

Automatic Mallampati Classification Using Active Appearance Models

Gabriel L. Cuendet, Anıl Yüce, Jean-Philippe Thiran
Signal Processing Laboratory (LTS5)
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
name.surname@epfl.ch

Matteo Sorci
nViso SA, Lausanne
Switzerland
matteo.sorci@nviso.ch

Patrick Schoettker, Christophe Perruchoud
Department of Anesthesiology,
University Hospital Center and University of Lausanne(CHUV), Switzerland
name.surname@chuv.ch

Abstract

Difficult tracheal intubation assessment is an important research topic in anesthesia as failed intubations are important causes of mortality in anesthetic practice. The modified Mallampati score is widely used, alone or in conjunction with other criteria, to predict the difficulty of intubation. This work presents an automatic method to assess the modified Mallampati score from an image of a patient with the mouth wide open. For this purpose we propose an active appearance models (AAM) based method and use linear support vector machines (SVM) to select a subset of relevant features obtained using the AAM. This feature selection step proves to be essential as it improves drastically the performance of classification, which is obtained using SVM with RBF kernel and majority voting. We test our method on images of 100 patients undergoing elective surgery and achieve 97.9% accuracy in the leave-one-out crossvalidation test and provide a key element to an automatic difficult intubation assessment system.

1. Introduction

Difficult tracheal intubation remains a constant and significant source of morbidity and mortality in anesthetic practice and failed intubations lead to complications ranging from airway trauma, severe hypoxemic brain injuries to death of the patient. Before anesthesia induction, accurate difficult airway identification allows optimal preparation, including dedicated equipment, experienced personnel or specific techniques in

order to avoid such complications.

Assessment of difficult intubation prior to anesthesia induction is an important research topic in anesthesia and several screening tests have been proposed. Among them, the modified Mallampati score [7, 9] is commonly used by physiologists to assess the difficulty of intubation. This score classifies the airway into 4 classes according to the visibility of the oro-pharyngeal structures observed on a patient opening the mouth and sticking his tongue out. The view is graded as follows (see fig.1): class I, soft palate, fauces, uvula, and pillars are visible; class II, soft palate, fauces, and uvula are visible; class III, soft palate and base of the uvula are visible; class IV, soft palate is not visible at all.

Although it has been shown to have little discriminative power in predicting tracheal intubation difficulty when used alone, the modified Mallampati test is still an important source of information when used in combination with other measures [6]. Among the various commonly used predictive models of difficult intubation, which use the modified Mallampati test, lies the Arné model where a simplified score is computed depending on certain physiological factors and the medical history of the patient [1]. Another similar scoring was put forward by Naguib et al. who performed a clinical study [8] to identify four risk factors that correlated with the difficult intubation, among which is the modified Mallampati score. As these and many other studies show, the Mallampati classification is an essential factor in the difficult intubation prediction, Mallampati score 1 and score 4 showing especially strong correlation with easy and difficult intubation respectively. Therefore, an automatic and objective classification of the modified Mallampati score is an important step in the process of

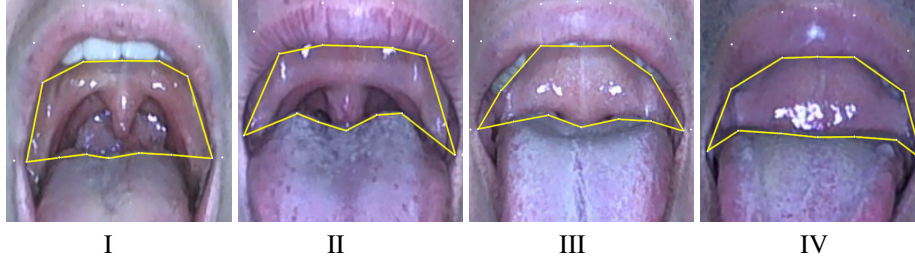


Figure 1. Modified Mallampati classification and AAM mask

developing an automatic difficult intubation assessment system. This will allow us to eliminate inaccurate classifications or inter-physician variations which are generally due to incorrect points of view.

In this work, we propose an effective method to assess the modified Mallampati score of patients from a frontal image of the head of the patient, with the mouth open and the tongue protruding to its maximum. For that purpose, we use active appearance models (AAM) to segment the inside of the mouth and describe the shape of the opening and the texture of the back of the throat. The most important coefficients of the projection of a new image on the AAM principal components are then used to perform classification using support vector machines (SVM).

The rest of the paper is organized as follows: Section II describes the proposed methodology, Section III contains information about the dataset and the data collection, Section IV details the results we obtain with our algorithm and finally Section V concludes the paper.

