A ROBUST APPROACH TOWARDS SEQUENTIAL DATA MODELING AND ITS APPLICATION IN AUTOMATIC GESTURE RECOGNITION

Sotirios Chatzis[1], *Dimitrios I. Kosmopoulos*[2], *Theodora Varvarigou*[1]

[1]National Technical University of Athens, 9, Iroon Polytechneiou Str., 15773 Athens, Greece
[2]NSCR Demokritos, P. Grigoriou & Neapoleos Str.,15310 Athens, Greece

Index Terms— Pattern classification, Pattern clustering methods, Pattern recognition

ABSTRACT

Hidden Markov models using finite Gaussian mixture models as their hidden state distributions have been applied in modeling of time series that result from various noisy signals. Nevertheless, Gaussian mixture models are well-known to be highly intolerant to the presence of outliers within the fitting sets used for their estimation. Finite Student's-t mixture models have recently emerged as a heavier-tailed, robust alternative to Gaussian mixture models, overcoming these hurdles. To exploit those merits of Student's-t mixture models, we introduce in this paper a novel hidden Markov chain model where the hidden state distributions are considered to be finite mixtures of multivariate Student's-t densities and we derive an algorithm for the model parameters estimation under a maximum likelihood framework. We apply this novel approach in automatic gesture recognition and we show that our model provides a substantial improvement in data representation performance and computational efficiency over the standard Gaussian model.

1. INTRODUCTION

The hidden Markov model (HMM) is increasingly being adopted in applications since it provides a convenient way of modeling observations appearing in a sequential manner and tending to cluster or to alternate between different possible components (subpopulations). Specifically, HMMs with continuous observation densities (continuous HMMs, CHMMs) have been used in a wide spectrum of applications in ecology, encryption, image understanding, speech recognition, handwriting recognition, emotion recognition based on facial expression classification, gesture recognition etc. [1].

The hidden observation densities associated with each state of a CHMM must be capable to approximate arbitrarily complex probability density functions. Finite Gaussian mixture models (GMMs) are the most common selection of hidden state distribution models in the CHMM literature, yielding the so-called Gaussian HMMs (GHMMs) [2]. Nevertheless, it is well-known that GMM estimation can be adversely affected by the presence of untypical data (outliers) in the data sets used for the model fitting. To address these issues, Peel et al. in [3] proposed the finite mixture of multivariate Student's-t distributions model (SMM) as a highly tolerant to outliers alternative to GMMs. It has been shown (see e.g. [3, 4]) that SMMs can model sufficiently well the hidden patterns of the data under examination, even under the presence of significant proportions of outliers, cases where the GMMs yield a relatively poor performance.

It is a natural consequence of the outlier intolerance of GMMs that CHMMs using GMMs as their hidden state densities do also suffer from the same outlier intolerance related issues. In modern CHMM literature, various efforts have been made towards the attenuation of these shortcomings of GHMMs (see e.g. [5, 6, 7]). However, all these methods have significant drawbacks, among which we might mention the heuristic nature of their majority as well as the application-specific nature of many of them. In this paper, we try to tackle these issues by proposing a novel CHMM where the hidden state distributions are modeled using finite mixtures of multivariate Student's-t densities, allowing for the exploitation of the outlier tolerance merits of SMMs in the context of sequential data modeling techniques using continuous hidden Markov chain models; the so-obtained Student's-t hidden Markov model (SHMM) provides the effective, computationally efficient and generic means for the outlier tolerant representation and classification of sequential data using CHMMs.

We consider the application of the SHMM model in automatic gesture recognition. Gaussian hidden Markov models are the standard statistical technique used to recognize and identify gestures [8]. Gesture recognition has many uses, such as helping surgeons perform operations and improving security, surveillance, and military applications. However, the related technology still faces major challenges one of the most significant being the problem of making these systems more accurate and robust to outliers, that inevitably comprise a significant proportion of the obtainable data, due to the nature of the gesture visual signals [8]. Under this motivation, we apply the SHMM in automatic gesture recognition, providing an insight in the advantages of the proposed model and showing that our model completely outperforms GHMMs.

The organization of the remainder of this paper is the following: in Section II the SHMM model is formulated. In Section III, a multiple token treatment of the SHMM under the maximum-likelihood framework is conducted, using the expectation-maximization algorithm. In Section IV, we apply the proposed model in automatic gesture recognition and show that it completely outperforms Gaussian HMMs. Finally, in the concluding section, the results of this paper are summarized and discussed.

