School of Computing Sciences, Postgraduate Research Day 2013: Book of Abstracts

Table of Contents

Papers

Heuristic Ensemble of Filters for Reliable Feature Selection

Ghadah Aldehim, Wenjia Wang and Beatriz de la Iglesia

School of Computing Sciences
{G.aldehim,Wenjia.Wang,B.Iglesia}@uea.ac.uk

Feature selection has become ever more important in data mining in recent years due to the rapid increasing in the dimensionality of data. Filters are more preferable in many applications as they are much faster than wrapper-based approaches, but the reliability and consistency of each individual filter vary considerably from data to data and yet no rule exists to indicate which should be used for a particular given dataset.

In this research, we have proposed a heuristic ensemble method that combines multiple filters with heuristic rules to improve the reliability and accuracy of feature selection. It consists of two types of filters: subset filters (SF) and ranking filters (RF) as each type gives different and complement information on the relevance of features. A subset filter only selects a subset of features in a dataset without ranking them, whilst a ranking filter just produces a ranking of all the features without selection. Our ensemble method combines their outputs by two aggregation methods: the frequency of the features selected by SFs and RFs, and then the weights of features determined by RFs. Firstly, the features selected by SFs and RFs are ranked based on their frequency. However, as the probability of any two features having the same frequency is high, we introduce an average ranking weight derived from RFs to resolve the ranking position of the features with the tied frequencies. After that we apply heuristic consensus rule to produce the final output of the ensemble.

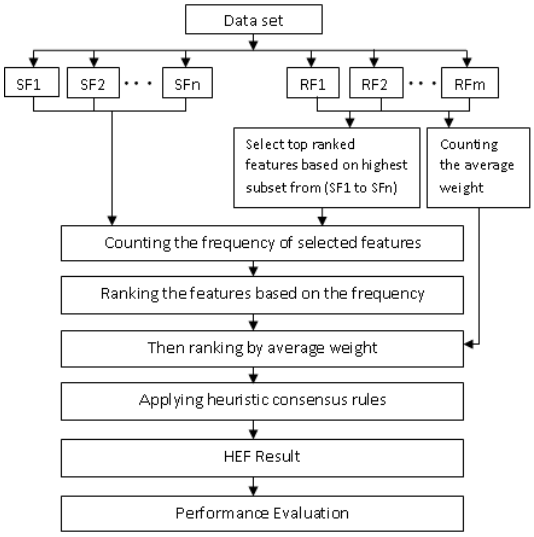


Figure 1: Framework of heuristic ensemble of filters (HEF) for feature selection.

The proposed heuristic ensemble of filters (HEF) framework is implemented in Java, primarily based on the modules available in an open source data mining package WEKA. The features selected by the HEF will be used to train classifier for a given dataset. The trained classifiers are tested on the test data and then compared with the classifiers that are trained with the features selected by the four individual filters, and also all the features (without selection). Eleven benchmark datasets from different domains are used in our experiments. For each dataset we use three kinds of model: Nave Bayesian Classifier, K-Nearest Neighbour and Support Vector Machine to test the model dependency of selected features. The 10-fold cross validation strategy is used for each classifier. Each experiment is then repeated ten times with different shuffling random seeds in order to assess the consistency and reliability of the results. In total, 23,100 models were built in the experiments. The experimental results demonstrate that our ensemble algorithm is more reliable and effective than individual filters as the features selected by the ensemble achieve better accuracy for typical classifiers on various datasets.

The plan for the future work is to implement certain methods for measuring the stability of our approach. Then attempt to improve the experiments, through investigate the ensemble results by using: different types of filters and different numbers of filters. Also, we plan to add a wrapper ensemble to improve the overall result and to assist in selecting the most important features.

