of the test signatures are randomly selected from the 15 authentic test signatures for the writer in question. A specific test set may therefore consist of only authentic signatures or only skilled forgeries. Consequently, each classifier (human expert or machine) authenticates $15 \times 51 = 765$ test signatures in total. Due to the random nature of this selection strategy, the total number of authentic and forged samples in the entire reduced test set is 432 and 333 respectively.

Human experts. The potential actions of a human expert is emulated by presenting each volunteer (a faculty member or a graduate student) with a training set (15 signatures) and corresponding test set (15 signatures) for all 51 writers. Twenty-four volunteers are utilised. Each volunteer is presented with all the writers' training and corresponding test sets on different sheets of paper. The volunteers are instructed to compare every test signature to the corresponding training set, and decide which of the test signatures are fraudulent. Each training set is scrutinised *as a whole*. Said volunteers are also instructed not to mull over each decision, so as to emulate the probable actions of a typical bank employee.

Machines. The signatures presented to the human volunteers, are also presented to two machines, i.e. a *writer-dependent* HMM-based classifier [13] and a *writer-independent* LDF-based classifier [14], as discussed below.

HMM-based classifier. Features based on the computation of the discrete Radon transform (DRT), are extracted from each signature image. These features are employed to train an HMM for each writer in the dataset. A questioned signature is matched with the appropriate HMM through Viterbi-alignment and a score is obtained. This score is then normalised through a strategy based on the z-norm.

LDF-based classifier. During signature modelling, a dissimilarity representation is achieved by employing a two-stage process. Binary signature images are first converted into feature sets using the DRT. Using a dynamic time warping algorithm these feature sets are matched to those extracted from writer-specific reference signatures, so that a set of dissimilarity vectors is obtained. The dissimilarity vectors obtained from signatures in the training set are used to train an LDF. During the implementation stage, questioned signatures from the ES are encoded into dissimilarity vectors, by comparing these signatures to the appropriate writer-specific reference signatures. The trained LDF is then used to predict class membership.

Cross-validation with repetition. The experimental protocol employs three-fold cross validation in conjunction with repetitive data randomisation. The experimental protocol is outlined as follows: (1) The dataset is partitioned into three equal subsets, where each subset contains signatures produced by 17 writers; (2) Each subset, in turn, is employed as an ES, that contains signatures produced by 17 writers, while the remaining two subsets constitute the OS, that contains signatures produced by the other 34 writers; (3) The order of the writers is randomly rearranged, and the procedure is repeated 10 times. The results for 30 trials are thus reported.

A set of 19 different cost gradients is specified as follows,

$$M = \left\{ \frac{1}{10}, \frac{1}{9}, \frac{1}{8}, \ldots, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, \ldots, 8, 9, 10 \right\}. \tag{3}$$

A specific trial is executed by considering all of the cost gradients in (3). For a specific cost gradient, the signatures in the OS are first used to select the optimal expert-machine hybrid (i.e. the hybrid that corresponds to the lowest expected cost) on each expert-specific MAROC curve. The signatures in the ES are then authenticated by combining the decisions of the optimal expert-machine hybrids through MV, and the expected cost is estimated. This process is repeated for every cost gradient in (3) so that the average expected cost over all cost gradients is reported for the trial in question.

In order to demonstrate the efficacy of the proposed protocol for randomly selected *subgroups* of human experts, the empirical protocol outlined in Algorithm 1 is adopted.

```
SizeOfWorkforce ← 24;
NrOfShuffles ← 10; NrOfFolds ← 3
for all NrOfSelectedExperts such that
NrOfSelectedExperts ∈ [1, SizeOfWorkforce] do
  Randomly select NrOfSelectedExperts humans
  from the workforce
  TrailNr ← 1
  for all ShuffleNr such that
  ShuffleNr ∈ [1, NrOfShuffles] do
    Randomly shuffle the 51 writers in the dataset
    for all Fold such that Fold ∈ [1, NrOfFolds] do
      Optimisation set ← 34 writers in the dataset
      Evaluation set ← 17 writers in the dataset
      Execute trial TrailNr
      TrailNr ← TrailNr + 1
    end for
  end for
end for
```

Algorithm 1: Experimental protocol.