2. Methodology

2.1. Active appearance models

Active appearance models [3] are statistical models of deformable objects which contain both the shape and texture variation among a set of training images of the object. The training process of AAMs consists first of obtaining statistical shape and texture models separately by applying the Principle Component Analysis (PCA):

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s \quad \text{and} \quad \mathbf{g} = \bar{\mathbf{g}} + \Phi_t \mathbf{b}_t \quad (1)$$

where $\bar{\mathbf{s}}$ and Φ_s represent the mean and eigenvectors of the covariance matrix of the shape, and $\bar{\mathbf{g}}$ and Φ_t represent those of the texture. In order to obtain a complete model of appearance the model parameter vectors \mathbf{b}_s and \mathbf{b}_t are concatenated and a third PCA is applied to this concatenated vector:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Q}_s \mathbf{c} \quad \text{and} \quad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_t \mathbf{c} \quad (2)$$

where \mathbf{c} is the complete appearance model parameters vector, and \mathbf{Q}_s and \mathbf{Q}_t are the principal modes of the combined variation, retaining a certain amount of the total variance.

Using this model a new instance of the object can be generated by alternating the model parameters \mathbf{c} . The idea of the AAM search algorithm is then to synthesize a new example by the adjustment of model parameters, and it is generally treated as a minimization problem of the difference between the synthesized image and the original unseen image.

In this work, we define an AAM consisting of 12 points located on the lower edge of the upper lip (or upper incisors, depending on their visibility) and on the line on the back of the tongue such that the parts defining the modified Mallampati score are included in the object (as shown by the yellow contour in fig1). The AAM fits perfectly to the Mallampati classification case, not only because it efficiently segments the object and models the shape and texture variations among different subjects but it also includes certain preprocessing steps such as shape alignment and texture warping which make us invariant to factors like translation, rotation and scaling.

We have manually annotated 100 images of different subjects and trained an AAM using these manual annotations. Then, we back-project these manually annotated points and the texture inside onto the three different eigenspaces defined by the model and obtain for each subject the model parameter vectors \mathbf{b}_s , \mathbf{b}_t , \mathbf{c} which constitute our complete set of features.

At this stage, we use the manual annotations of the mouth to calculate the model parameters to exclude the effect of model fitting accuracy in the classification process. In the future, using AAM will allow to automatically segment the contour of the mouth by fitting the model, providing full automatization of the system.

2.2. Feature selection and classification

Once we obtain the full set of features (the three different model parameter vectors), we perform a selection

of features on these three sets separately. By discarding irrelevant and redundant features, feature selection provides performance improvement in classification. This is due to the fact that the AAM parameters are ordered depending on the ratio of the total variation they explain and since this variation is not necessarily caused by the different Mallampati classes, certain coefficients introduce noise if taken into account. Feature selection is thus a crucial step in the classification process.

In order to select the most relevant subset of features, we train linear support vector machines (SVM) in a recursive manner, removing one feature at each iteration (backward feature elimination), similarly to what is done in [4]. Linear SVM is a supervised learning method used for binary classification. The model resulting from a linear SVM is a hyperplane of the form:

$$\mathbf{w} \cdot \mathbf{x} - a = 0 \quad (3)$$

which maximizes its distance to the nearest training data point of both classes. The normal vector to the hyperplane, \mathbf{w} , can be seen as features weights where the highest weight indicates the feature that contributes the most to separating the two classes. At each iteration ordering the features in decreasing order of weight w_i and eliminating the feature with the lowest weight allows obtaining a ranking of the features. As linear SVM is a binary classifier, six different classifiers are trained, in a 1-against-1 fashion, resulting in six different rankings of features.

Then once we obtain the feature subsets using these rankings, we train six different SVM with RBF kernel using the publicly available LibSVM implementation [2]. Once again the SVM are trained in a 1-against-1 fashion as better results are generally obtained by this method, compared to other multiclass SVM such as 1-against-all [5]. Details of the cross-validation and parameter optimization are presented in the results section. The final classification of the modified Mallampati score is then obtained by the majority voting of these 6 classifiers.

3. Dataset

The dataset used is composed of 100 images of different subjects, equally balanced between classes. The images are acquired at the University Hospital in Lausanne (CHUV), and the subjects are actual patients who undergo the regular preoperative assessment for anesthesia prior to their elective surgeries. The recording process of the images is part of a larger project on the automatic assessment of difficult intubation. The subjects included in the dataset are aged between 24 and 81 and the proportion of female subjects is 39%.

The assessment of the ground truth for the modified Mallampati score is then performed by experienced anesthesiologists only on the basis of these images. The Mallampati classification depends highly on the angle of view of the mouth in the images. The images were taken by trained staff such that the head is positioned to obtain the best visibility of the oropharyngeal features. Once we obtain an accurate image based Mallampati classification, we will use videos of patients rotating their heads vertically to classify each frame and assess the lowest score obtained in the video, which corresponds to the optimal view, for that patient.

4. Results and discussion

In this section we report the results of the classification using the leave-one-subject-out cross validation method. For each of the 100 subjects we train six different support vector machines (one class against another). Each time the kernel and regularization parameters of the SVM are optimized using a 5-fold cross validation on the 99 samples in the training set. The corresponding sample that was left out is then classified by the six binary SVM and the final modified Mallampati score is obtained by majority voting.

Feature selection is a key step in the proposed method as explained in Section 2.2. Indeed, we see from the analysis of the feature rankings that, in general, the coefficients corresponding to principle modes explaining a very small portion of the total variance are assigned higher weights. The optimal number of features used by each of the six SVM is experimentally determined by comparing the overall accuracy obtained by using different numbers of features, which is shown in fig. 2.