Clustering Ensemble Method based on Neighborhood Structure

Tahani Alqurashi and Wenjia Wang

Machine Learning and Statistics Laboratory
{T.Alqurashi, Wenjia.Wang}@uea.ac.uk

Clustering is the process of discovering natural groups of similar elements in datasets. Although, cluster analysis has success in several applications, a significant challenges need to be met. Due to the inherent difficulties in clustering algorithms, most of them produce different result for the same dataset. Ensemble clustering can overcome all of these issues as well as its ability to improve the clustering performance. It combines multiple partitions (ensemble members) from the same dataset in order to produce a single improved final clustering without accessing the features or the algorithms that determined these partitions. Therefore, the main problem here is how to combine a multiple partitions. Many clustering ensemble methods have been proposed and the most popular one based on the pairwise similarity. This method is overlooked the issue of handling the similarity between uncertain pair of objects, which they have been assigned to the same cluster by approximately half of the members in the ensemble and assigned to different clusters by the other half, and between certain dissimilar pair of objects, which have not been assigned in the same cluster by all members and its similarity score equal to 0. In fact, when the uncertain and certain dissimilar pairs are large, they can mislead the ensemble clustering to produce an inappropriate partition of the data. Our method is deal with these issues by consider the neighborhood structure as a measure of similarity. The neighborhood is a region in the object space covering the object in question. The framework of clustering ensemble based on neighborhood structure (CENS) generates heterogeneous ensemble members using different clustering algorithms. In the integration phase the undirected graph is constructed (V objects, W edge weight which is similarity matrix). W calculates the co-occurrence of each pair over all ensemble members. Then based on W we identified the three types of pair, certain similar pair, which has been assigned in the same cluster by all partitions and its value equal to 1, uncertain similar pair and certain dissimilar pair. Based on these types, we defined two rules to deal with them:

- Rule-1 Keep the similarity score for the certain similar pair.
- Rule-2 Recalculate the similarity score for uncertain similar and certain dissimilar pair by considering their average neighborhood over all members.

Then we applied the spectral clustering algorithm to obtain the final partition. Experiments conducted on three artificial data and three real data (Glass, BreastCancer, Wine data) and firstly we generated for each dataset six ensemble members using different algorithms with the pre-specified k cluster. Then the final clustering result obtained using CENS and we evaluated the quality using Normalised Mutual Information (NMI). Moreover, We compared it with Cluster-based Similarity Partitioning Algorithm (CSPA) the HyperGraph-Partition Algorithm (HGPA), and the Meta-Clustering Algorithm (MCLA) proposed by [1]. The results show that HGPA has the worst performance in all datasets except D2C2. In the presence of an unbalanced dataset, such as D2C2 or D2C3, CENS performs better than other methods. We also obtained the average performance of the generated ensemble members, and we observed that CENS leads to consistent performance over the ensemble members in all cases. HGPA appears to be the least stable method.

In conclusion, clustering ensemble has emerged as a prominent method for improving the robustness, stability, and accuracy of the clustering result. In this work, we developed a clustering ensemble method based on neighborhood structure in order to overcome the problems associated with uncertain pairs of object and certain dissimilar pairs of objects. From our experiment, we conclude that CENS performs better than the CSPA, HGPA, and MCLA methods as well as the generated ensemble members. We have limited the ensemble size to six, but in future research, we will investigate the impact of different ensemble sizes on the overall performance.

References

[1] A. Strehl and J. Ghosh, "Cluster ensembles-a knowledge reuse framework for combining multiple partitions," *The Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2003.

Generating a Domain Specific- Context Inspection Method through an Adaptive Framework

Roobaea Alrobaea and Pam Mayhew

School of Computing Science
{R.Alrobaea,P.Mayhew}@uea.ac.uk

The growth of the Internet and related technologies has enabled the development of a new breed of dynamic websites and applications that are growing rapidly in use and that have had a great impact on many businesses. These websites need to be continuously evaluated and monitored to measure their efficiency and effectiveness, to assess user satisfaction, and ultimately to improve their quality. Nearly all the studies have used Heuristic Evaluation (HE) and User Testing (UT) methodologies, which have become the accepted methods for the usability evaluation of User Interface Design (UID); however, the former is general, and unlikely to encompass all usability attributes for all website domains. The latter is expensive, time consuming and misses consistency problems. To address this need, new evaluation method is developed using traditional evaluations (HE and UT) in novel ways.