C. Results

The dataset introduced in Section V-A is now considered, and the level of experimental complexity is increased in a step-wise fashion. This approach clarifies the *methodology* and demonstrates the *efficacy* of the collaboration protocol

A *single* human expert is first considered and the OS is employed to illustrate how the optimal expert-machine hybrid, for a *specific* cost gradient, is selected (see Figure 3). *Three* human experts are then considered and the ES is employed to show that, when the selected *three* expert-machine hybrids are *again* combined through MV, the expected cost associated with the *combined* hybrid classifier is lower than that of the optimal unaided threshold-specific machine for the cost gradient in question (see Figure 4). The expected cost associated with the *combined* hybrid classifier is *also* lower than that of the unaided human workforce, when the individual human decisions are combined through MV. The results in Figures 3 and 4 are generated for a *single* cost gradient during a *single* trial, while only the HMM-based machine is considered.

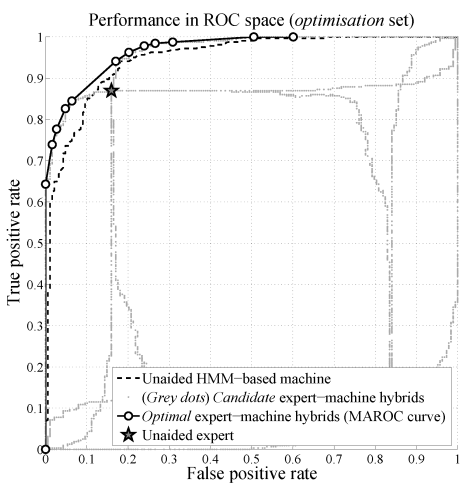

Fig. 3. MAROC curve generation for a *single* expert using the *OS*. The estimated performance (in ROC space) of an *unaided* human expert and an *unaided* HMM-based machine (see Section V-B) are depicted by the pentagram and the dashed line respectively. The performance of the candidate expert-HMM hybrids, when *all* the Boolean fusion functions described in Section II are used to combine the expert decision with every threshold-specific machine decision, is represented by grey dots, while the MAROC-curve for the candidate expert-HMM hybrids is denoted by black circles.

Since only the cost gradient/ratio M is specified in the remaining experiments, and not the individual error costs ($S(+|-)$ and $S(-|+)$), the constraint, $S(+|-)+S(-|+) = 1$, is imposed. Since $M = S(+|-)/S(-|+)$, (2) simplifies to

$$E_A = \frac{0.5}{M+1}\left[(1 - t_A^+) + (M \cdot f_A^+)\right],$$

for an arbitrary classifier C_A. This is convenient for plotting the expected cost associated with C_A as a function of the cost gradient M. The relaxation of the above-mentioned constraint has no impact on the shape of the $E_A - M$ graph. In fact, said relaxation only results in a re-calibration of the E_A-axis.

In Figure 5 the average expected cost over all 30 trials is shown as a function of the cost gradient M, with the

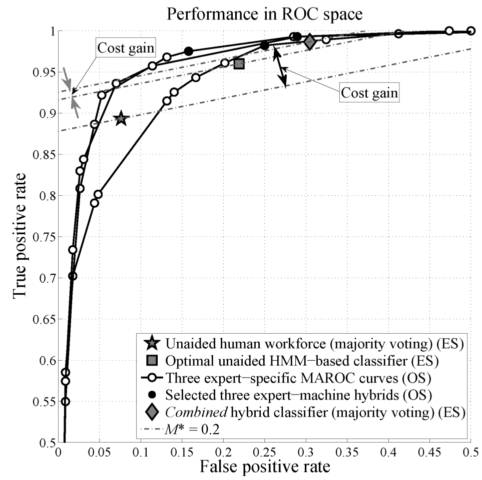

Fig. 4. Dynamic classifier selection, MV and performance evaluation for *three* human experts using the *ES*. When an unseen questioned signature associated with a cost gradient of $M^* = 0.2$ is to be authenticated, only *one* optimal hybrid is selected on *each* of the three expert-specific MAROC curves (three black dots). These hybrids are used to authenticate the questioned signature, after which the three decisions are combined through MV. When *all* the questioned signatures in the *ES* are authenticated in this way, the performance is depicted by the diamond. The estimated cost associated with the aforementioned classifier is lower than that of the optimal *unaided* threshold-specific HMM-based classifier (square)—the gain in cost is depicted by the grey double-arrow. The cost depicted by the diamond is *also* lower than that of the *unaided* human workforce (pentagram), when the individual human decisions are combined through MV—the gain in cost is depicted by the black double-arrow.

different values of M specified in (3). The same three human experts that relate to Figure 4 are again considered here. For a specific trial and cost gradient, the OS is used to select the three optimal expert-machine hybrids, while the ES is used to estimate the expected cost when these hybrids are combined through MV. The combined hybrid classifier outperforms the optimal unaided HMM-based classifier *and* the unaided human workforce (when the individual human decisions are combined through MV) for *all* cost gradients.

