Visualization Techniques for Tongue Analysis in Traditional Chinese Medicine

Binh Pham[a], Yang Cai[b]
[a]Faculty of IT, Queensland University of Technology, GPO Box 2434 Brisbane Q 4001 Australia;
[b]Human-Computer Interaction Institute, Carnegie Mellon University, USA.

ABSTRACT

Visual inspection of the tongue has been an important diagnostic method of Traditional Chinese Medicine (TCM). Clinic data have shown significant connections between various viscera cancers and abnormalities in the tongue and the tongue coating. Visual inspection of the tongue is simple and inexpensive, but the current practice in TCM is mainly experience-based and the quality of the visual inspection varies between individuals. The computerized inspection method provides quantitative models to evaluate color, texture and surface features on the tongue. In this paper, we investigate visualization techniques and processes to allow interactive data analysis with the aim to merge computerized measurements with human expert's diagnostic variables based on five-scale diagnostic conditions: Healthy (H), History Cancers (HC), History of Polyps (HP), Polyps (P) and Colon Cancer (C).

Keywords: visualisation, Traditional Chinese Medicine, tongue analysis, parallel coordinates, fuzzy set.

1. INTRODUCTION

For over two thousand years, visual inspection of the tongue has been a unique and important diagnostic method of Traditional Chinese Medicine (TCM). Observing the abnormal changes in the tongue proper and the tongue coating can aid in diagnosing diseases. Clinic data have shown significant connections between various viscera cancers and abnormalities in the tongue and the tongue coating [Yao, 1996]. In 1987, the China TCM Society, China Cancer Society and TCM Diagnosis Association conducted a national project that included cases of 12,448 cancerous patients, 1,628 non-cancerous patients and 5,578 normal patients. The results statistically showed that there are significant changes of colour, coating, shape and dorsum shape of the tongues of cancerous patients versus those tongues of non-cancerous patients or normal subjects.

Visual inspection of the tongue is simple, non-invasive and inexpensive. However, the current practice in TCM is mainly experience based, and the quality of the visual inspection varies between medical professionals. Furthermore, the skills of a small number of good experts are not easily transferable to other less experienced professionals. Thus, it is beneficial to devise more objective approaches and quantitative models to evaluate color, texture and surface features on the tongue and correlate these features to patients' health conditions. Watsuji et al. [1999] developed a systematic diagnostic approach using several logic scores based on fuzzy theory. They examined two conditions: coldness-heat, and deficiency-excess. For the condition of coldness-heat, they used five grades of color (pale white, pale, light red, reddened, deep red), five grade of color of fur (white, slight yellow, yellow, dark yellow, black), and five grades of wetness-dryness. Similarly, for the condition of deficiency-excess, the shape of the tongue and the thickness of fur were used. For each condition, seven grades of diagnostic outcomes were labelled with input from medical experts (e.g. seriously cold, moderately cold, slightly cold, normal, slightly hot, moderately hot, seriously hot). Fuzzy rules were then constructed to link different tongue conditions to diagnostic labels. The authors tested these rules for 20 cases for the coldness-heat condition and found that the results coincide with those obtained by subjective inspections for 18 cases.

More recently, there have been a number of attempts to develop automated digital tongue diagnostic systems using image analysis. Jang et al. [2002] constructed a database of similar features to those used by Watsuji et al.: tongue color (light pink, rose pink, red, purple or blue), tongue thickness, tongue shape (round, triangle), fur color (white, yellow, black), fur thickness, and degree of wetness. They used a multi-layered feed forward neural network and error back propagation learning rule with R,G,B histograms as inputs to analyse tongue color which they believed is the most important factor. Cai [2002] also constructed a digital imaging system to capture tongue images and extracted various features for analysis. The features extracted are a^* and b^* (chromatic dimensions of $L^* a^* b^*$ color space), and the energy and entropy functions computed from the grey level co-occurrence matrix [Reed, 1993]. Li and Cai [2003] later devised algorithms to extract two more features which indicate the roughness and the amount of cracks on the tongue. They extracted features for 34 tongues of 5 categories: (Healthy (HE), History Cancers (HC), History of Polyps (HP), Polyps (PO), Colon Cancer (CC)). We will use these results as test data for our visualization techniques to be discussed in later sections.