The lack of a methodological framework that can be used to generate a domain-specific evaluation method, which can then be used to assess and improve the usability assessment process for a product in any chosen domain, represents a missing area in usability testing. This research proposes an adapting framework (see Figure 1) and evaluates it by generating an evaluation method for assessing and improving the usability of a product, called Domain Specific Inspection (DSI), and then analysing it empirically by applying it on the educational domain and social network domain. Also, it presents an adaptive Domain Specific Inspection (DSI) check-list from DSI method as a tool that can be used by evaluators, designers, developers, instructors, and website owners to design an interactive interface. In addition, It identifies the usability problem areas in both domains, and on that basis the DSI method will be classified according to these areas. Furthermore, the role of number of evaluators with their types and users in such testing will be examined. Moreover, the correlation between set of factors and methods used will be explored.

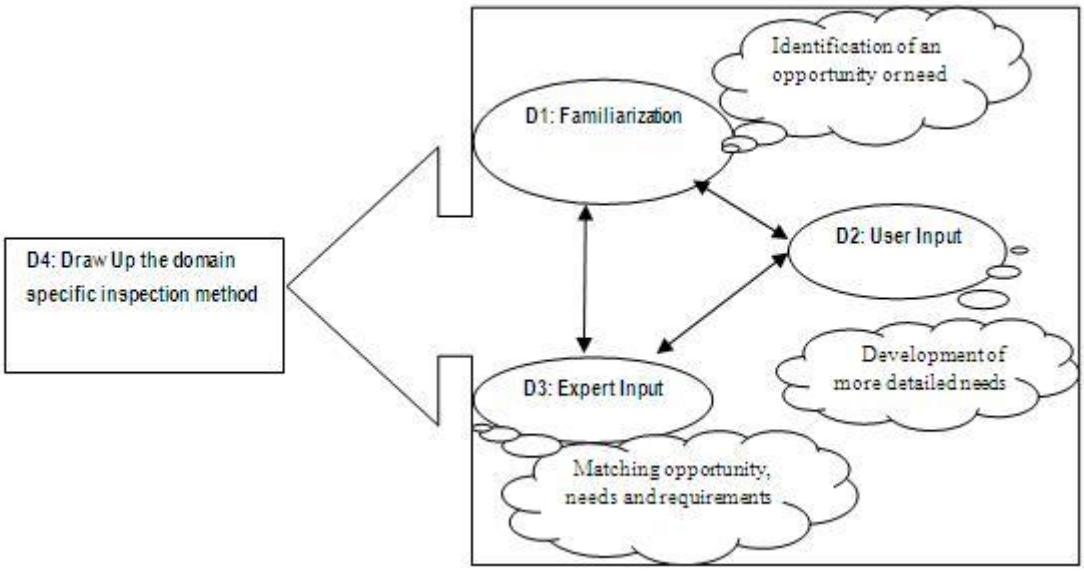


Figure 1: The adaptive framework for generating DSI method.

Profiling the Effect of Secondary Structure on sRNA Expression in YRNAs

Matthew Beckers, Carly Turnbull, Tamas Dalmay and Vincent Moulton

School of Computing Sciences, School of Biological Sciences
{m.beckers@uea.ac.uk,t.dalmay}@uea.ac.uk

YRNAs are non-coding RNAs of around 84 nucleotides in length that fold into a tight hairpin-like secondary structure. Once folded, they are cleaved by endonucleases to produce small RNAs (sRNAs) that are around 31 nucleotides in length [1]. These YRNA-derived sRNAs (YsRNAs) are only expressed when the cell is stressed and in the presense of a number of proteins binding to the YRNA sequence such as Ro60 and La (figure 1a). Furthermore, YsRNAs are not processed by dicer-like proteins or involved in the RNA interference pathway [2]. This is unlike the function of other sRNAs such as miRNAs, whose precursor RNA folds in a similar fashion to YRNAs. Instead, little is known about the function of YsRNAs or how they are cleaved from their YRNA precursor.

