Chapter #2

EXPERIMENTS ON DECISION FUSION FOR DRIVER RECOGNITION

Subtitle

Hakan Erdoğan, Aytül Erçil and Hüseyin Abut
Sabancı University, Istanbul, Turkey

Abstract: In this work, we study the individual as well as combined performance of various driving behavior signals on identifying the driver of a motor vehicle. We investigate a number of classifier fusion techniques to combine multiple channel decisions. We observe that some driving signals carry more biometric information than others. When we use trainable combining methods, we can reduce identification error significantly using only driving behavior signals. Classifier combination methods seem to be very useful in multi-modal biometric identification in a car environment.

Key words: Biometric person identification, driver recognition, speaker recognition, face recognition, driving signals, driver behavior modeling

1. INTRODUCTION

Studies and technological advances in biometric person identification promises a world with no keys or passwords where smart devices or systems around us can identify us from our biological or behavioural traits. Biometric person identification would also be useful in a car where most of us spend long hours every day. It may be beneficial to identify the driver of a vehicle for safety and comfort purposes [1].

In-car person identification is a relatively new area where a few academic studies exist. Earlier, we have studied techniques to combine information from video, audio and driving signals to identify a driver of a vehicle using a 20 person subset of the Nagoya University CIAIR database [1,2]. In that study, we have used feature fusion of acceleration and brake pedal pressure signals to perform identification. Thus, the separate effects of acceleration and brake pedal pressure were not clear. However, in this paper, we focus on individual identification performance of five different driving signals and their various combinations. We compare feature versus decision fusion in this scenario as well.

This paper is organized in the following way. In section 2, we describe the types of driving behaviour signals of the CIAIR database [2]. We present our statistical GMM models for the driving signals in section 3. Fusion methods are explained in section 4. We present our experimental results in section 5 followed by the conclusions and future plans in the final section.

2. TYPES OF DRIVING SIGNALS

The Center for Integrated Acoustic Information Research (CIAIR) at Nagoya University has built a multi-modal corpus inside a vehicle, where each driver subject was required to carry out conversations with three different dialog systems while driving [2]. Data from 12 audio channels and 3 video channels have been recorded for over 800 drivers, both female and male. They have also collected five different "driving behavior signals" during theses sessions. Driving behavior data is collected by 5 analog channels, each sampled at 1.0 kHz with an unsigned 16-bit format.

- Brake pedal pressure in Kgforce/cm.sq.: 0 - 50 kgf/cm2 is mapped to 0 - 5.0V and linearly digitized in the range 0 to 32767.
- Accelerator pedal pressure in Kgforce/cm.sq.: 0 - 50 kgf/cm2 is mapped to 0 - 5.0V and linearly digitized in the range 0 to 32767.
- Engine speed in rpm.: 0 - 8,000 rpm is mapped to 0 - 5.0V and linearly digitized in the range 0 - 32767.
- Vehicle speed in km/h.: 0 - 120 km/h is mapped to 0 - 5.0V and linearly digitized in the range 0 - 32767.
- Steering wheel angle in -1800 degrees to +1800 degrees; i.e., five CW and 5 CCW revolutions is linearly digitized in the range -32769 to 32767.

In this paper, we use these signals to try to identify driver identities. To extract features from these signals, we perform smoothing and noise removal in time domain followed by decimation. We also extract dynamic features by computing the first difference of time-domain samples. Fourier domain or cepstral features are normally used for speech. For driving signals, however, we use time domain signals after a smoothing stage to reduce noise. Unlike speech, there is no evidence of periodic (pitch) and frequency-related information in these driving signals. Frequency domain processing could be useful to remove noise in these signals. However, noise removal could be performed in time-domain as well. Thus, we only use the time-domain signal directly in this work.

We use statistical modelling to model these driving signals and their first differences. We provide the details of our modelling approach in the following section.