Figures 6 and 7 show the average expected cost for the *unaided* human workforce (using MV), the optimal *unaided* threshold-specific machine, the human-machine collaboration strategy proposed in [10], and the human-machine collaboration strategy proposed in *this* paper, as a function of the number of *available* experts, by considering the HMM-based and LDF-based machines (described in Section V-B), respectively. By considering the ES only, the average expected cost is obtained by calculating the mean cost over all 30 trials (and all specified cost gradients). For the HMM-based machine, *both* of the above-mentioned collaboration strategies outperform the unaided human workforce *and* the optimal unaided threshold-specific HMM-based classifier when more than *two* experts are available. For the LDF-based machine, *only* the collaboration strategy proposed in *this* paper outperforms the unaided human workforce *and* the optimal unaided threshold-specific LDF-based classifier when more than *five* experts are available.

Since the human-machine collaboration protocol presented in *this* paper enhances the performance of *both* the *less*

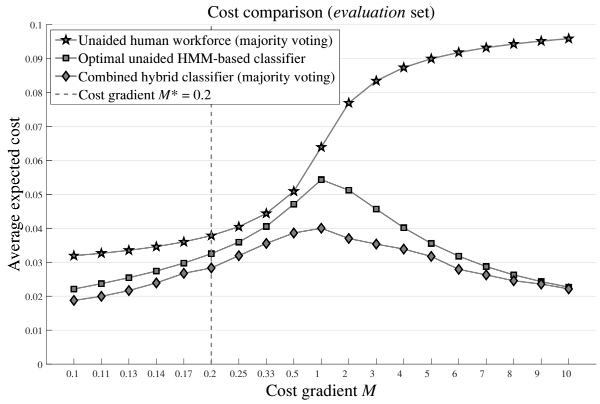

Fig. 5. The *average* expected cost over *all* 30 trials as a function of the cost gradient M. The *same* three human experts that relate to Figure 4 are again considered here. The combined hybrid classifier outperforms the optimal unaided threshold-specific HMM-based classifier *and* the unaided human workforce (when the individual human decisions are combined through MV) for *all* cost gradients. The cost gradient considered in Figure 4, i.e. $M^* = 0.2$, is indicated for reference.

proficient HMM-based machine *and* the *more proficient* LDF-based machine, when a reasonable number of experts are kept in the loop, the protocol presented here is *superior to*, and *more robust than*, the one proposed in [10].

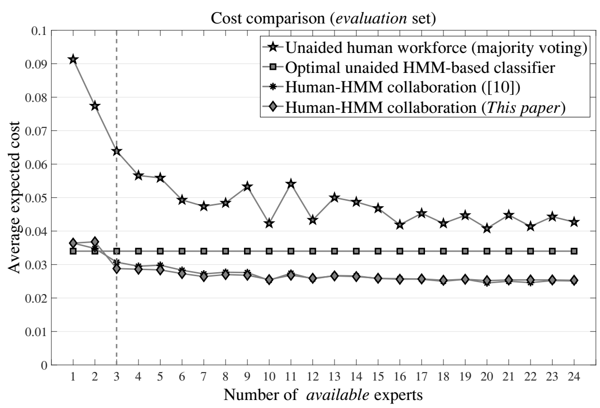

Fig. 6. The average expected cost estimated on the *ES* (for 19 different cost gradients and for *all* 30 trials) of the *unaided* human workforce (using MV), the optimal *unaided* threshold-specific machine, the human-machine collaboration strategy proposed in [10], and the human-machine collaboration strategy proposed in *this* paper, as a function of the number of *available* experts, when an *HMM-based* machine is considered. *Both* of the above-mentioned collaboration strategies outperform the unaided human workforce *and* the optimal unaided threshold-specific HMM-based classifier when more than *two* experts are available. Since *three* experts are available for the scenario depicted in Figure 5, the average expected costs associated with this scenario is specifically indicated.