2. THE PROPOSED MODEL

The adoption of the multivariate Student's-t distribution provides a way to broaden the Gaussian distribution for potential outliers. The probability density function (pdf) of a Student's-t distribution with mean vector $\boldsymbol{\mu}$, positive definite inner product matrix $\boldsymbol{\Sigma}$, and ν degrees of freedom is [9]

$$t(\boldsymbol{y}_t; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\boldsymbol{\Sigma}|^{-1/2} (\pi\nu)^{-p/2}}{\Gamma(\nu/2)\{1 + d(\boldsymbol{y}_t, \boldsymbol{\mu}; \boldsymbol{\Sigma})/\nu\}^{(\nu+p)/2}} \quad (1)$$

where p is the dimensionality of $\boldsymbol{y}_t$, $d(\boldsymbol{y}_t, \boldsymbol{\mu}; \boldsymbol{\Sigma})$ is the squared Mahalanobis distance between $\boldsymbol{y}_t, \boldsymbol{\mu}$ with covariance matrix $\boldsymbol{\Sigma}$, and $\Gamma(s)$ is the Gamma function.

The definition of the proposed Student's-t hidden Markov model (SHMM) is derived by assuming a finite state-space hidden Markov chain model where the hidden state densities are considered to be finite mixtures of Student's-t distributions.

The multiple token ML treatment of the SHMM model can be conducted by using the EM algorithm [10]. Let us consider M independent sequences of fitting data. We assume for convenience, that all the sequences have the same length T, i.e. comprising T data points, without any loss of generality. Let the m-th sequence be $\boldsymbol{y}_m = \{\boldsymbol{y}_{mt}\}_{t=1}^{T}, \; m = 1, ..., M$, where $\boldsymbol{y}_{mt}$ stands for the t-th data point of the m-th fitting sequence. Let us denote as $\boldsymbol{s}_{mt}$ the state indicator vectors of the observable data, where $\boldsymbol{s}_{mt} = (s_{1mt}, ..., s_{gmt})$, where s_{imt} is one or zero, according to whether $\boldsymbol{y}_{mt}$ is viewed as being emitted, or not, from the i-th state of the model $(i = 1, .., g)$. Let us also denote as $\boldsymbol{z}^i_{mt}$ the state-component indicator vectors of the observable data, where $\boldsymbol{z}^i_{mt} = (z^i_{1mt}, ..., z^i_{nmt})$, and $z^i_{jmt} = 1$ if, given that $\boldsymbol{y}_{mt}$ is emitted from the i-th state of the model, it holds that it is particularly generated from the j-th component density of the i-th state's hidden distribution, $z^i_{jmt} = 0$ otherwise.

Then, the probability density function of an observation $\boldsymbol{y}_{mt}$, given it is emitted from the i-th state of the model, is given by

$$p(\boldsymbol{y}_{mt}; \boldsymbol{\Theta}_i) = \sum_{j=1}^{n} c_{ij} t(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}, \nu_{ij}) \tag{2}$$

where, c_{ij}, $\boldsymbol{\mu}_{ij}$, $\boldsymbol{\Sigma}_{ij}$ and ν_{ij} are the mixing proportion, mean, precision and degrees of freedom of the j-the component density of the i-th state of the model, respectively. Furthermore, from the properties of the Student's-t distribution [9] it can be shown that, equivalently, it holds

$$p(\boldsymbol{y}_{mt} | \{u_{ijmt}\}_{j=1}^{n}; \boldsymbol{\Theta}_i) = \sum_{j=1}^{n} c_{ij} \mathcal{N}(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}/u_{ijmt}) \tag{3}$$

where, $\mathcal{N}(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij})$ stands for a Gaussian distribution, and u_{ijmt} is a Gamma-distributed precision scalar of the observable data point $\boldsymbol{y}_{mt}$, given it is generated from the j-th component density of the i-th hidden state distribution, and it holds

$$u_{ijmt} \sim \mathcal{G}\left(\frac{\nu_{ij}}{2}, \frac{\nu_{ij}}{2}\right) \tag{4}$$