3. MODELLING TIME-SERIES SIGNALS

As many time-series signals are slowly varying, it is natural to assume quasi-stationarity for modeling purposes. This naturally leads to Hidden Markov Model type dynamic generative models to model time-series data. In biometric identification from time-series data, the underlying state topology of the signal is usually unclear (except in text-dependent speaker recognition) and single-state probabilistic models perform well. This is true given that we use a parametric continuous distribution function with multiple modes to cover variations in a time-series signal. Gaussian mixture models (GMMs) are models that can approximate any smooth distribution, even if it has multiple modes. As long as we use enough number of mixtures in a GMM, we can obtain a good statistical model of time-series data. GMM modeling for driving signals were first used in [3] and we also used GMM models in [1] for modeling driving behavior.

In GMM modeling, features are considered as independent identically distributed random vectors drawn from a GMM distribution:

$$\mathrm{f}(\mathbf{x} \mid S_i) = \sum_{k=1}^{K} \pi_k \, \mathrm{N}(\mathbf{x}, \boldsymbol{\mu}_\mathbf{k}, \boldsymbol{\Sigma}_\mathbf{k}). \tag{1}$$

where $\mathbf{x}$ represents the feature vector, π_k are mixture weights and $\mathrm{N}(\mathbf{x}, \boldsymbol{\mu}_\mathbf{k}, \boldsymbol{\Sigma}_\mathbf{k})$ are individual Gaussians for representing a particular subject under study, S_i. For computational purposes, $\boldsymbol{\Sigma}_\mathbf{k}$ are chosen to be diagonal matrices. GMMs have been used in text-independent speaker recognition with great success [4]. A popular way of using GMMs in speaker recognition is to train a large background speaker model, i.e., 1024 Gaussians, and adapt this model to each speaker using that particular speaker's data. GMM training is performed using the well-known EM algorithm [5].

In this paper, we train a GMM for each person's time-series data from scratch and we have used eight (8) mixtures, which resulted in satisfactory performance in this application. During the testing phase, the per-frame log-likelihood value of observed data $\mathbf{X} = (\mathbf{x}_j)_{j=1}^{N}$ under the model of a particular person S_i can be computed as:

$$
\begin{aligned}
L_i(\mathbf{x}) &= \frac{1}{N}\sum_{j=1}^{N}\log \mathrm{f}(\mathbf{x}_j \,|S_i) \\
&= \frac{1}{N}\sum_{j=1}^{N}\left(\log \sum_{k=1}^{K}\pi_k \ \mathrm{N}(\mathbf{x}_j, \boldsymbol{\mu}_{\mathbf{k}}, \boldsymbol{\Sigma}_k)\right).
\end{aligned} \tag{2}
$$

We also train a background model, one more GMM, with twice the number of mixtures. Background GMM is required for normalization in likelihood ratio testing for biometric *verification*. The log-likelihood of the observed data under the background model, $L_g(\mathbf{x})$ can also be computed in a similar way.

For verification task, the Bayesian decision amounts to the comparison of the log-likelihood-ratio, $L_i(\mathbf{x})$ -$L_g(\mathbf{x})$ to a threshold. For different thresholds, we trace the receiver operating characteristics (ROC) curve which plots false-accept rate versus false-reject rate.

For *identification* problem, however, we need to obtain posterior probabilities of identities given test data and choose the largest one as the identity of the test segment. The posterior probabilities can be found from:

$$
p(S_i \mid \mathbf{x}) = \frac{e^{L_i(\mathbf{x})}}{\sum_{j=1}^{M} e^{L_j(\mathbf{x})}}, \tag{3}
$$

where we assume equal priors for each class. These probabilities can also be called *scores*. We discuss how to combine these scores from different modalities in the following section.

4. FUSION METHODS

Combining multiple classifiers is a new research area that has attracted great interest [6] during the past few years. Combination methods can be divided into two categories: fixed and trained.

Fixed methods have simple fixed rules to combine information from a set of classifiers. On the other hand, trainable combination methods have some free parameters that can be trained on a separate part of training data (validation or held-out data). Trainable combiners are typically classifiers themselves. They classify in the score space rather than the original feature space.

