Multiple feature sets based categorization of laryngeal images

A. Verikas\textsuperscript{a,b,\ast}, A. Gelzinis\textsuperscript{a}, D. Valincius\textsuperscript{a}, M. Bacauskiene\textsuperscript{a}, V. Uloza\textsuperscript{c}

\textsuperscript{a}Department of Applied Electronics, Kaunas University of Technology, Studentu 50, LT-51368 Kaunas, Lithuania
\textsuperscript{b}Intelligent Systems Laboratory, Halmstad University, Box 823, S-301 18 Halmstad, Sweden
\textsuperscript{c}Department of Otolaryngology, Kaunas University of Medicine, LT-50009 Kaunas, Lithuania

\textbf{A B S T R A C T}

This paper is concerned with an automated analysis of laryngeal images aiming to categorize the images into three decision classes, namely healthy, nodular, and diffuse. The problem is treated as an image analysis and classification task. Aiming to obtain a comprehensive description of laryngeal images, multiple feature sets exploiting information on image colour, texture, geometry, image intensity gradient direction, and frequency content are extracted. A separate support vector machine (SVM) is used to categorize features of each type into the decision classes. The final image categorization is then obtained based on the decisions provided by a committee of support vector machines. Bearing in mind a high similarity of the decision classes, the correct classification rate of over 94\% obtained when testing the system on 785 laryngeal images recorded at the Department of Otolaryngology, Kaunas University of Medicine is rather promising.

© 2006 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The diagnostic procedure of laryngeal diseases in clinical practice is rather complex and based on evaluation of patient's complaints, history, and data of instrumental and histological examination. During the last 2 decades clinicians and researchers have developed a variety of imaging techniques for examination of the larynx and objective measurements of voice quality [1,2]. Evaluation of larynx has improved significantly with the establishment of the computed tomography (CT) and magnetic resonance imaging (MRI), as the technologies provide insights into the endoscopically blind areas and reveal depth of tumour infiltration as well as cartilage and bone marrow invasion. The technologies may be beneficial in staging larynx carcinoma and planning the most appropriate surgical procedure [3–6]. Ultrasonography is useful in cases of larger laryngeal lesions and may have some role in screening unilateral vocal fold pathologies. At the same time, further fine-tuning of the technique may be necessary [7,8].

However, visualization of the larynx, by performing indirect video laryngostroboscopy and direct micro-laryngoscopy remains one of the most important diagnostic procedures, especially when making a primary diagnosis and then planning other more sophisticated diagnostic actions and the treatment. A physician evaluates colour, shape, geometry, contrast, irregularity, and roughness of the visual appearance of vocal folds. This type of examination is rather subjective and to a great extent depends on physician’s experience.
Availability of objective measures of these features would be very helpful for assuring objective analysis of laryngeal images and creating systematic databases for education, research, and health care purposes.

In addition to the data obtained from one particular patient, experience plays also a very important role in the decision-making process. However, a physician interpreting the data from a particular patient may have a limited knowledge and experience in the analysis of the data. Moreover, to exploit the experience, a tedious and time-consuming analysis of a large amount of data (images for example) may be required sometimes. In such a situation, a decision support system for an automated analysis and interpretation of medical data is of great value. Therefore, physicians would very much appreciate having a system able to automatically categorize the laryngeal images into several decision classes corresponding to different diseases. Such a system would enable a physician to effectively exploit a priori information available from a database of laryngeal images in screening laryngeal diseases.

Due to a large variety of appearance of vocal folds, the categorization task is sometimes difficult even for a trained physician [9,10]. Fig. 1 presents three examples of laryngeal images. The image placed on the right-hand side of the figure comes from the diseased class, while the other two are taken from the healthy vocal folds. In this case, the only discriminative feature is the slightly convex vocal fold edges in the upper part of the image coming from the diseased class.