The computerized diagnostic approach, which provides quantitative models to evaluate different features of the tongues and deduce the patients' illness, is still at an early stage of development. Based on these features, commonly used clustering and classification methods such as K-means, C4.5, SVM and neural networks may be used to group tongue images into diagnostic classes. These classes can then be evaluated for their accuracy, ie. how well they match with experts' classification. However, a practical problem is that TCM doctors often do not easily accept numerical measurements and classification results that are not readily aligned to their intuitive representation. Thus, it is essential to provide visualization techniques that allow them to interactively explore and gain insights into the special characteristics of the data, and their correlations and impacts on the diagnosis. The aim of this paper is to investigate such techniques and the exploration process in order to merge computerized measurements with human expert's diagnostic variables so that we can have a more holistic picture of the diagnosis/symptom. Section 2 gives a brief overview of the traditional methods for tongue analysis in TCM and image features for automated analysis. Section 3 presents different visualisation techniques for preliminary analysis and rule discovery. We also discuss how fuzzy set theory can be integrated into these techniques to improve visual perception and analysis. Summary discussions and future work are given in the final section.

2. IMAGE FEATURES FOR TONGUE ANALYSIS

In Chinese medicine, inspection of the vital signs: eyes, tongue, facial expressions and general appearance is one of the main techniques of diagnosis. The tongue is specially believed to contain special physiological information about the human body. To determine a patient's condition and diseases, Chinese (or more broadly, Oriental) doctors have routinely used information on the color, degree of wetness and coarseness, and shape of the tongue. There have also been various clinical studies to establish the correlation between the condition of tongues and different stages of major illnesses, in particular, different types of cancer. The goal of providing an automated system for tongue analysis is not to replace conventional diagnostic methods, but to assist doctors with their decision-making by giving an early alert signal that can lead to further diagnosis by other methods such as MRI, CT, X-Ray and colonoscopy. However, there are a number of important issues that need to be addressed. The first issue is how to capture the tongue images and effectively segment the tongues and calibrate their colors. This problem has been addressed by a few research groups (e.g. Jang et al. 2002, Cai 2002). The second issue is how to select image features that are relevant and how to represent these features? For example, although the color of the tongue is generally accepted as the most important factor in the diagnosis, it is not clear which color space is more appropriate, and whether color histograms or average colors should be used? The third issue is whether to use machine learning techniques solely to train and classify the data based on these features or to allow doctors to gain insights into data characteristics by interactive exploration via appropriate visualization techniques? The fourth issue is how to deal with the fuzzy characteristics of the data. The judgements on both the tongue condition and disease diagnosis are subjective and imprecise, while feature measurements are not. Should these measurements be fuzzified appropriately, and would such a move improve the accuracy of the rules obtained? In this paper, we focus on the visualization aspect regarding the third and fourth issues. Our intention is not to provide static visualization results, but a dynamic visualization process to facilitate interactive discovery of salient characteristics of data, their correlation, and rules that underlie the characteristics of different clusters.

To demonstrate this visualization process, we use a case study based on a set of test data obtained by the tongue image capture system constructed by [Cai, 2002]. In this system, the colors of these images are calibrated using a Munsell ColorChecker [McCamy,1976]. The tongue area is cropped by using an active contour model to match a deformable model to an image region by means of energy minimization [Akgul, 1999]. The shadows and highlights were then removed by firstly converting color intensity values from RGB to HSV, then using appropriate thresholds as cutoff criteria. Cai captured and processed a set of 19 tongue images consisting of 8 healthy patients and 11 patients with