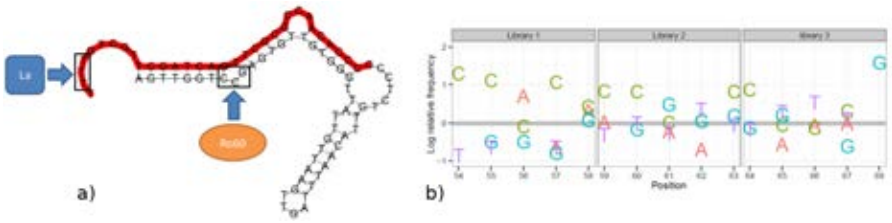


Figure 1: a) The hy5 YRNA sequence folded into a tight hairpin secondary structure (predicted by RNAfold). b) Relative log-odds of nucleotides over all mutation positions when compared to the background frequency.

High-throughput sequencing (HTS) of RNAs provides a vital resource for the bioinformatic analysis of sRNA expression levels within an organism's transcriptome [3]. In this study, we have applied HTS techniques to the hy5 YRNA in order to further understand the biogenesis of YsRNAs. This unique experiment involved mutating the YRNA at three different five-nucleotide regions of the YRNA in the wet lab. This produced pools of YRNAs containing all possible mutations at each position, which were transfected in to mouse cells. The cells were placed under stress to initiate YsRNA production and the resulting sRNA transcriptome was extracted and sequenced.

The sequenced sRNA dataset was used in a bioinformatic analysis to assess the effect of mutating the YRNA sequence, and changing its secondary structure, on the expression of sRNAs. The analysis involved mapping the resulting sRNAs to all possible hY5 reference sequences. sRNA sequences relating to the 3' YsRNA were aggregated by the mutation that they mapped to. Sequence motif and secondary structure motif analyses were performed in order to search for a pattern of sRNA expression when compared across mutants.

The results showed that certain mutations expressed sRNAs at a significant level when compared to a random sampling of all possible mutations. Sequence motif analysis of these mutations revealed that certain nucleotides were conserved in order to generate the sRNAs (figure 1b). Furthermore, analysis of the secondary structures showed that the signifcantly expressed mutations were cleaved at a position dictated by the secondary structure. This suggests that the cleavage mechanism of the endonucleases involved in processing the YRNA to YsRNAs function by targeting structural motifs of the precursor YRNA.

References

[1] A. E. Hall, C. Turnbull, and T. Dalmay, "Y rnas: recent developments," *BioMolecular Concepts*, vol. 4, no. 2, pp. 103–110, 2013.

[2] F. E. Nicolas, A. E. Hall, T. Csorba, C. Turnbull, and T. Dalmay, "Biogenesis of y RNA-derived small RNAs is independent of the microRNA pathway," *FEBS Letters*, vol. 586, no. 8, pp. 1226–1230, Apr. 2012.

[3] N. Fahlgren, C. M. Sullivan, K. D. Kasschau, E. J. Chapman, J. S. Cumbie, T. A. Montgomery, S. D. Gilbert, M. Dasenko, T. W. Backman, and S. A. Givan, "Computational and analytical framework for small RNA profiling by high-throughput sequencing," *Rna*, vol. 15, no. 5, p. 9921002, 2009.

Benchmarking Software Reverse Engineering Approaches with RED-BM

David Cutting and Joost Noppen

Software Engineering
david.cutting@uea.ac.uk, j.noppen@uea.ac.uk

Organisations often have a need to understand existing software to order to maintain it or make changes such as adding new features. This existing software frequently has only partial or missing documentation as even if originally developed from clear designs, organic updates made during the life of the system can mean these designs have little relation to the current state. An approach to regaining comprehensible designs of software is reverse engineering; the process of analysing the current source code and reconstituting understandable design information such as UML class diagrams.