Attempts made to develop computer-aided systems for analyzing vocal fold images are few. In Ref. [11], a system for automated categorization of manually marked suspect lesions into healthy and diseased classes is presented. The categorization is based on textural features extracted from co-occurrence matrices computed from manually marked areas of vocal fold images. A correct classification rate of 81.4% was reported when testing the system on a very small set of 35 images. In our previous studies [12,13], a committee of multi-layer perceptrons employed for categorizing vocal fold images into three decision classes correctly classified over 90% of test set images.

In this study, aiming to obtain a comprehensive description of laryngeal images, we extract multiple feature sets coming from different processing approaches. Image colour distribution, distribution of the image intensity gradient direction, parameters characterizing the geometry of edges of vocal folds, distribution of the spectrum of the Fourier transform of the colour image complex representation, and parameters calculated from multiple co-occurrence matrices are the feature types used to describe laryngeal images. By extracting different types of features, we aim to exploit various image domains that could prove useful for image categorization into three decision classes. Description of the classes will be given shortly. Aiming to better exploit information contained in extracted features, the kernel principal components [14,15] are used for classification instead of the original features. A separate support vector machine (SVM) is used to categorize kernel principal components computed from features of each type into the decision classes. The final image categorization is then obtained based on the decisions provided by a committee of support vector machines.

The variety of feature types used to characterize laryngeal images, the KPCA based preprocessing, and the SVM committee based classification are the main differences between the technique utilized in this work and the studies presented in Refs. [11–13]. In Ref. [11], ordinary co-occurrence matrix based features have been utilized. The Fourier power spectrum, image intensity gradient, edge geometry, and the co-occurrence matrix based features are the new feature types utilized in this study if compared to those exploited in Refs. [12,13]. The processing approach applied allowed to increase the classification accuracy of laryngeal images.

The remainder of the paper is organized as follows. In the next section, we briefly describe the data used. Section 3 outlines the analysis techniques employed. Section 4 presents the results of the experimental investigations. Finally, conclusions of the work are given in Section 5.

2. Images used

A set of 785 laryngeal images recorded at the Department of Otolaryngology, Kaunas University of Medicine during the period from October 2002 to December 2003 is used in this study. The images were acquired during routine direct micro-laryngoscopy employing the Möller-Wedel Universa 300 surgical microscope. The 3-CCD Elmo colour video camera of 768 × 576 pixels was used to record the images. To lessen the influence of variation of the image capturing conditions on image appearance, we apply the multi-scale retinex theory-based colour image enhancement [16,17]. Details on how the enhancement has been applied can be found in Ref. [12].

A rather common, clinically discriminative group of laryngeal diseases was chosen for the analysis, i.e. mass lesions of vocal folds. In the study group, diagnosis of the mass lesions of the vocal folds – the gold standard – was based on routine clinical examination of the patient, including patient’s complaints, history, voice assessment, indirect and direct micro-laryngoscopy, and finalized (confirmed) by the histological examination.

Mass lesions of vocal folds could be categorized into six classes, namely polypus, papillomata, carcinoma, cysts, kerato-
sis, and nodules. This categorization is based on clinical signs and a histological structure of the mass lesions of vocal folds [18,9,10]. In this study, the task was to differentiate between the healthy (normal) class and two groups of mass lesions of vocal folds, i.e. nodular – nodules, polyps, and cysts – and diffuse – papillomata, keratosis, and carcinoma – lesions. Thus, including the healthy class, we have to distinguish between three classes of images. Amongst the 785 images available, there are 49 images from the healthy class, 406 from the nodular class, and 330 from the diffuse class. Certainly, the healthy class is underpopulated. However, the database is continuously updated and this deficiency will be mitigated in the near future. Moreover, our experience shows that, in general, the variability of the healthy class images is much lower than those coming from the other two classes. Visual signs of typical representatives of vocal fold mass lesions (colour, shape, surface, margins, size, localization) are rather typical, clinically evident, and descriptive. However, it is worth noting that due to the large variety of appearance of vocal fold mass lesions, the classification task is difficult even for a trained physician [9,10]. Fig. 1 provides an example of such a task. Fig. 2 presents characteristic examples from the three decision classes considered, namely nodular, diffuse, and healthy.