	New Measures		*Old Measures*			
	Db	**CI**	**a**	**b**	**energy**	**entropy**
abnormal\04.16_1415_1.HC.tif	1.51	0.45	15.47	3.76	0.32	4.38
abnormal\04.16_1500_3-P.tif	1.65	0.58	13.59	6.14	0.37	4.48
abnormal\04.23_1530_2-C.tif	1.71	0.58	7.17	12.73	0.27	5.00
abnormal\04.30_0900_1-HC.tif	1.55	1.25	13.29	6.36	0.32	3.68
abnormal\04.30_1300_1-P.tif	1.72	1.95	15.43	4.18	0.28	4.68
abnormal\05.06_1255_2-P.tif	1.67	0.23	12.25	4.95	0.32	4.72
abnormal\05.06_1335_2-P.tif	1.73	1.92	7.75	4.73	0.33	4.59
abnormal\05.14_0930_2-HP.tif	1.66	0.97	16.21	7.75	0.27	4.23
abnormal\05.14_1400_1-HP.tif	1.70	0.68	10.08	4.96	0.32	4.80
abnormal\05.28_0840_2-P-HC.tif	1.61	3.83	11.84	5.92	0.29	5.06
abnormal\05.28_0900_1-HC.tif	1.63	1.15	12.57	6.24	0.27	4.47
abnormal\05.28_0940_1-HP.tif	1.61	0.64	11.11	6.16	0.31	4.70
abnormal\06.04_1300_1-P.tif	1.64	0.55	9.98	5.35	0.33	4.66
abnormal\06.04_1420_2-HP.tif	1.68	0.62	18.76	9.97	0.25	4.76
abnormal\06.18_0930_3-HC.tif	1.68	0.84	9.33	5.76	0.36	4.74
abnormal\06.18_1020_2-HP-OC.tif	1.64	1.23	13.07	5.84	0.19	5.12
abnormal\06.25_0835_1-HP.tif	1.67	1.77	15.12	5.01	0.23	5.04
abnormal\11.27_1030-C.tif	1.63	0.10	11.55	8.77	0.33	4.91
healthy\04.16_0845_1.tif	1.61	0.61	24.56	6.08	0.16	4.64
healthy\04.23_0930_1.tif	1.62	1.17	11.53	11.34	0.26	4.46
healthy\04.23_0930_3-STRANGE.tif	1.59	3.32	13.62	-6.54	0.33	4.22
healthy\04.23_1300_1.tif	1.67	1.42	11.13	-1.34	0.28	4.37
healthy\04.23_1530_1.tif	1.70	1.57	6.05	11.83	0.25	4.29
healthy\04.30_0925_2.tif	1.61	0.84	16.47	6.44	0.30	4.77
healthy\05.06_0845_3.tif	1.66	1.24	11.54	5.61	0.32	5.11
healthy\05.06_0945_2.tif	1.71	0.79	8.33	5.55	0.30	4.67
healthy\05.06_1300_1.tif	1.70	0.53	14.87	5.18	0.32	4.59
healthy\05.06_1430_2.tif	1.62	0.97	17.83	5.83	0.27	4.34
healthy\05.14_1100_2.tif	1.65	0.54	10.29	7.13	0.31	4.47
healthy\05.14_1300_1.tif	1.70	0.14	11.44	5.34	0.30	5.05
healthy\05.14_1430_1.tif	1.55	0.08	15.58	6.93	0.24	4.95
healthy\06.04_0825_1.tif	1.73	0.26	12.40	9.10	0.27	4.45
healthy\06.25_0920_2.tif	1.66	2.10	12.26	3.64	0.28	5.22
healthy\06.25_1020_2.tif	1.60	2.01	16.06	6.94	0.28	4.85

Table 1. Extracted feature values for 34 tongue images [Li & Yang, 2003].

stomach cancer. Four features were calculated: a^* and b^* (chromatic dimensions of $L^* a^* b^*$ color space), and two texture features: the energy and entropy functions computed from the grey level co-occurrence matrix [Reed, 1993].

Yang reported that a 3D cluster plot of the data for the dimensions (energy, a^* , b^*) indicates two distinct clusters of healthy and cancerous patients.

