AN ALGORITHM FOR PSEUDO-3D REPRESENTATION
OF THE CONTOUR OF THE TONGUE WHILE PLAYING THE DIDGERIDOO

Claudia Lindner, Wolfgang Angerstein, Volker Aurich, Thomas Massing

ABSTRACT

In 1991, the linearity of ultrasound images for the analysis of the movement of the tongue was shown for the first time. Since then, algorithms have been developed for examining the ultrasound data with technical support. About a quarter century later the demand for a tongue contour segmentation algorithm based on current technical potential arose. Method: The ultrasound images are recorded as a plane of ultrasound cross sections with a high transverse and longitudinal resolution. Image processing steps smoothing, segmentation and pseudo-3D representation are applied to cross section stacks. The main step (segmentation) deals with edge detection, optimisation and completion of edges in grey value ultrasound images. As a result, pseudo-3D representation stands for the illustration of an image stack in the colour of time rather than for a virtual three-dimensional representation. Results: Applying the algorithm to images recorded while playing the didgeridoo shows satisfying results for arbitrary vertical as well as for horizontal veins. But, under certain conditions artefacts might occur due to an unsatisfying signal-to-noise ratio. Conclusions: Results show improvement opportunities concerning technical optimisation like advanced instrumentation or further development of algorithms. Moreover, the algorithm is applicable to other wind instruments and also for medical implications like in auscultation, swallowing or running speech.

INTRODUCTION

In 1991, Sieroni et al. investigated the effectiveness of ultrasound images for the analysis of the movement of the tongue (Sieroni, 1991). The conclusion was that ultrasound imaging of the tongue is high for the significance of examinations of tongue movements during running speech. At that time, the tongue contour was still manually determined in ultrasound images. Some years later in 1995, Fleck et al. presented a fully automated system for the automatic determination of the tongue edge in ultrasound images (Fleck, 1995). Meanwhile, the presented algorithm estimated the edge contour in a pseudo-3D representation. Since then, the algorithm has been based on the technological advances in the same field. Nowadays, ultrasound images are recorded with a high transverse and longitudinal resolution as cross section stacks (see Figure 2). In this context, pseudo-3D representation is applied for the illustration of an image stack in the colour of time rather than for a virtual three-dimensional representation. In terms of processing, ultrasound image sequences recorded while playing the didgeridoo are applied for the algorithm’s performance.

METHOD

The algorithm is based on the processing of bidimensional-bi mode images of the tongue. These cross sections are recorded with 30 to 40 millivolts. The procedure that generates a suitable pseudo-3D representation from a noisy ultrasound image sequence comprises three main steps: smoothing, segmentation and pseudo-3D representation. Firstly, the signal-to-noise ratio is improved by the application of appropriate smoothing algorithms. Afterwards, the contour of the tongue is extracted based on the processed image stack. Thus, ultrasound image data is converted from a two-dimensional grey value image to a pseudo-3D contour representation. The smoothing, as well as the segmentation, is done for each image separately, based only on the grey value of the image. Although the underlying algorithms are applied on single cross section images, the particular stage bases a complete cross section sequence. In this manner, an image sequence always gets weighted to that point. That means the grey value range is divided into three, corresponding to the colour black, and the pseudo-3D representation.

Smoothing

In this step, the ultrasound image sequences are smoothed in a two-dimensional space. This means that the number of edges in the contour is decreased and the contour is blurred. Thus, the ultrasound images appear more uniform to the human eye. This step has the advantage that the ultrasound images get smoother. Consequently, a well-known border established between tongue and oral cavity would become more visible. Usually, ultrasound data is much noisier than as a supportive intermediate step. However, the selection of the respective filter parameters influences the results, and provides scope for adaptation to several use cases. For a more detailed explanation of the filter parameters the reader is referred to Lindner (2005; Aurich, 1995; Weule, 1994).

Segmentation

The edge detection performs the segmentation of the ultrasound image sequences. Therefore, the ultrasound image sequences are converted from a two-dimensional space into a pseudo-3D contour representation. The following steps are applied for each ultrasound sequence. Firstly, each single cross section is treated line by line beginning in the top left corner. Every column is evaluated separately. Then, a sum of column means is calculated and compared to a threshold. If the sum of column means is below the threshold, the column is assumed to be part of the tongue. If the sum of column means is above the threshold, the column is assumed to be part of the oral cavity. Consequently, the contour of the tongue is extracted in the ultrasound image sequence.

Pseudo-3D representation

In the following, the ultrasound image sequences are converted from a two-dimensional grey value image to a pseudo-3D contour representation. The pseudo-3D representation is based on the non-linear Gaussian filter. It is defined as the composition of the following two filter steps.

Depending on the quality of the images and the particular parameter selection there are at least two different contour boundary edges. These two cases are not to be distinguished in order to get a full illustration of the tongue edge. Therefore, the edge detection results are processed in a second step, which is denoted as completion of the tongue edge. This algorithm is based on the non-linear Gaussian filter. It is defined as the composition of the following two filter steps.

RESULTS

The described algorithms provide a pseudosurface chain for obtaining a pseudo-3D representation of the contour of the tongue as illustrated in Figure 4. For achieving the requested results several different algorithms are applied in this chain of algorithms. The described algorithms are based on an individual algorithm chain which is designed for the given ultrasound images. This algorithm chain is based on the non-linear Gaussian filter. It is designed as the composition of the following two filter steps.

Below the results of the described algorithm chain are discussed. Thereby, the mentioned algorithms were reviewed on the basis of ultrasound images recorded while playing the didgeridoo. The corresponding results are satisfying for arbitrary vertical as well as for horizontal veins. However, under certain conditions insufficiently short cross sections in the original ultrasound images, artefacts might occur on account of an unsatisfying signal-to-noise ratio.

CONCLUSIONS

Although the algorithm show satisfying results, there is still space for improvement opportunities in terms of advanced instrumentation. Therefore, the described algorithm may be improved in two ways. First of all, the signal-to-noise ratio can be improved by the application of advanced instrumentation, which nor a development to be improved for improving the algorithmic effectiveness when processing images of poor quality. Further, at the moment the processing chain is not only not batch processing, but also the particular algorithm related processing time for each image. Consequently, the smoothing step is the time-consuming step of the processing chain. So, the described algorithm might be optimised for improving its processing performance. Therefore, the described algorithm would be improved in terms of signal-to-noise ratio. The described algorithm could be improved for improving the algorithmic effectiveness when processing images of poor quality. For that reason, the signal-to-noise ratio can be improved by the application of advanced instrumentation. The described algorithm might be optimised for improving its processing performance. Therefore, the described algorithm would be improved in terms of signal-to-noise ratio.