There are a wide number of techniques and tools to perform this reverse engineering [1] but there is no standard with which to compare them or to validate new techniques [2]. To address this we have created the Reverse Engineering to Design Benchmark (RED-BM) [3], a standard which allows the systematic comparison of reverse engineering techniques.

The primary component of the benchmark is a set of 16 standard software artefacts to be used as targets within a reverse engineering process. These artefacts offer a varied range of size, complexity, and architectural styles. Size varies from the smallest artefact (ASCII Art Example A) containing 10 classes with 124 lines of code to the largest (Libre Office) with 414 classes and 39,896 lines of code.

In addition to the artefacts themselves we provide:

- Reference UML diagrams and documentation where available
- A number of initial metrics and measures to gauge performance of reverse engineering
- A ”gold standard” of expected output for each performance measure and artefact, allowing performance to be ranked
- A weighted compound measure which summarises performance measures into a single overall value for ranking
- Output from common reverse engineering approaches for each of the artefacts for comparison and use in validation in new techniques

For future use and adaption the benchmark is designed to be extensible through altering of weights as well as definition of new individual and compound measures. This allows any user of the benchmark to change focus to a new area of reverse engineering such as design pattern detection.

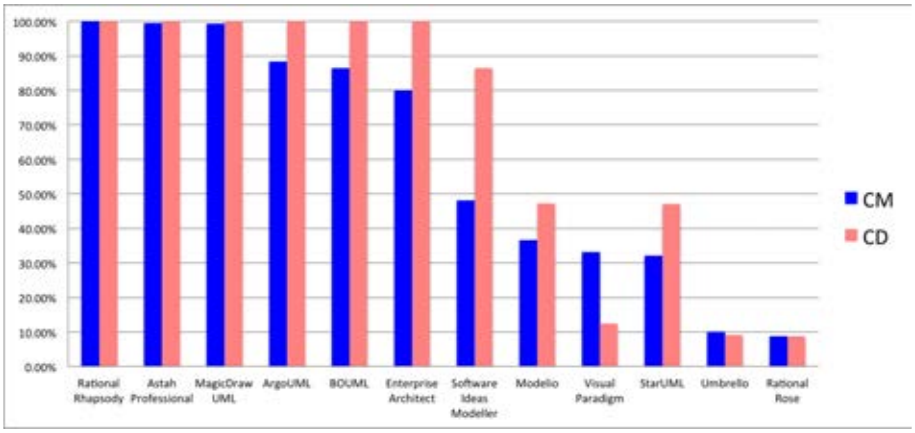


Figure 1: Overall performance for Compound Measure (CM) and Class Detection (CD)

When applied to industry tools a wide variation in output format and successful recreation of design information could be seen with performance scores ranging from 8.82% to 100%. Such a range of performance scores (Figure 1) demonstrates the effectiveness of RED-BM to differentiate between tools and underlines the complexity and coverage of the artefacts included. The analysis tools provided in the benchmark show very high accuracy when compared against validated manual analysis. Key findings with regard to the tools compared by the benchmark were a very high variance in tool performance with a difference of 91.18% between best and worst performers. Although the top performing tools were commercial, freeware such as ArgoUML outperformed the majority of commercial tools, including the widely used and expensive IBM Rational Rose.

References

[1] J. Roscoe, “Looking forwards to going backwards: An assessment of current reverse engineering,” *Current Issues in Software Engineering*, 2011.

[2] S. E. Sim, S. Easterbrook, and R. C. Holt, “Using benchmarking to advance research: A challenge to software engineering,” in *Proceedings of the 25th International Conference on Software Engineering*. IEEE Computer Society, 2003, pp. 74–83.

[3] UEA, “Reverse engineering to design benchmark,” http://www.uea.ac.uk/computing/machine-learning/traceability-forensics/reverse-engineering, 2013, [Online; accessed May 2013].