As far as the application of the EM algorithm for the ML multiple token treatment of the SHMM model is concerned, letting the complete data corresponding to the m-th sequence, $\boldsymbol{y}_m^{comp}$, be the observable data $\boldsymbol{y}_{mt}, \; t = 1, ..., T$, $m = 1, ..., M$, their state indicator vectors, $\boldsymbol{s}_{mt}$, their state-component indicator vectors, $\boldsymbol{z}^i_{mt}$, and their corresponding precision scalars, u_{ijmt}, the complete data log-likelihood of the model reads

$$\begin{aligned} \log L_c(\boldsymbol{\Psi}) = \sum_{m=1}^{M} \sum_{i=1}^{g} \left[s_{im1} \log \pi_i + \sum_{t=1}^{T} s_{imt} \log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) \right] \\ + \sum_{m=1}^{M} \sum_{h=1}^{g} \sum_{i=1}^{g} \sum_{t=1}^{T-1} s_{hmt} s_{im,t+1} \log \pi_{hi} \end{aligned} \tag{5}$$

where $\boldsymbol{\Psi}$ is the parameter vector of the model, containing the c_{ij} and ν_{ij}, and the elements of the $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\Sigma}_{ij}$, $\boldsymbol{y}_{mt}^{comp}$ stands for the complete data corresponding to the t-th observation of the m-th sequence, $\boldsymbol{y}_{mt}$, and $\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i)$ is the complete data log-likelihood of the hidden distribution (SMM) of the i-th state of the SHMM model, with respect to the observation $\boldsymbol{y}_{mt}$. Concerning this latter quantity, from eq. (3) - (4) we yield

$$\begin{aligned} \log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) = \sum_{j=1}^{n} z^i_{jmt} \left\{ -\log \Gamma\left(\frac{\nu_{ij}}{2}\right) + \frac{\nu_{ij}}{2} \left[\log\left(\frac{\nu_{ij}}{2}\right) \right.\right. \\ \left. + \log u_{ijmt} - u_{ijmt} \right] - \frac{u_{ijmt}}{2} d(\boldsymbol{y}_{mt}, \boldsymbol{\mu}_{ij}; \boldsymbol{\Sigma}_{ij}) - \frac{1}{2} \log |\boldsymbol{\Sigma}_{ij}| \\ \left. + \log c_{ij} \right\} \end{aligned} \tag{6}$$

The E-step on the (k+1)-th iteration of the EM algorithm requires the calculation of the quantity

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = E_{\boldsymbol{\Psi}^{(k)}}(\log L_c(\boldsymbol{\Psi}) | \boldsymbol{y}) \tag{7}$$

which is the conditional expectation of the complete data log-likelihood given the fitting data, $\boldsymbol{y} = \{\boldsymbol{y}_m\}_{m=1}^{M}$, where $\boldsymbol{\Psi}^{(k)}$ denotes the *current* estimator (obtained by the k-th iteration of the EM algorithm) of $\boldsymbol{\Psi}$. Using (5) we obtain

$$\begin{aligned} Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = \sum_{m=1}^{M} \sum_{h=1}^{g} \left[\gamma_{hm1}^{(k)} \log \pi_h + \sum_{i=1}^{g} \sum_{t=1}^{T-1} \gamma_{himt}^{(k)} \log \pi_{hi} \right] \\ + \sum_{m=1}^{M} \sum_{i=1}^{g} \sum_{t=1}^{T} \gamma_{imt}^{(k)} E_{\boldsymbol{\Psi}^{(k)}} \left(\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i) | \boldsymbol{y} \right) \end{aligned} \tag{8}$$

where $\gamma_{imt}^{(k)}$ denote the current estimators of the state emission posterior probabilities, defined as

$$\gamma_{imt} \triangleq p(s_{imt} = 1 | \boldsymbol{y}) = p(s_{imt} = 1 | \boldsymbol{y}_m) \tag{9}$$

$(t = 1, ..., T)$ and $\gamma_{himt}^{(k)}$ denote the current estimators of the state transition posterior probabilities, defined as

$$\gamma_{himt} \triangleq p(s_{im,t+1} = 1, s_{hmt} = 1 | \boldsymbol{y}) \tag{10}$$

$(t = 1, ..., T - 1)$ for $m = 1, ..., M$, $h, i = 1, ..., g$.