More recently, Li & Cai [2003] constructed algorithms to calculate two more features which indicate the roughness and the amount of cracks on the tongues. The roughness feature is measured by the Differential-Box-Counting Dimension Db which is an estimator of fractal dimension [Kraft, 1995]. The crack index CI is obtained by firstly identifying the areas containing cracks, then computing the percentage of such areas versus the total area of the tongue. Detailed algorithms for calculating these six features were given in [Cai, 2002; Li & Cai, 2003], hence are not repeated here. The first two features and the last two features match more intuitively with the physical meaning of the tongue condition. The texture features although provide some discriminatory power in terms of higher order statistics, do not correspond directly with any physical property. Li & Cai extracted the values for these six features for a set of 34 tongues which belong to people of 5 different diagnostic categories : Healthy (H), History of Cancers (HC), History of Polyps (HP), Polyps (P), Colon Cancer (C)) [Li & Cai, 2003]. Table 1 lists these feature values together with their corresponding diagnostic categories. We use this data set to demonstrate the visualization techniques and process in the next section.

3. VISUALISATION TECHNIQUES AND PROCESS

3.1. Preliminary Analysis Using 2D and 3D Cluster Plots

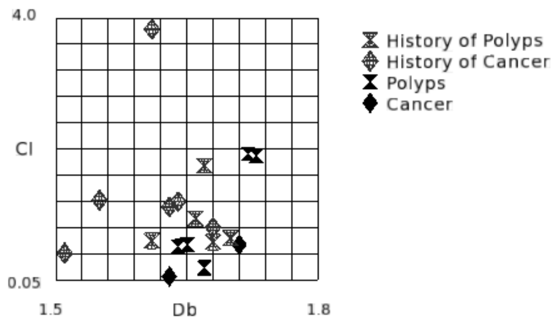
The first task of data exploration is to investigate if there are some obvious correlations between different features. This can be achieved by 2D and 3D cluster plots. We have found that there are certain definite patterns for some categories as summarized in Table 2. Fig. 1 shows a 2D cluster plot of Db-CI which shows some clear patterns for HP and HC cases, where HP corresponds to low Db and high CI while PC corresponds to low Db and low CI.

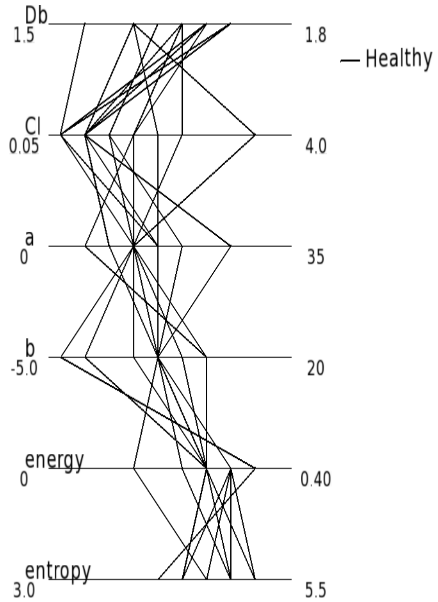
Diagnostic category	**Db**	**CI**	**a^***	**b^***	**Energy**
HP	Low	High	Narrow mid range	Mid range	Low
P	High			Mid range	High
HC	Low	Low	Narrow mid range	Narrow midrange	
C			Low	Very high	
H	either low Db	or high CI			

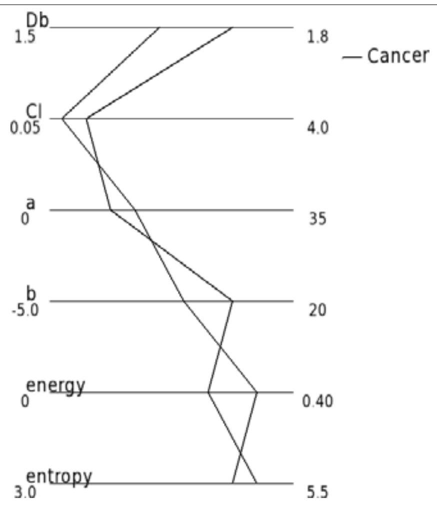
Table 2. Some observed patterns obtained from 2D and 3D cluster plots

As a^* and b* are two chromatic dimensions in the L^* a^* b^* color space, observed definite ranges of values for these features for certain diagnostic conditions indicate that there is a strong correlation between the tongue color and these diagnostic conditions. We also observe two outliers with very high crack index, one with Healthy condition and one with Polyps and History of Cancer. This suggests that these are special cases where these tongues normally have lots of cracks, and the amount of cracks therefore does not reflect on the disease condition.