Language Identification by Textual, Audio and Visual Feature Mapping

Zhuoyi Dai, Richard Harvey, Stephen Cox, Yuxuan Lan

Speech Lab

{Zhuoyi.Dai, R.W.Harvey, S.J.Cox, Y.Lan}@uea.ac.uk

During last three decades, researching on language identification are mainly focus on text and audio, and most rubust models are produced on TLID(Text language Identification) and ALID(aduio language identification). However, the tasks on visual feature language are only started for several years and need to be more improved. In VLID(Visual Language Identification), the problem we got is that we do not have that many languages. What we concentrate is the whether we can use text and audio identification result to improve the identification performance on video by mapping.

Figure 1 describes a brief detail about the main idea. Assuming there are several languages exists in visual feature and some languages like L_1,L_2,L_3 could be key languages. We assume that we get the feature spaces for all languages and then income an unknown visual language. We would measure distances from unknown language to all languages by using nearest neighbor clustering. The question is whether we can develop a mapping from VLID distance result to ALID distance result or even more, to TLID distance result. The mapping of transferring from x, L_1, L_2, L_3 on VLID to ALID should be Sammon mapping in this case.

The other question is whether we can measure distances between languages with reasonable and reliable and make visual distance to audio distance mapping works. Through generating MFCCs features, we force those features through a Vector Quantisor with 64 bins and learn the frequency of occurrence in those various bins. For unseen language feature, we force it through the same vector quantisor and compare the distance between unseen language and exists languages which are previously trained. The distance between languages could prove whether language is correctly identified.

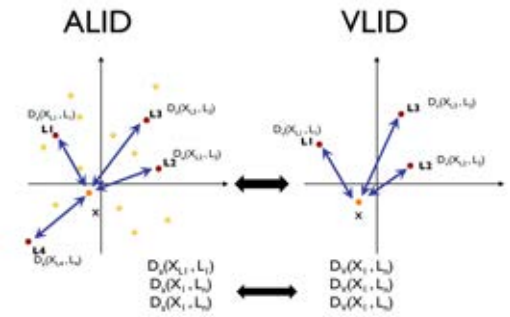


Figure 1: Description

We use two resource for our tasks. The NIST(National Institute of Standards and Technology) data corpus is a standard audio dataset and has been used in a large number of researches. It consists of telephone speech from 21 languages which are Arabic, Cantonese, Czech, Farsi, German, Hindi, Hungarian, Japanese, Korean, Malay, Mandarin, Italian, Polish, Portuguese, Russian, Spanish, Swedish, Swahili, Tamil, Vietnamese, and English. The audio feature is presented by MFCCs(Mel-Frequency Coefficients) for displaying spectral shape.

The video dataset created by Jacob Newman is focusing on lip boundary movement. He use "The Universal Declaration of Human Rights" as transcript and collect data from English, Mandarin and Arabic speakers. We track videos by image sequences and train AAMs (Active Appearance Models) for feature extraction. We currently interest in lip movement and AAMs aim to solving the boundary definition by mainly focus on lip area in our case.

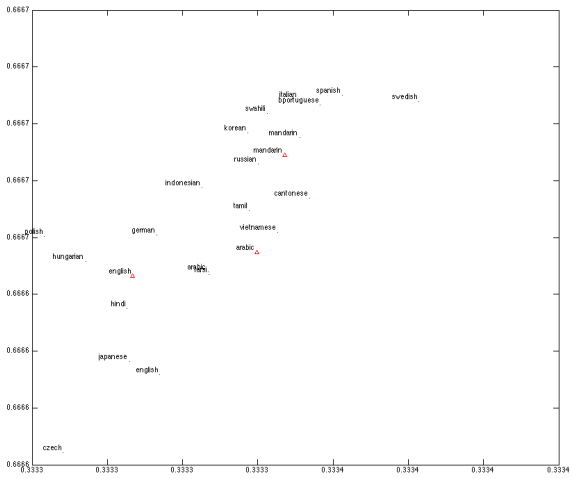


Figure 2: Bigram distance comparison result.