3. Methods

The approach adopted in this paper can be summarized in the following three steps:

(1) Comprehensive image characterization by multiple sets of features obtained by applying different processing approaches.
(2) Multiple image categorization into the decision classes using separate SVM for each type of the extracted features.
(3) Final image categorization based on the decisions provided by the committee of the SVMs.

3.1. Feature extraction

Features of each type are collected into a separate vector. Thus, the values of measurements related to image colour (c), texture (t), geometry (g), and frequency content (f1), (f2) are collected into six separate vectors: \( \xi_c, \xi_t, \xi_g, \xi_f1, \) and \( \xi_f2 \). Observe that there are two types of frequency content based features. The feature vectors \( \xi_c, \xi_t, \xi_g, \xi_f1, \) and \( \xi_f2 \) are then obtained by applying the kernel principal component analysis (KPCA) [14,15] separately for each of the spaces spanned by \( \psi_c, \psi_t, \psi_g, \psi_f1, \) and \( \psi_f2 \).

3.1.1. Kernel principal components

Assume that \( \kappa \) is a Mercer kernel [15] and \( \phi \) is a mapping of \( \psi \) onto a feature space \( F \), such that \( \kappa(\psi_i, \psi_j) = \langle \phi(\psi_i), \phi(\psi_j) \rangle \), where \( \langle \cdot, \cdot \rangle \) stands for the inner product and \( \phi(\psi) \) denotes a centered data point in the feature space \( F \). How the centering is performed in the feature space can be found in Ref. [15]. The features \( \xi \) are then given by the kernel principal components computed as projections of \( \phi(\psi) \) onto the eigenvectors

\[
v = \sum_{i=1}^{M} a_i \phi(\psi_i)\tag{1}\]

of the covariance matrix \( \kappa_{ij} = \langle \phi(\psi_i), \phi(\psi_j) \rangle \), where \( M \) is the number of data points and the expansion coefficients \( a_i \) of the eigenvector are found from the eigenvalue problem \( \lambda \alpha = K \alpha \) \( \tag{2} \)

where the matrix \( K \) is given by \( \kappa_{ij} \). Thus, the feature \( \xi \) is given by

\[
\xi = \langle v, \phi(\psi) \rangle = \sum_{i=1}^{M} a_i \kappa_{ij}\tag{3}\]

Polynomial kernels

\[
k_{ij}(\psi_i, \psi_j) = (1 + \psi_i^{T} \psi_j)^n \tag{4}\]

of order \( n = 1, 2, 3 \) have been tested in both KPCA and SVM based classification. The second order kernel provided the best performance. In this study, we provide the results obtained using the second order polynomial kernel.

Unlike linear PCA, KPCA allows extracting a number of components, which can exceed the input dimensionality. KPCA can find up to \( M \) non-zero eigenvalues. It has been demonstrated experimentally that an SVM trained using a large number of kernel principal components (exceeding the input dimensionality) can significantly outperform the one trained using any number of linear principal components [14]. By applying the KPCA we do not aim reducing input dimensionality. The aim is to extract features by applying the non-linear analysis. Nonetheless, having kernel principal components extracted, we attempt reducing the dimensionality of the new transformed space, which is equal to the number of training samples. The feature selection procedure we apply in this study is very simple. Starting from the component corresponding to the smallest eigenvalue, components are eliminated one by one. The number of components retained

---

Fig. 2 – Images from the nodular (left), diffuse (middle), and healthy (right) classes.
is equal to the number providing the highest correct classification rate on the cross-validation set. The same procedure is applied for all the feature types. By no doubts, a component corresponding to a small eigenvalue can be important for classification. Thus, applying a more advanced feature selection procedure one can probably increase the classification accuracy. However, feature selection process for large feature sets is rather time consuming. Therefore, we resorted to this simple feature selection procedure in this study.