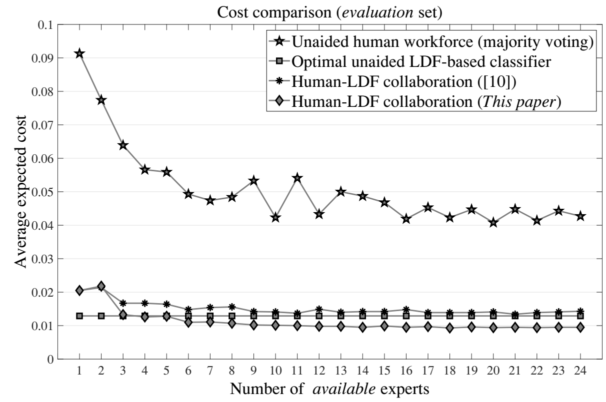

Fig. 7. The average expected cost estimated on the *evaluation* set (for 19 different cost gradients and for *all* 30 trials) of the *unaided* human workforce (using MV), the optimal *unaided* threshold-specific machine, the human-machine collaboration strategy proposed in [10], and the human-machine collaboration strategy proposed in *this* paper, as a function of the number of *available* experts, when an *LDF-based* machine is considered. The collaboration strategy proposed in *this* paper outperforms the unaided human workforce *and* the optimal unaided threshold-specific LDF-based classifier when more than *five* experts are available.

REFERENCES

[1] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28 – 44, 2013.

[2] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *Proceedings of the 12th International Conference on Computer Vision*, 2009, pp. 221–228.

[3] V. Frinken, T. Peter, A. Fischer, H. Bunke, T.-M.-T. Do, and T. Artieres, "Improved handwriting recognition by combining two forms of hidden markov models and a recurrent neural network," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, X. Jiang and N. Petkov, Eds. Springer Berlin Heidelberg, 2009, vol. 5702, pp. 189–196.

[4] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[5] C. Drummond and R. C. Holte, "Cost curves: An improved method for visualizing classifier performance," *Machine Learning*, vol. 65, no. 1, pp. 95–130, 2006.

[6] M. Nilsson, J. van Laere, T. Susi, and T. Ziemke, "Information fusion in practice: A distributed cognition perspective on the active role of users," *Information Fusion*, vol. 13, no. 1, pp. 60 – 78, 2012.

[7] M. P. Jenkins, G. A. Gross, A. M. Bisantz, and R. Nagi, "Towards context aware data fusion: Modeling and integration of situationally qualified human observations to manage uncertainty in a hard + soft fusion process," *Information Fusion*, vol. 21, pp. 130 – 144, 2015.

[8] J. Coetzer, B. M. Herbst, and J. A. du Preez, "Off-line signature verification: A comparison between human and machine performance," in *Proceedings of the Tenth International Workshop on Frontiers in Handwriting Recognition*, 2006, pp. 481 – 485.

[9] X. Fan, S. Sun, M. McNeese, and J. Yen, "Extending the recognition-primed decision model to support human-agent collaboration," in *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems*, 2005, pp. 945–952.

[10] J. Coetzer, J. P. Swanepoel, and R. Sabourin, "Efficient cost-sensitive human-machine collaboration for off-line signature verification," in *Proceedings of the 19th International Conference on Document Recognition and Retrieval*, C. Viard-Gaudin and R. Zanibbi, Eds., vol. 8297, 2012, pp. 82 970J–1–82 970J–8.

[11] W. Khreich, E. Granger, A. Miri, and R. Sabourin, "Iterative boolean combination of classifiers in the ROC space: An application to anomaly detection with HMMs," *Pattern Recognition*, vol. 43, no. 8, pp. 2732–2752, 2010.

[12] J. G. A. Dolfing, E. H. L. Aarts, and J. J. G. M. van Oosterhout, "On-line signature verification with hidden Markov models," in *Proceedings of the 14th International Conference on Pattern Recognition*, vol. 2, 1998, pp. 1309–1312.

[13] J. Coetzer, B. M. Herbst, and J. A. du Preez, "Off-line signature verification using the discrete Radon transform and a hidden Markov model," *Eurasip Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 559–571, 2004, Special issue: Biometric Signal Processing.

[14] J. P. Swanepoel and J. Coetzer, "Writer-specific dissimilarity normalisation for improved writer-independent off-line signature verification." in *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition*, 2012, pp. 391–396.