Let us begin with the posterior probabilities γ_{imt} and γ_{himt}. The updates of these quantities can be computed on the (k+1)-th iteration of the EM algorithm utilizing the forward-backward recursions algorithm. It holds [2, 1]

$$\gamma_{himt}^{(k)} = \frac{a_{hmt}^{(k)} \pi_{hi}^{(k)} p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_i^{(k)}) b_{im,t+1}^{(k)}}{\sum_{\upsilon=1}^{g} \sum_{\phi=1}^{g} a_{\upsilon mt}^{(k)} \pi_{\upsilon\phi}^{(k)} p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_\phi^{(k)}) b_{\phi m,t+1}^{(k)}} \tag{11}$$

$(t = 1, ..., T - 1)$ and

$$\gamma_{imt}^{(k)} = \frac{a_{imt}^{(k)} b_{imt}^{(k)}}{\sum_{h=1}^{g} a_{hmt}^{(k)} b_{hmt}^{(k)}} \; (t = 1, ..., T) \tag{12}$$

where

$$a_{im1}^{(k)} = \pi_i^{(k)} p(\boldsymbol{y}_{m1}; \boldsymbol{\Theta}_i^{(k)}) \tag{13}$$

$$a_{im,t+1}^{(k)} = p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_i^{(k)}) \sum_{h=1}^{g} a_{hmt}^{(k)} \pi_{hi}^{(k)} \; (t = 1, .., T - 1) \tag{14}$$

$$b_{hmT}^{(k)} = 1 \tag{15}$$

$$b_{hmt}^{(k)} = \sum_{i=1}^{g} \pi_{hi}^{(k)} p(\boldsymbol{y}_{m,t+1}; \boldsymbol{\Theta}_i^{(k)}) b_{im,t+1} \ (t = T-1, ..., 1) \quad (16)$$

Finally, concerning the term $E_{\boldsymbol{\Psi}^{(k)}}\left(\log p(\boldsymbol{y}_{mt}^{comp}; \boldsymbol{\Theta}_i)|\boldsymbol{y}\right)$, from eq. (6) it can be shown that the derivation of this quantity is eventually reduced to the computation of the conditional posterior probabilities that $\boldsymbol{y}_{mt}$ is generated from the j-th component distribution of the i-th state of the SHMM, given that it is emitted from the i-th state of the model, yielding

$$\begin{aligned} \xi_{ijmt}^{(k)} &\triangleq E_{\boldsymbol{\Psi}^{(k)}}(z_{jmt}^{i}|\boldsymbol{y}_{mt}, s_{imt}=1) \\ &= \frac{c_{ij}^{(k)} t(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ij}^{(k)}, \boldsymbol{\Sigma}_{ij}^{(k)}, \nu_{ij}^{(k)})}{\sum_{h=1}^{n} c_{ih}^{(k)} t(\boldsymbol{y}_{mt}; \boldsymbol{\mu}_{ih}^{(k)}, \boldsymbol{\Sigma}_{ih}^{(k)}, \nu_{ih}^{(k)})} \end{aligned} \quad (17)$$

as well as the computation of the posterior expected values of the precision scalars, u_{ijmt}, of the observable data

$$\begin{aligned} u_{ijmt}^{(k)} &\triangleq E_{\boldsymbol{\Psi}^{(k)}}\left(u_{ijmt}|\boldsymbol{y}_{mt}\right) \\ &= \frac{\nu_{ij}^{(k)} + p}{\nu_{ij}^{(k)} + d(\boldsymbol{y}_{mt}, \boldsymbol{\mu}_{ij}^{(k)}; \boldsymbol{\Sigma}_{ij}^{(k)})} \end{aligned} \quad (18)$$

Further, the M-step of the multiple token EM fitting of the SHMM model is derived as follows

$$\pi_i^{(k+1)} = \frac{1}{M} \sum_{m=1}^{M} \gamma_{im1}^{(k)} \quad (19)$$

$$\pi_{hi}^{(k+1)} = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T-1} \gamma_{himt}^{(k)}}{\sum_{m=1}^{M} \sum_{t=1}^{T-1} \gamma_{hmt}^{(k)}} \quad (20)$$

$$c_{ij}^{(k+1)} = \sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} / \sum_{m=1}^{M} \sum_{t=1}^{T} \gamma_{imt}^{(k)} \quad (21)$$

$$\boldsymbol{\mu}_{ij}^{(k+1)} = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} u_{ijmt}^{(k)} \boldsymbol{y}_{mt}}{\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} u_{ijmt}^{(k)}} \quad (22)$$