Given test data **x**, let *S(i,j)* represent the score of person *i* in modality *j*. We drop **x** from our notation for brevity. Our goal is to obtain a single score

S(i) for person *i* using a combination method. We identify various classifier combination methods below:

4.1 Fixed combiners:

We use simple fixed rules to combine scores from different classifiers (modalities). Fixed rules are suboptimal since classifiers are not differentiated among each other. Thus, fixed rules do not take into account the variability of reliabilities of different classifiers.

Some fixed rules of classifier combination are listed below:

- Max Rule: $S(i)=\max_j S(i,j)$
- Min Rule: $S(i)=\min_j S(i,j)$
- Mean (sum) Rule: $S(i)=\mathrm{sum}_j S(i,j)$
- Product Rule: $S(i)=\mathrm{prod}_j S(i,j)$
- Median Rule: $S(i)=\mathrm{median}_j S(i,j)$

4.2 Trainable combiners:

In trainable classifier combination, we form a vector of all scores computed using all the classifiers available. The entries of the vector is given as the elements of the following set $S=\{S(i,j): i=1..N_p, j=1..N_c\}$ where N_p and N_c denote the number of people (classes) and classifiers respectively.

We use this score vector as a new feature vector for classification. Thus, we can use any classification method as a second classifier. We must note that the second classifier should be trained using a different (held-out or validation) data set from the original training data to avoid overtraining.

We list the types of combining classifiers we used in this work in the following:

- Nearest mean combiner (NMC): A simple linear combiner that chooses nearest class mean as the classifier output.
- Fisher combiner (Fisher): A linear classifier that minimizes least squares error in mapping features to class labels in a one-vs-all fashion.
- Linear discriminant combiner (LDC): Another linear classifier that models each class by a Gaussian that shares the same covariance matrix with other classes.
- Naïve Bayes combiner (NB): Naïve Bayes classifier. Assumes that class-conditional probabilities of the feature vector coordinates are statistically independent. Each coordinate is modeled with a nonparametric binning distribution model with 10 bins.
- Parzen combiner: Parzen density based combiner.

5. EXPERIMENTS AND RESULTS

We have carried out experiments on decision fusion for driver recognition using CIAIR database from Nagoya University. We used a 20 person subset of the database that we also used in an earlier paper [1]. In this study, we extract all driving signals and evaluate their performance individually as well as after combination.

From each driver, 50 image frames, 50 seconds of non-silence audio and around 600 seconds of driving signals were utilized. We extracted features from this dataset and divided all features into 20 equal length parts for each driver and modality and number the parts from one to 20. When we have formed the multimodal test-sets, we have assumed that each modality part was associated with the parts that have the same number in other modalities. Smoothed and sub-sampled driving signals and their first derivatives were used as features for modeling driving behavior of the drivers. Thus, each driving signal has 2 dimensional feature vectors.

We have then performed a leave-one-out training procedure, where for each single testing part, seventeen parts were used for training and two parts were held-out for validation to optimize normalization parameters and fusion weights. This gave us 20 tests for each person (each time the training data is different although not independent), leading to 400 (20x20) genuine tests in total. GMMs were driven with eight, one, and eight mixture components for speech, face, and driving signals, respectively. Background GMM models were trained for each modality as well [6].

We performed closed set identification with the data for this study. We used *prtools* [7] software library for evaluating the results and combining the classifiers. In Table 1, we present individual performance results for each (possibly feature combined) modality.

Table -1. Individual performance results for different modalities

Modality	Percent Error (%)
Acceleration (A)	42.5
Brake (B)	31.7
Engine Speed (E)	84.2
Vehicle Speed (V)	81
Steering wheel angle (W)	88.7
A+B+E+V+W	31.2
A+B[1]	10.2
A+B+W	16.5

[1] The results for A+B, F and S features were found in [1]. For A+B driving features, we re-estimated the GMMs. Due to random initialization, the results are slighty different than the ones reported in [1].