Figure 2 is an interpolation from visual distance result to audio distance result. This is an improved result based on bigram. From the image we can find visual Mandarin and visual Arabic distance are close to their audio results. But English is far from itself compare to Mandarin and Arabic and Language group are slightly not preserved.

From this task, we assume that unigram is not accurate enough to map visual and audio data and other factors such like gender, age and accent might effect on results. Since the result is not good enough, we assume VQ(Vector Quantisation) might not be a good method for language classification tasks.

Our further work would introduce zipping into our model. Since zipping performs good clustering results on TLID [1], we hope it works on visual and audio language classification. What is more, we also need to collect more data from different languages especially on visual part for our tasks.

References

[1] D. Benedetto, E. Caglioti, and V. Loreto, "Language trees and zipping," *Physical Review Letters*, vol. 88, no. 4, p. 048702, 2002.

Accent Information In Speaker Identification

Andrea DeMarco and Stephen Cox

Audio, Speech and Language Processing Laboratory
{a.de-marco,s.j.cox}@uea.ac.uk

In traditional speaker identification (SID) systems, features from speech utterances are used to build speaker models during training, which are then used to infer a speaker identity at testing time. However, the speech waveform contains a lot of information about various speaker characteristics. In this project, we are proposing a multi-layered approach to SID that includes the use of gender identification (GID) and accent identification (AID).

The first objective of the project was to build a robust GID system. The classifier constructed is a robust, unsupervised method of automatic gender identification from speech. We first design a baseline gender classifier based on MFCC features, and add a second classifier that uses context-dependent but text-independent pitch features. The results of these classifiers are then examined for disagreements in gender classification. Any disagreements are resolved by the use of a novel pitch-shifting mechanism applied to the utterances. We show how the acoustic-context classifier provides very good gender identification results, and how these are further enhanced by the pitch-shifting process. Furthermore, this enhancement is preserved across a set of different corpora [1].

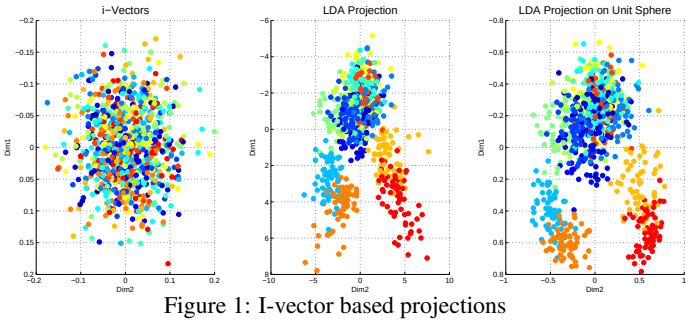


Figure 1: I-vector based projections

The second research objective was to devise an unsupervised AID algorithm that is accurate enough to be used for SID purposes. The crux of our work is based on the I-vector model. In traditional MAP adaptation, we can create an utterance specific GMM by adapting a UBM with utterance data. The resuling GMM is speaker/channel/utterance dependent. We can describe the entire GMM by the mean super-vector, which is of very high dimensionality. We can then apply PCA to this super-vector space to get to low dimensionality. The resulting subspace is known as *eigenvoices*. The coordinates of the super-vector relative to this basis is the theoretical *I-vector*. The components of the I-vector represent high-level characteristics of the utterance, irrespective of phonetic content (the UBM models phonetic variability). The problem with this concept is that the GMM super-vector is not observable. The I-vector model helps by constructing a prior model of latent super-vector variation by simple factor analysis over Baum-Welch statistics from an utterance. We can then achieve class-dependent modelling (accent-dependent in our case) by discriminative linear projections such as LDA, as in Figure 1.