$$\begin{aligned} \boldsymbol{\Sigma}_{ij}^{(k+1)} &= \sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} u_{ijmt}^{(k)} (\boldsymbol{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)})(\boldsymbol{y}_{mt} - \boldsymbol{\mu}_{ij}^{(k+1)})^T \\ &\times \left[\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} \right]^{-1} \end{aligned} \quad (23)$$

while, the degrees of freedom, ν_{ij}, are given by the solution of the equation

$$\begin{aligned} &1 - \psi\left(\frac{\nu_{ij}}{2}\right) + \log\left(\frac{\nu_{ij}}{2}\right) + \psi\left(\frac{\nu_{ij}^{(k)} + p}{2}\right) - \log\left(\frac{\nu_{ij}^{(k)} + p}{2}\right) \\ &+ \frac{1}{\sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)}} \sum_{m=1}^{M} \sum_{t=1}^{T} r_{ijmt}^{(k)} \left(\log u_{ijmt}^{(k)} - u_{ijmt}^{(k)}\right) = 0 \end{aligned} \quad (24)$$

where, $\psi(s)$ is the digamma function and $r_{ijmt}^{(k)}$ is the joint posterior probability that $\boldsymbol{y}_{mt}$ is generated from the i-th state of the model and particularly from its j-th component distribution, i.e. it holds

$$r_{ijmt} \triangleq p(s_{imt} = 1, z_{jmt}^{i} = 1|\boldsymbol{y}) = \gamma_{imt} \xi_{ijmt} \quad (25)$$

A final issue concerns the conception of a computationally efficient solution to the calculation of the likelihood $p(\boldsymbol{y}_1^{'}, ..., \boldsymbol{y}_T^{'}; \hat{\boldsymbol{\Psi}})$ of a sequence of data, $\boldsymbol{y}^{'} = \{\boldsymbol{y}_t^{'}\}_{t=1}^{T}$, with respect to a trained SHMM. Such a result, holding for every finite state-space hidden Markov chain model, ignorantly to the particular selection of its hidden state distributions, is obtained by application of the forward-backward algorithm; in particular, it holds [1, 2]

$$p(\boldsymbol{y}_1^{'}, ..., \boldsymbol{y}_T^{'}; \hat{\boldsymbol{\Psi}}) = \sum_{i=1}^{g} \hat{a}_{iT} \quad (26)$$

where, the $\hat{a}_{iT}$ are obtained by application of the forward-backward recursions algorithm for the given data sequence, $\boldsymbol{y}^{'}$, using the already obtained SHMM model parameters estimator, $\hat{\boldsymbol{\Psi}}$.

3. GESTURE RECOGNITION USING THE SHMM

We apply the proposed method in the problem of bimanual gesture recognition. For this purpose we use a test set comprising 16 different bimanual gestures, which are characterized by several self occlusions. More specifically, we experiment with the American Sign Language gestures for the words: "against", "aim","balloon", "bandit", "cake", "chair", "computer", "concentrate", "cross", "deaf", "explore", "hunt", "knife", "relay", "reverse" and "role".

The considered data set, publicly and freely available through the site *http://www.iit.demokritos.gr/∼dkosmo/downloads/gesture*, has been obtained by four different persons executing each one of these gestures. It comprises a training set, including 30 videos of variable duration per gesture, and a test set composed of 10 videos per gesture. Using these raw data, we extract a number of representative features suitable to formulate the observation vectors required to train and evaluate statistical models.

These considered features represent the relative position of the hands and the face in the images, as well as the shape of the respective skin regions and are derived using the complex Zernike moments (see e.g., [11]). The Zernike moments have been selected as region descriptors due to their high representation capabilities and noise resiliency. More specifically the observation is given by the vector

$$\boldsymbol{y} = (y_1, ..., y_8, y_9, ..., y_{16})$$

where the variables y_1 to y_8 represent the left hand region, and the rest the right hand. The y_1, y_2 are the distance and the angle of the left hand region from the center of gravity of all skin regions. y_3 to y_8 are the non-constant norms and arguments of the complex Zernike moments, up to third order. The rest vector elements for the right hand region are calculated in the same fashion.