We have performed the tests using the coefficients obtained from the shape model, texture model and the combined appearance model separately to identify which type of features is the most efficient in discriminating the different Mallampati classes. For each model we keep a number of principal components explaining more than 99.99% of the total variance, resulting in a total of 23 shape features, 100 texture features and 99 combined features, which are then ranked using the linear SVM method explained in Section 2.2.

Intuitively, the information about the modified Mallampati score is contained mainly in the texture rather than the shape of the mouth opening. This hypothesis is confirmed by the poor results obtained using only the coefficients \mathbf{b}_s modelling the variations in the shape, while using only the coefficients \mathbf{b}_t leads to performances of the same quality as using the coefficients \mathbf{c} , modelling the complete appearance. It can thus be con-

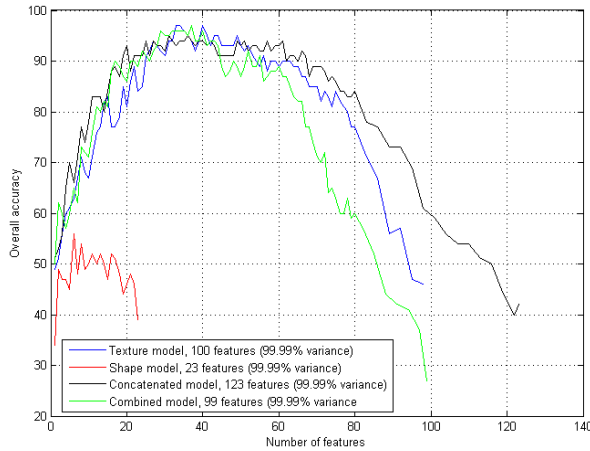


Figure 2. Classification accuracy vs number of features

cluded that taking into account the shape does not help to improve the classification performance (see fig. 2).

The best classification performance is obtained using 33 features of the texture model. Table 1 presents the confusion table for the corresponding leave-one-out cross validation test. The classification of 3 of the 100 samples was ambiguous due to an equal number of votes in the majority voting scheme. These samples are discarded in the calculation of the final accuracy and not included in the confusion table. In order to avoid such ambiguities, future work will include a probabilistic weighting of each classifier in the voting scheme. 95 of the rest of the 97 are correctly classified, corresponding to a 97.94% overall accuracy and 100% recall and precision for Mallampati class 4, which is an important indicator of difficult intubation.

5. Conclusion

In this paper, we proposed an AAM based method to assess the modified Mallampati score of patients from an image of the mouth cavity. We select the relevant features obtained by the AAM using linear SVM and obtain the classification by the majority voting of six different binary SVM classifiers. We perform tests on images of 100 patients and show that with the optimal number of features we can correctly classify 95% of the total samples, taking into account the 3 samples that were ambiguously classified.

To our knowledge this is the first work proposing an automatic system to assess the modified Mallampati score from images. The modified Mallampati score is often criticized for the lack of objectivity in the way practitioners assess it, especially due to the angle of view. This leads to different scores on the same patient,

Table 1. Confusion Table, OA=0.979

	1	2	3	4	
1	21	2	0	0	0.913
2	0	25	0	0	1
3	0	0	24	0	1
4	0	0	0	25	1
	1	0.926	1	1	

when examined by different practitioners. The proposed image based method can be extended to analyse videos and will allow objectively assessing the modified Mallampati score. This work thus provides an essential element to an automatic difficult intubation assessment system.

Acknowledgements

This project is supported by the Commission for Technology and Innovation (CTI) of the Swiss Confederation (Project No: 12636.1).

References

- [1] J. Arne, P. Descoins, and J. Fusciardi. Preoperative assessment for difficult intubation in general and ENT surgery: predictive value of a clinical multivariate risk index. *British Journal of Anaesthesia*, 80:140–146, 1998.
- [2] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):681–685, 2001.
- [4] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46:389–422, 2002.
- [5] C. Hsu and C. Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE transactions on*, 13(2):415–425, 2002.
- [6] L. H. Lundstrom, M. Vester-Andersen, A. M. Moller, S. Charuluxananan, J. L’Hermite, and J. Wetterslev. Poor prognostic value of the modified mallampati score: a meta-analysis involving 177 088 patients. *British Journal of Anaesthesia*, 107(5):659–667, 2011.
- [7] S. Mallampati, S. Gatt, and L. Gugino. A clinical sign to predict difficult tracheal intubation: a prospective study. *Canadian Journal of Anesthesia*, 32:429–434, 1985.
- [8] M. Naguib, T. Malabarey, R. A. AlSatli, S. A. Damegh, and A. H. Samarkandi. Predictive models for difficult laryngoscopy and intubation. a clinical, radiologic and three-dimensional computer imaging study. *Canadian Journal of Anesthesia*, 46(8):748–759, 1999.
- [9] G. L. T. Samsoun and J. R. B. Young. Difficult tracheal intubation: a retrospective study. *Anaesthesia*, 42(5):487–490, 1987.