### 3.1.2. Colour features

The approximately uniform L’a’b’ colour space [19] was employed to represent colours. We characterize the colour content of an image by the probability distribution of the colour represented by the 3-D colour histogram of N = 4096 (16 × 16 × 16) bins and consider the histogram as an N-vector. Most of bins of the histograms were empty or almost empty. Therefore, to reduce the number of components of the N-vector, the histograms built from a set of training images were summed up and the N-vector components corresponding to the bins containing less than Nc hits in the summed histogram were left aside. Hereby, when using Nc = 10 we were left with 733 bins—a ψc vector of 733 components. The colour features ξc are then given by the kernel principal components computed using ψc.

### 3.1.3. Texture features

Regarding characterization of texture of vocal fold images, the multi-channel 2-D Gabor filtering, co-occurrence matrices, run-length matrices, and the singular value decomposition based approaches have been applied in previous studies [11–13,20]. Amongst those, the Gabor filtering and the co-occurrence matrices based approaches proved to provide the best and approximately the same performance. Since the co-occurrence matrices based approach is much less time consuming, we resorted to this type of texture features in this work.

It is common practice to utilize the 14 well-known Haralick’s coefficients [21] as the co-occurrence matrix based features. The coefficients are usually calculated from the average co-occurrence matrix obtained by averaging the matrices calculated for 0°, 45°, 90°, and 135° directions. The matrices are computed for one or several, experimentally selected, distance parameter values. In this work, a different approach to exploiting information available in the co-occurrence matrices was adopted. A polynomial p(x) of degree n

\[ p(x) = p_0 + p_1 x + \ldots + p_{n-1} x^{n-1} + p_n x^n \]

has been fitted to the values of each of the 14 coefficients calculated from the co-occurrence matrices evaluated for several distance parameter values. The distance parameter d values utilized to calculate the co-occurrence matrices were: d = 1, 3, 5, 7, 9, 11, 13, 15. Parameters of the polynomials were then used as components of the measurement vector ψf. Thus, the number of components in the measurement vector ψf is equal to (n + 1) × 14. Performance of the first to fourth order polynomials, defined by Eq. (5), have been investigated.

The set of texture features defined by the second order polynomial provided the best performance. Fig. 3 presents an example of the plot of the fit of second, third, and fourth order polynomials to the values of one of the coefficients (difference variance) computed using images from the three decision classes.

### 3.1.4. Fourier spectrum based features

There are two types of frequency content based features ξf1 and ξf2. Let us assume that Z(u, v) is the Fourier transform of the image

\[ z(x, y) = L(x, y) \exp(jH_{ab}(x, y)) \]

where z(x, y) is a complex colour representation of the colour image L(x, y), a’(x, y), b’(x, y), and H_{ab}(x, y) = \arctan[b’(x, y)/a’(x, y)] is the CIE hue-angle [19]. We define the Fourier spectrum of the image z(x, y) as:

\[ P(u, v) = ||Z(u, v)||^2 = R^2(u, v) + I^2(u, v) \]

where R(u, v) and I(u, v) are the real and imaginary parts of Z(u, v), respectively. Figs. 4–6 present the Fourier spectra calculated for the images shown in Fig. 2.

To compute the feature vector ξf1, the upper part of the frequency plane is divided into M equidistant wedges W_i and the average power

\[ \bar{P}_i = \frac{1}{N_{W_i}} \sum_{u,v \in W_i} P(u, v) \quad i = 1, \ldots, M \]

is computed in each of the wedges, where N_{W_i} is the number of distinct frequencies in the wedge W_i. The \bar{P}_i values constitute the measurement vector ψf1. Fig. 7 presents an example of the upper part of the Fourier spectrum obtained from a laryngeal image along with two lines delineating a wedge around the vertical axis.