The following assumptions additionally hold: (a) if a hand does not appear in the image, the related fields are set to zero, (b) if the hands occlude each other or both occlude the face, the observations corresponding to the two hands are identical and equal to the values extracted from the "common" region (c) if a hand occludes the face, the corresponding observations are calculated based on the "common" region. Thus we are able to represent the observations also in the case of occlusions. The skin segmentation can be done using a skin model and the region labeling using a method such as the one presented in [12].

In our experiments we use directly ground truth data to eliminate the effect of labeling errors. We model the considered gestures using one 3-state SHMM per gesture, considering diagonal precision matrices. The trained models are evaluated by classifying the test data to the trained models under a maximum a posteriori (MAP) probability classification fashion. To obtain some comparative results, we also train and evaluate a 3-state GHMM per gesture with

Table 1. Average Obtained Recognition Error Rate Over 30 Runs of the EM Algorithm

Gesture	SHMM with $n = 9$	GHMM with $n = 10$
against	0.23%	1.2%
aim	0.27%	0.89%
balloon	5.21%	7.94%
bandit	0.16%	0.92%
chair	15.24%	29.8%
cake	14.87%	20.35%
computer	7.1%	8.03%
concentrate	13.82%	20.07%
cross	19.7%	30.47%
deaf	0.48%	4.69%
explore	5.03%	7.73%
hunt	20.75%	33.12%
knife	13.53%	27.64%
relay	0.54%	6.85%
reverse	4.55%	10.05%
role	0.55%	2.23%
Average	7.62%	13.25%

diagonal precision matrices. This experiment is repeated for various numbers of component densities per state, n, and the lowest number yielding the highest recognition rate for each model is determined. In Table 1 we provide for each gesture the average recognition error rate obtained over 30 runs of the EM algorithm with different starting points, using the SHMM and GHMM models for the number of component densities per state, n, yielding their highest classification performance.

As we notice, using SHMMs to model the considered gestures we yield a considerably higher recognition performance comparing to the performance obtained by using GHMMs for a lower number of component densities per state. On average, the SHMM models yield a 42.49% lower recognition error rate comparing to the GHMMs.

4. CONCLUSIONS

In this paper we proposed a novel hidden Markov model, considering the selection of the HMM hidden state densities as being of multivariate Student's-t form. The proposed model is especially suitable for the statistical modeling of signals inherently containing significant proportions of artifacts and outliers, cases where the conventional Gaussian HMMs yield a rather poor performance.

Several applications that require modelling and classification of time series that contain outliers are expected to benefit from our work. The investigation of such applications is part of our future work. Here, we considered the application of this model in automatic gesture recognition, a challenging computer vision application where artifacts and outliers have significant effects in classification performance. Our results have provided the tangible evidence about the efficacy of our novel approach.

We finally outline that the proposed model imposes a computational burden comparable to that of the conventional GHMMs, as its major computational load stems from the calculation and inversion of empirical precision matrices, as is also the case with the GHMMs.

5. REFERENCES

[1] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics, New York, 2005.

[2] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 245–255, 1989.

[3] D. Peel and G. McLachlan, "Robust mixture modeling using the t distribution," *Statistics and Computing*, vol. 10, pp. 335–344, 2000.

[4] S. Chatzis, D. Kosmopoulos, and T. Varvarigou, "Signal modeling and classification using a robust latent space model based on t distributions," *IEEE Transactions on Signal Processing*, Accepted for future publication.

[5] L.M. Arslan and J.H.L. Hansen, "Selective training for hidden Markov models with applications to speech classification," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, pp. 46–54, 1999.

[6] C.-H. Lee and S.-Y. Lee, "Noise-robust speech recognition using top-down selective attention with an HMM classifier," *IEEE Signal Processing Letters*, vol. 14, no. 7, pp. 489–491, 2007.

[7] Jen-Tzung Chien and Chih-Pin Liao, "Maximum confidence hidden Markov modeling for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Accepted for future publication.

[8] D. Geer, "Will gesture-recognition technology point the way?," *Computer*, vol. 37, no. 10, pp. 20–23, 2004.

[9] C. Liu and D. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19–39, 1995.

[10] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, B*, vol. 39, no. 1, pp. 1–38, 1977.

[11] R. Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis: Theory and Applications*, World Scientific, 1998.

[12] D. Kosmopoulos and I. Maglogiannis, "Hand tracking for gesture recognition tasks using dynamic Bayesian network," *International Journal of Intelligent Systems and Applications*, vol. 1, no. 3/4, pp. 359–375, 2006.