For 3D cluster plots, we use different colors and glyphs for each category and allow users to rotate them around each axis to facilitate viewing, Rotating the plots increases the sense of depth and the perception of clusters and correlations. We also observe that the high precision of the raw data might have adverse effects on the cluster viewing (in other words, the precision is higher than what is meaningful and required). We therefore try different quantization scales and finally choose a 10-interval scale which is the coarsest scale that can still distinguishes each data point. The plots for each triplet of features (with a^* and b^* being always kept together) re-confirm the observations summarized in Table 2.







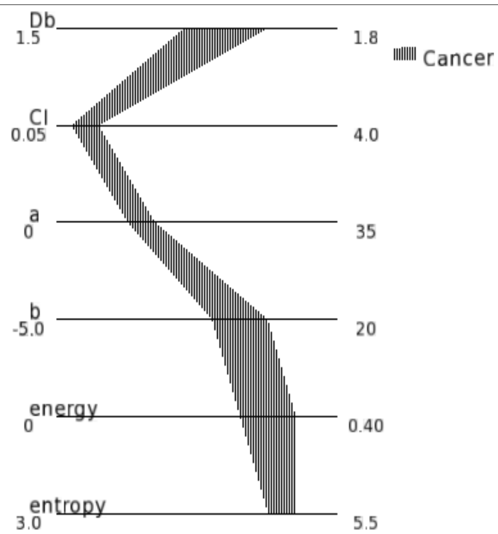


Fig. 1. 2D cluster plot of all cases. Fig.2. Parallel coordinate plot of Healthy cases.
Fig. 3. Parallel coordinate plot of Cancer cases. Fig.4. Parallel coordinate plot of Cancer cases as shaded areas.

Another type of information we wish to extract from these 3D cluster plots is to determine which triplets of features give better discriminating power and use these as a starting point to gradually explore rules for clustering in higher dimensions using parallel coordinates. Fig. 5 displays a 3D cluster plot for (Db, a^*, b^*) for all data which shows better separation between cases. Fig. 6 displays only disease cases (HP, P, HC, C). However, we observe that although there appear to be some groupings, these groupings not separable and the rules underlying these groupings are not simple. We then investigate this problem further using another visualization method based on parallel coordinates, which allows the simultaneous viewing of multi dimensions (more than 3 dimensions).

Fig. 5. 3D cluster plot of all cases.

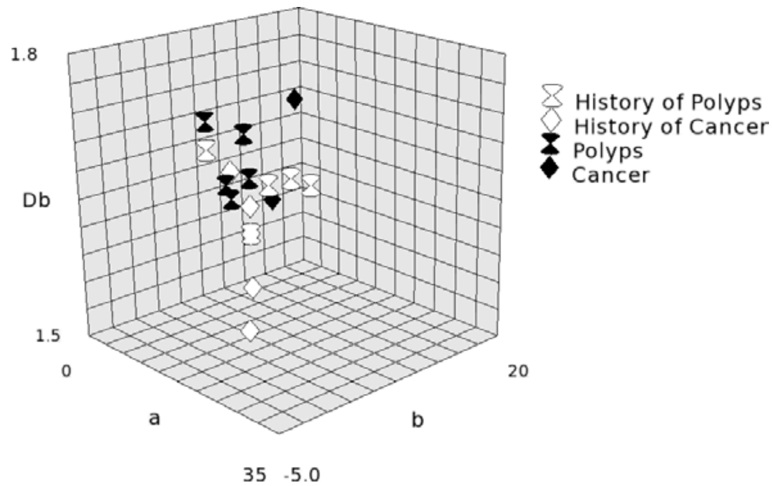


Fig. 6. 3D cluster plots of HP, P, HC, C cases.