To extract the image frequency content based features of the second type, the frequency plane is divided into several rings R_i of different average frequency. The following way of

---

**Fig. 3** – The second, third, and fourth order polynomials fitted to the values of the “difference variance” coefficient using images from the three decision classes.
partitioning has been applied:

\[ u, v \in R_i \text{ if } iW \leq \log^2(\sqrt{u^2 + v^2}) < (i + 1)W \tag{9} \]

where

\[ W = \frac{\log^2(\sqrt{N^2/2})}{N_{\text{ring}}} \tag{10} \]

with \( N_{\text{ring}} \) being the number of rings of different average frequency and \( N = \max(N_x, N_y) \), where \( N_x \) and \( N_y \) is the image size in the horizontal and vertical direction, respectively. The partitioning can be viewed as band selection in the frequency domain. The Chi-square \( \chi_i \) and the entropy \( M_i \) of the Fourier power are then computed in each of the rings and used as the image frequency content based measurements of the second type (F2).

The Chi-square value in the \( i \)th frequency ring is computed as

\[ \chi_i = N_i^2 \sum_{u,v \in R_i} \left( \frac{P_i(u,v) - \frac{1}{N_i}}{N_i} \right)^2 \tag{11} \]

where \( N_i \) is the number of distinct frequencies in \( R_i \) and

\[ P_i(u,v) = \frac{P_i(u,v)}{\sum_{u,v \in R_i} P_i(u,v)} \tag{12} \]

The Chi-square measures the difference between the power spectrum of the ring \( R_i \) and the “white noise spectrum”. When the texture is random the Chi-square value approaches zero. For the ordered texture, the \( P_i(u,v) \) values spread further from \( 1/N_i \) and the Chi-square value approaches \( N_i^2 \).

The entropy measure we use is given by

\[ M_i = -\frac{1}{\log N_i} \sum_{u,v \in R_i} P_i(u,v) \log P_i(u,v) \tag{13} \]

The \( \chi_i \) and \( M_i \) values computed for each of the \( R_i \) rings constitute the measurement vector \( \Phi_2 \). The Fourier spectrum based
features $\xi_1$ and $\xi_2$ are given by the kernel principal components computed using $\psi_{f1}$ and $\psi_{f2}$, respectively.

3.1.5. **Image intensity gradient based features**
The features are based on the distribution of the image intensity gradient direction. The direction angle $\alpha(x, y)$ of the image intensity gradient vector $\nabla I$ is given by:

$$\alpha(x, y) = \tan^{-1} \left( \frac{G_x}{G_y} \right)$$

where $\nabla L = [G_x \ G_y]^T = [\partial I/\partial x \ \partial I/\partial y]^T$, the Sobel operator, which consists of a pair of $3 \times 3$ convolution masks [22] has been used to estimate the image intensity gradient. We use a histogram to represent the distribution of the angle $\alpha(x, y)$. The histogram vector $\psi_G$ is then projected onto the space spanned by the eigenvectors obtained from the kernel principal component analysis and the vector of the principal components $\xi_G$ is utilized as a feature vector of this type.

3.1.6. **Geometrical features**
Geometrical features we use are mainly targeted to characterize the shape of edges of two vocal folds. Observe that laryngeal images recorded by the system used in this study always show approximately the same view. Two thirds of the lower part of laryngeal images are used in the analysis. In general the upper part of the images is more informative than the lower one. However, the upper part is relatively well characterized by the colour and texture features. Moreover, usually there are many spurious edges in the upper part of the images. Convex regions of vocal fold edges in the lower part of the images are the information sources utilized in the analysis. Various techniques can be applied to extract colour edges [23]. In this work, we exploited the technique based on the difference vector operators [23]. The pixel $(x_0, y_0)$ in the gradient image $g(x, y)$ is assumed to be an edge pixel if $g(x_0, y_0) > g_e$, where $g_e$ is a threshold specific for each image analyzed. To find the threshold $g_e$, gradient values computed for a given image were normalized to fall into the interval $[0–255]$ and a histogram of the normalized values was constructed. The threshold $g_e$ was then given by

$$g_e = k + \delta$$

where

$$k = \arg \max_{i=0 \ldots 255} h(i)$$

where $h(i)$ stands for the histogram and $\delta$ is a parameter chosen experimentally. The value of $\delta = 4$ has been used in all the tests. There were many spurious edge pixels left even after the thresholding. Therefore, in the next step, all the edge pixels were sorted out into connected components [22] and the small ones were eliminated. A connected component containing less than 30 pixels was considered as being small. Two large connected components, one on the left- and the other on right-hand side of the image being analyzed were then found and used in further analysis. When looking for the large connected components, the knowledge of the fact that there are no or very few edge pixels between pixels belonging to the large components was exploited.