Modality	Percent Error (%)
Speech (S)	2
Face (F)	11

In this table, + sign denotes feature fusion, that is A+B means acceleration and brake features are concatenated and a bigger feature vector of dimension 4 is obtained. The results show that individually, each driving signal is not appropriate for biometric identification. However, feature fusion of acceleration and brake signals (A+B) achieves a respectable 10.2% error rate. This was also observed in [1].

In Table 2, we present decision fusion using fixed rules for various modalities. In this table the comma (,) sign indicates decision fusion, that is, classifier posterior probabilities are combined.

Table -2. Error rates (%) when using fixed combination rules for combining different modalities.

Modalities	Max	Min	Median	Mean	Product
A,B	28.2	14.5	14	14	11.2
A,B,E,V,W	43	31	41.5	22.7	23.5
A,B,W	38	22.5	30.7	18.7	16.2
A,B,F,S	9	3.2	0	0	1

The fixed combination rules are usually suboptimal since they do not consider relative reliability of individual modalities. We observe that among the fixed combiners, product and mean rules perform the best in general. Since acceleration and brake are more reliable amongst driving signals, it is possible to achieve better results by just using them instead of using all driving signals. When we include face and speech modalities, we can easily achieve close to 0% error even with these suboptimal fixed combining rules.

In Table 3, trainable combiner results are presented. The trainable combiners are trained using validation data that was set aside from training and testing data in the cross-validation procedure described above.

Table -3. Error rates (%) when using trainable combination methods

	NMC	Fisher	LDC	NB	Parzen
A,B	12.7	11.7	10.7	6.5	0.2
A,B,E,V,W	10.5	5.2	3	5	0
A,B,W	10	8.2	7	6.7	2
A,B,F,S	0	0	0.2	0	0

In trainable combiners, we achieve lower error rates in most cases due to validation data training. In this study, it appears that the validation and test data are very similar and overtraining combiners such as Parzen density based combiner work is clearly the best. It is well known that, Parzen classifier overfits to training data and does not easily generalize. However in this case, it seems that it is the most promising choice in comparison with other methods tried. Linear combiners such as LDC and Fisher also achieve respectable performance especially when the number of input classifiers are large.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have studied the performance of various combination methods for driver identification using driving behavior signals collected in a real-world scenario. The results show that individual driving signals are not largely indicative of the person by themselves. However, when we combine decisions from GMM classifiers of each driving signal using trainable combiners, we can achieve very low error rates in identifying the driver of the vehicle.

In the future, we plan to test these models in a larger subset of the CIAIR database.

ACKNOWLEDGEMENTS

The authors would like to thank Professors, Kazuya Takeda, Nobuo Kawaguchi, and Fumitada Itakura for providing the CIAIR database.

REFERENCES

[1] H Erdogan, A Ercil, HK Ekenel, SY Bilgin, I Eden, M Kirisci, H Abut, "Multi-modal person recognition for vehicular applications," N.C. Oza et al. (Eds.): MCS 2005, LNCS 3541, pp. 366 – 375, Monterey CA, Jun. 2005.

[2] N. Kawaguchi, S. Matsubara, I. Kishida, Y. Irie, H. Murao, Y. Yamaguchi, K. Takeda and F. Itakura, “Construction and Analysis of the Multi-layered In-car Spoken Dialogue Corpus,” Chapter 1 in *DSP in Vehicular and Mobile Systems,* Springer, New York, NY, 2005.

[3] K. Igarashi, C. Miyajima, K. Itou, K. Takeda, H. Abut and F. Itakura, “Biometric Identification Using Driving Behavior,” Proceedings IEEE ICME 2004, June 27-30, 2004, Taipei, Taiwan.

[4] D.A. Reynolds, “Speaker identification and verification using Gaussian mixture speaker models,” Speech Communications, 17, 91-108, 1995.

[5] A Dempster, N Laird, M Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm”, J. Royal Statistical Soc., 39, 1, 1978.

[6] J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining classifiers,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226-239, 1998.

[7] R.P.W. Duin, P. Juszczak, D. de Ridder, P. Paclik, E. Pekalska, and D.M.J. Tax, PRTools, a Matlab toolbox for pattern recognition, http://www.prtools.org, 2004.