3.2. Discovery of Rules Using Parallel Coordinates Plots

High dimensional data is often transformed or projected to 2D or 3D representations for visualization. However, this practice usually causes a loss of information. Parallel coordinates allow n-dimensional data to be displayed in 2D

[Inselberg & Dimsdale, 1994]. In this method, n Cartesian coordinates are mapped into n parallel coordinates, and an n-dimensional point becomes a series of (n-1) lines connecting the values on n parallel axes. Berthold and Holve [2000] used this technique to visualize fuzzy rules underlying 3 classes of Iris which resulted from a training set of 75 data samples with 4 features (petal length, petal width, sepal length, sepal width). Pham & Brown [2003] extended this technique to 3D to provide better visualization of the membership function of the fuzzy sets and insight into the strength of the clustering. We now show how to apply these techniques to analyse the tongue data.

Since the set of data available at this stage only consists of 34 tongues (6HP, 5P, 5HC, 2C, 16H), there is not sufficient data for training and evaluation using commonly used classification techniques such as neural networks and C4.5. Instead, we attempt to cluster this data set through their display in 6 parallel coordinates: Db, CI, a, b, energy and entropy. Fig. 2 displays the points which corresponds to all Healthy cases, where each point is represented by 5 linear segments connecting 6 values. Fig. 3 displays the points which correspond to the cases of Cancer. The order of these coordinates do not change the results, although it might affect the perceptability. For example, a large number of intersections might cause confusion and make it difficult to discern the clusters. Thus, we provide tools to swap coordinates in order to choose the order with best perceptability. We observe that the Cancer cases form a narrow band for all 6 coordinates as seen in Fig. 4.

To determine if there are any observable patterns for different cases, we also display all data for each diagnostic category superimposed on one plot. Figure 7 shows the results for 4 categories: HP, P, HC and C. The shaded area representing each category is obtained by plotting the extent covered by the extreme values for each coordinate. It can be readily seen that these areas although are overlapped, are distinguishable from each other.

We also observe that for Healthy cases, the variability in feature values is much greater than in the disease cases (Fig. 2). However, this fact can only be confirmed when more data on disease cases is available. If a much larger set of data is available, it would also be possible to provide a more sophisticated visualization by integrating fuzzy sets and using 3D parallel coordinates. We will discuss how this can be achieved in the next subsection.

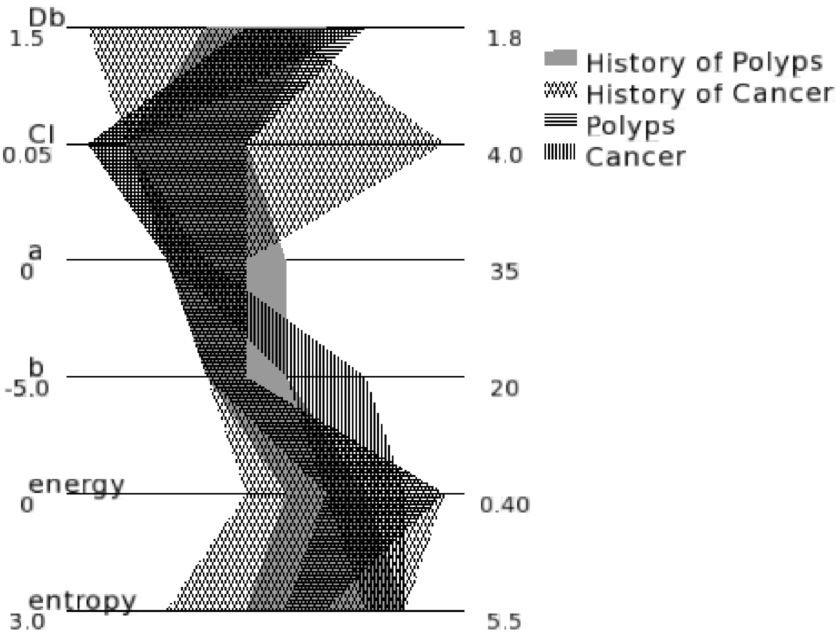


Fig. 7. Parallel coordinate plot of 4 diagnostic categories: HP, P, HC, C.