Having the two connected components, three polynomial curves given by Eq. (17) – one of the first, one of the second, and one of the third order – were fitted to pixels of each of the components

$$p(x) = p_0 + p_1 x + \cdots + p_{n-1} x^{n-1} + p_n x^n \quad n = 1, 2, 3$$

Thus, in total, we have 18 parameters $p_i$ characterizing the six curves. The parameters $p_i$ constitute the measurement vector $\psi_C$. The vector of geometrical features $\xi_G$ is given by the kernel principal components computed using $\psi_C$. Fig. 8 presents two examples of laryngeal images coming from the healthy and nodular classes along with the third order polynomial curves found.

3.2. **Pattern classifiers**
Classification rules of two types, namely a support vector machine, and a committee of SVMs have been utilized in this work. The SVM is known as having an ability to solve a classification task at hand by maximizing the classifier margin. The discriminant function of a two-class – binary – SVM is given by

$$f(\xi) = \text{sgn}(\hat{y})$$

where

$$\hat{y} = \sum_{j=1}^{M} \alpha_j^* y_j (\xi_j, \xi) + b$$

where $\phi(\xi, \xi)$ is a kernel, $M$ the number of data points, $\text{sgn}$ stands for the sign function, $y_j$ is a target value ($y_j = \pm 1$), and values of the parameters $b$ and $\alpha_j^*$ are found as a solution to the optimization problem defined by the type of SVM used. In this work, we used the 1-norm soft margin SVM [15]. The parameters $\alpha_i$ satisfy the following constrains:

$$\sum_{i=1}^{M} \alpha_i y_i = 0, \quad \sum_{i=1}^{M} \alpha_i = 1, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, M$$

with $C$ being the regularization constant. To find the optimal value of the constant, we applied the Simplex method [24].
Support vector machines with the polynomial kernel of degree one to three have been investigated as pattern classifiers. An SVM with the first degree kernel implements a linear decision boundary in the transformed space. Since the transformed space is obtained by applying the kernel principal component analysis with the second degree polynomial kernel, the obtained decision boundary is non-linear in the initial space. Thus, one can expect that it is enough to use a linear SVM, with the first degree polynomial kernel, for example. However, our experimental investigations performed have shown, that SVMs with the second and third degree polynomial kernels outperformed the ones built using the first degree kernel. Since SVMs with the second and third degree kernels provided approximately the same performance, the second degree kernel has been used in further tests.

To distinguish between three classes of images, we utilized the one versus one pair-wise classification scheme. The following rule has been used to calculate the output value for the ith class \( \hat{y}_i(\xi) \) – the estimate of the probability of a sample \( \xi \) to belong to the class i – based on the output values obtained from the binary SVMs:

\[
\hat{y}_i(\xi) = \frac{\sqrt{\text{card}(S_i(\xi))} \sum_{k \in S_i(\xi)} |\hat{y}_k(\xi)|}{\sum_{m=1}^{Q} \sqrt{\text{card}(S_m(\xi))} \sum_{k \in S_m} |\hat{y}_k(\xi)|}
\]

where Q is the number of classes, \( \hat{y}_k(\xi) \) the output value of the kth binary SVM calculated using Eq. (19), \( S_i(\xi) \) the set of binary SVMs that have assigned \( \xi \) to the ith class, and card stands for the cardinality of the set.

Numerous previous studies on classification committees have demonstrated that, in most of the cases, a classification committee outperforms a single member of the committee [25–28]. Therefore, a committee of SVMs has also been used in this study. However, an efficient committee should consist of members that are not only very accurate, but also diverse in the sense that the classifier errors occur in different regions of the input space [29]. Splitting or splitting and weighting a data set by clustering [30], bootstrapping [31], adaBoosting [32], pasting votes [33], employing different subsets of variables and different architectures are the most popular approaches used to achieve the diversity of classifiers. In this work, input features of each type are analyzed by a separate classifier for achieving the diversity.

A variety of schemes have been proposed for combining multiple classifiers into a committee [27,28,31,34–36]. The weighted averaging technique is often found to be amongst the best approaches to classifier fusion [37]. In this work, we explored three weighted averaging schemes for aggregating the classifiers into a committee:

1. Weighted averaging using one weight for one classifier. Given an image \( \xi \), the winning class \( k \) is found according to the following rule:

\[
k = \arg \max_{i=1,...,Q} \frac{\sum_{j=1}^{L} w_j \hat{y}_j(\xi)}{\sum_{j=1}^{Q} w_j \hat{y}_j(\xi)}
\]

where \( L \) stands for the number of classifiers aggregated into a committee, \( w_j \) the jth classifier weight, and \( \hat{y}_j(\xi) \) is given by Eq. (21), where the index \( j \) was added to address a feature type.

2. Weighted averaging using one weight for one classifier. Let us assume that \( p_j \) is the correct classification rate of the jth classifier. The aggregation weight \( w_j \) is then given by \( w_j = p_j^y \), with \( y \) being a parameter. The only difference between the first and the second aggregation schemes is the way of obtaining the aggregation weights.

3. Weighted averaging using weights specific for each classifier and class:

\[
k = \arg \max_{i=1,...,Q} \frac{\sum_{j=1}^{L} w_{ij} \hat{y}_j(\xi)}{\sum_{j=1}^{Q} w_{ij} \hat{y}_j(\xi)}
\]

where \( w_{ij} \) is the jth classifier weight when considering a possibility of assigning the pattern \( \xi \) to the ith class.

The parameter \( y \) and the aggregation weights used in the weighted averaging approaches (1) and (3) have been found by cross-validation using the Simplex algorithm. A cross-validation has been run for each choice of the parameter under analysis. Values of the parameter \( y \) ranged from 4 to 7.

### 4. Experimental investigations

In all the tests, we used 200 different random ways to partition the data set into Training-\( D_l \) and Test-\( D_t \) sets. The mean values and standard deviations of the test set correct classification rate presented in this paper were calculated based on those 200 trials. Out of the 785 images available, 650 images were assigned to the set \( D_t \) and 135 to the test set \( D_l \).

#### 4.1. Classification results

Table 1 summarizes the test data set correct classification rate obtained from the separate SVMs trained using features of a single type (kernel principal components) for each SVM. In the parentheses, standard deviations of the estimates are provided. The number of the kernel principal components available is equal to the number of the training samples. By applying cross-validation, the average number of the principal components providing the best performance exceeds the dimensionality of the measurement space. As it can be seen from the table, the number of the principal components providing the best performance is far below the maximum number of the components available equal to 650. However, for the relatively low dimensional measurement spaces, the number of the kernel principal components providing the best performance exceeds the dimensionality of the measurement space, except for the geometrical features. As it can be seen from Table 1, the colour distribution based features are the most discriminating ones. It is worth noting that the co-occurrence matrix based features utilized in this study provided approximately a 4% higher classification accuracy, if compared to the accuracy obtained using the ordinary co-occurrence matrix based fea-
Table 1 – The average test data set correct classification rate obtained for the different feature types when using a separate SVM for each type of features

<table>
<thead>
<tr>
<th>Feature type</th>
<th>N# initial features</th>
<th>N# kernel PC</th>
<th>N# selected kernel PC</th>
<th>Classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient</td>
<td>1000</td>
<td>650</td>
<td>210 (10.4)</td>
<td>78.86 (3.82)</td>
</tr>
<tr>
<td>Texture</td>
<td>42</td>
<td>650</td>
<td>230 (12.1)</td>
<td>82.61 (3.97)</td>
</tr>
<tr>
<td>Frequency (F1)</td>
<td>180</td>
<td>650</td>
<td>71 (11.3)</td>
<td>79.18 (3.74)</td>
</tr>
<tr>
<td>Frequency (F2)</td>
<td>40</td>
<td>650</td>
<td>180 (12.8)</td>
<td>76.35 (3.52)</td>
</tr>
<tr>
<td>Geometrical</td>
<td>18</td>
<td>650</td>
<td>18 (03.4)</td>
<td>76.01 (3.73)</td>
</tr>
<tr>
<td>Colour</td>
<td>733</td>
<td>650</td>
<td>208 (11.4)</td>
<td>91.73 (2.50)</td>
</tr>
</tbody>
</table>