3.3. Integration of Fuzzy Sets to Visualization

Fuzzy logic has been used extensively and successfully in many areas, especially in social sciences and engineering. While mathematical models are based on algebraic operations (e.g. equations, integrals), logic models rely on logic-type connectives (and, or, if-then), often with linguistic parameters, which give rise to rule-based and knowledge-based systems. Fuzzy logic models can combine both of these types of modelling via the fuzzification of algebraic and logical operations. There are three common classes of fuzzy logic models: information processing model which describes probabilistic relationship between sets of inputs and outputs; control models which control the operations of systems governed by many fuzzy parameters; and decision models which model human behaviour incorporating subjective knowledge and needs, using decision variables. For some applications, fuzzy systems often perform better than traditional systems because of their capability to deal with non-linearity and uncertainty. One reason is that while traditional systems make precise decisions at every stage, fuzzy systems retain the information about uncertainty as long as possible and only draw a crisp decision at the last stage. Another advantage is that linguistic rules, when used in fuzzy systems, would not only make tools more intuitive, but also provide better understanding and appreciation of the outcomes.

As more tongue data is available, it would be more appropriate to treat the extent of each feature value for each diagnostic category as a fuzzy set. The membership function for this fuzzy set can be computed from the frequency of each value. This membership value gives an indication of the confidence level that each value belongs to this set. Hence, the level of overall confidence that a given case belongs to a particular diagnostic category is the minimum of the membership values for all features. It is worthwhile to note that as each pair (a^*, b^*) corresponds to a specific color on the color space, it is more appropriate to map pairs of (a^*, b^*) to fuzzy description of different colors, rather than to fuzzy quantitative data on two separate dimensions. On the other hand, four other features can be mapped directly to intervals (a simple form of fuzzy representation) or fuzzy sets with descriptive representations and hedges.

Since we have not yet obtained a large enough set of tongue data available, we demonstrate this techniques using the Iris data example. Fig. 8 shows a 3D parallel coordinates display for this data set. The advantages of integrating fuzzy sets are two-fold. Firstly, it provides an intuitive match with the way doctors fuzzily assess the condition of the tongues. Secondly, it is possible to select the tightness of clusters through the use of an alpha cut plane to discard those cases whose feature values have too low membership values (ie. the level of confidence that a particular case belong to a specific class is low).

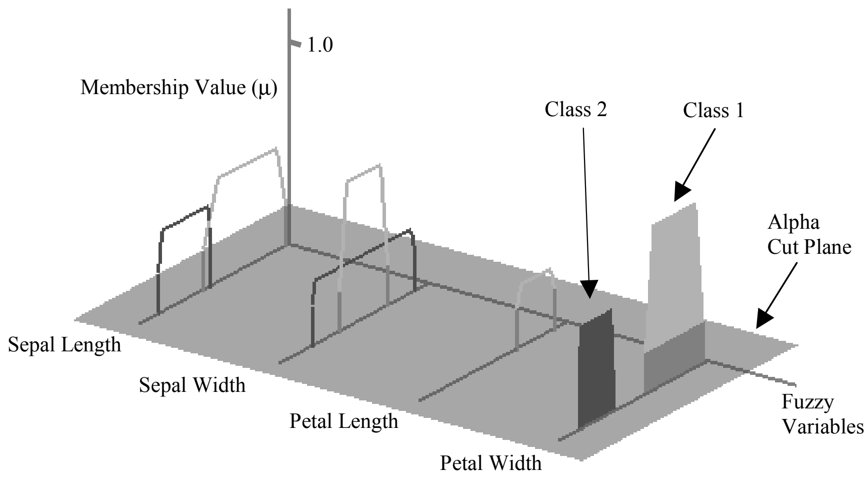


Fig. 8. A 3D parallel plot of Iris data example [Pham & Brown, 2003].