Fig. 9 – Three misclassified vocal fold images from the healthy (left), the nodular (middle), and the diffuse (right) classes.

<table>
<thead>
<tr>
<th>Aggregation alternative</th>
<th>Correct classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted averaging (1)</td>
<td>94.03 (1.35)</td>
</tr>
<tr>
<td>Weighted averaging (2)</td>
<td>93.88 (1.36)</td>
</tr>
<tr>
<td>Weighted averaging (3)</td>
<td>94.11 (1.34)</td>
</tr>
</tbody>
</table>

Table 3 – Confusion matrix

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Healthy</th>
<th>Nodular</th>
<th>Diffuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>0.90 (0.060)</td>
<td>0.00 (0.000)</td>
<td>0.00 (0.000)</td>
</tr>
<tr>
<td>Nodular</td>
<td>0.09 (0.056)</td>
<td>0.96 (0.018)</td>
<td>0.08 (0.024)</td>
</tr>
<tr>
<td>Diffuse</td>
<td>0.01 (0.008)</td>
<td>0.04 (0.018)</td>
<td>0.92 (0.024)</td>
</tr>
</tbody>
</table>

The estimated probabilities for an image coming from a particular class to be assigned by the classifier to the correct and incorrect classes.

5. Conclusions

Aiming to obtain a comprehensive description of laryngeal images, multiple feature sets of different origin are extracted run 200 times using different random partitioning of the available data into the training and test sets. Table 3 presents the estimated probabilities for an image coming from a particular class to be assigned by the classifier to the correct and incorrect classes. The standard deviations of the estimates calculated from these 200 trials are given in the parentheses. Most of the classification errors occur when discriminating between images coming from the nodular and diffuse classes.

One can wonder how the misclassified images look like. Are they very similar? An example of three misclassified images coming from the three decision classes is given in Fig. 9. The image shown on the left-hand side of Fig. 9 comes from the healthy class and, in some experiments, was erroneously assigned to the nodular class. While the image shown on the left-hand side and that in the middle of Fig. 9 are not very alike, observe that images coming from these two classes can be very similar. The image shown in the middle of Fig. 9 comes from the nodular class and was erroneously assigned to the diffuse class. The representative of the diffuse class shown on the right-hand side of Fig. 9 was, in some experiments, erroneously assigned to the healthy class. A trained physician has no big difficulties to correctly categorize these three images.
and preprocessed by applying the kernel PCA. The image colour distribution, distribution of the image intensity gradient direction, parameters characterizing the geometry of edges of vocal folds, distribution of the spectrum of the Fourier transform of the colour image complex representation, and parameters calculated from multiple co-occurrence matrices are the feature types used to describe laryngeal images. A committee of support vector machines is then utilized to categorize the descriptions into the healthy, nodular, and diffuse classes. As expected, when used alone, the colour features provided the highest correct classification rate amongst all the types of features tested. Only a moderate increase of the average correct classification rate is obtained from the committees if compared to the case of using the colour features based SVM. However, the variance of the correct classification rate is reduced considerably. A correct classification rate of over 94% was obtained when classifying a set of unseen images into the three decision classes. Bearing in mind the high similarity of the decision classes, the correct classification rate obtained is rather encouraging. However, the fact that some of the misclassified images are quite easily categorized by an experienced physician proves the possibility of further improvements.

Acknowledgement

We gratefully acknowledge the support we have received from the Lithuanian State Science and Studies Foundation (EUREKA Project E!3681).

REFERENCES