Another improvement can be made by integrating fuzzy sets is to ask doctors to provide the assessment of diagnostic categories with fuzzy grading. For example, instead of Healthy, 3 grades are introduced: Very Healthy, Moderately Healthy, and Slightly Healthy. Similarly, disease conditions can be expressed in 3 grades: Very Serious, Moderately Serious, Slightly Serious. Such fuzzy assessment would match more faithfully with real diagnosis practice. By linking fuzzy values for the color and texture features of the tongue with fuzzy diagnostic categories, it is envisaged that more accurate classification of cases would result. However, in order to achieve this, we will need to collect more cases and more detailed diagnosis from doctors for each case.

4. CONCLUSION AND FUTURE WORK

We have presented a method for interactive exploration of patterns for the automated analysis of tongues of Healthy condition and of different types of disease state. The analysis is based on the color and texture features extracted from a set of 34 tongue images. We have found that the clustering patterns are not simple and are not easily perceived on cluster plots, even when color glyphs and rotation capability were provided. Cluster plots are also of limited use when dealing with more than three features. On the other hand, parallel coordinate plots offer a better alternative to allow simultaneous viewing of all features, and a clearer grouping of cases with the same diagnostic category.

We have also discussed how fuzzy sets can be used to represent the fuzzy characteristics nature of both features and diagnostic conditions in order to model more faithfully to the real diagnosis process. Due to the difficulty of accessing a large number of disease cases, the evaluation of the integration of fuzzy sets to parallel coordinate visualization is still pending. Once a larger set of data is available, we also plan to apply objective classification methods C4.5 and SVM to this data set. A comparative analysis of the classification results obtained by these methods, by experts, and by visualization methods will then be carried out.

Acknowledgement. The authors wish to thank Dr. Robert Scheon from University of Pittsburgh Medical Center for his assistance with collecting the clinical data and Alexander Streit for his programming assistance.

REFERENCES

1. Y.S. Akgul et at, Automatic Extraction and Tracking of the Tongue Contours, *IEEE Trans. on Medical Imaging*. Vol.18, No.10, October, 1999.
2. Y. Cai, A novel imaging system for tongue inspection, *IEEE Instrumentation and Measurement Technology Conference*, AK, USA, May 2002.
3. L. Hall and M. Berthold, Fuzzy Parallel Coordinates, in Fuzzy Information Processing Society, *NAFIPS. 19th International Conference of the North American*, Atlanta, GA, USA, (2000) 74 -78.
4. A. Inselberg and B. Dimsdale, Multidimensional lines i: representation. *SIAM J. Applied Math,* 54(2), 559-577, 1994.
5. J.H. Jang, J.E. Kim, K.M. Park, S.O. Park, Y.S. Chang, B.Y. Kim. "Development of the Digital Tongue Inspection System with Image Analysis". *Proceedings of the Second Joint EMBS/BMES Conference*. Houston, TX, USA. October 23-26, 2002.
6. R. Kraft. "Fractals and Dimensions". Munich University of Technology. 1995.
7. G. Li and Y. Cai, Texture analysis for tongue analysis, *Technical Report BV-2003-2*, School of Computer Science, Carnegie Mellon University, May 2003.
8. C.S. McCamy et al, A Color Rendition Chart, Journal of Applied Photographic Engineering, Summer Issue 1976, Vol.2, No.3. page 95-99
9. B. Pham and R. Brown, Multi-agent approach for visualisation of fuzzy systems*, ICCS'03 International Conference on Computational Science*, Melbourne, June 2003, 995-1004.

10. T.R. Reed, and J.M.H. DuBuff (1993). *A review of recent texture segmentation and feature extraction techniques.* CVGIP: Image Understanding, v. 57, n. 3, pp 359-372.
11. T. Watsuji, S. Arita, S. Shinohara, T. Kitade. "Medical Application of Fuzzy Theory to the Diagnostic System of Tongue Inspection in Traditional Chinese Medicine". *IEEE International Fuzzy Systems Conference Proceedings*, pp. 145-148, 1999.