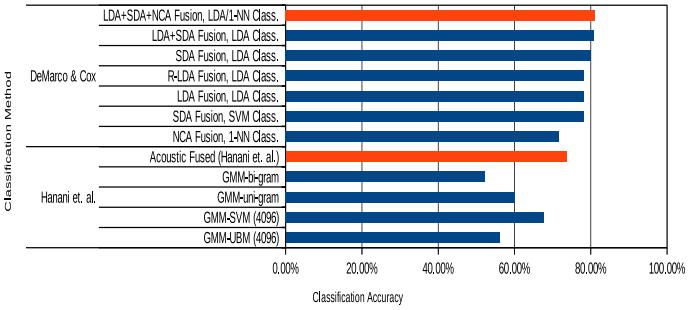


Figure 2: State-of-the-art unsupervised AID

Our work has presented a comprehensive analysis of the use of I-vector based classifiers for the classification of unlabelled acoustic data as native British accents [2, 3]. We demonstrated the different behaviours of various dimensionality reduction techniques that have been previously used in problems such as speaker and language classification. Our results (Figure 2) show that a fusion of I-vector based systems gives state-of-the-art performance (in comparison to previous results in [4]) for unlabelled classification of British accent speech data, reaching $\sim$81% accuracy. This AID system was also employed in accent and speaker adapted ASR systems in collaborative work with the University of Birmingham [5]. The final stage of our work will involve applying our unsupervised GID and AID classifiers to aid and improve SID results.

References

[1] A. DeMarco and S. J. Cox, "An accurate and robust gender identification algorithm," in *INTERSPEECH*, 2011.

[2] ——, "Iterative classification of regional british accents in i-vector space," in *MLSLP*, 2012.

[3] ——, "Native accent classification via i-vectors and speaker compensation fusion," in *INTERSPEECH*, 2013.

[4] A. Hanani, M. J. Russell, and M. J. Carey, "Human and computer recognition of regional accents and ethnic groups from British English speech," *Computer Speech and Language*, 2012.

[5] M. Najafian and A. DeMarco, "Supervised and unsupervised adaptation to regional accented speech using limited data for automatic speech recognition," submitted to: ICASSP 2014.

Visual Prosody in Speech Animation

David Greenwood and Barry-John Theobald

Graphics Laboratory
{david.greenwood, b.theobald}@uea.ac.uk

Speech animation involves transforming and deforming a human-like model, temporally synchronised to an audible utterance, to give the appearance that the model is speaking. The problem is challenging, as human viewers are very sensitive to natural human movement.

Practical applications of speech animation, for example computer games and animated films, often rely on motion capture devices or laborious hand animation. Both of these approaches are expensive and time consuming.

Recent research has led to the development of systems that can generate convincing automatic lip-syncing [1]. Currently, these systems are trained on expressionless data, i.e. an actor does not reveal any emotion or expression when delivering the utterance. Other visual cues: gaze direction, head movement and body posture are not considered, leaving these elements to conventional methods. Although claims have been made of completely realistic human animation, this is arguable, with near realistic results falling into the *Uncanny Valley* [2].

Linguists define prosody as "the rhythm, stress, and intonation of speech". Prosody can impart emotion, gravitas, emphasis or context that is additional to the vocabulary or grammar of the utterance. *Visual* prosody is the combination of hand gestures, eye gaze, head position, body posture or other visual cues during verbal discourse. These cues can semantically supplement the utterance, impart emphasis in the same way as spoken inflection and contribute to the natural appearance of human communication. Research spanning the past two decades considers why we make gestures [3], and the contribution of co-speech gesture to comprehension [4].

Figure 1: *Hand gestures are part of the natural process of communication.*

References

[1] S. L. Taylor, M. Mahler, B.-J. Theobald, and I. Matthews, "Dynamic units of visual speech," in *Proceedings of the 11th ACM SIGGRAPH/Eurographics conference on Computer Animation*. Eurographics Association, 2012, pp. 275–284.

[2] M. Mori, "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[3] A. Kendon, "Do gestures communicate? A review," *Res. Lang. Soc. Interact.*, 1994. [Online]. Available: http://www.tandfonline.com/doi/pdf/10.1207/s15327973rlsi2703_2

[4] K. Alahverdzhieva, "Alignment of speech and co-speech gesture in a constraint-based grammar," Ph.D. dissertation, Edinburgh, 2013. [Online]. Available: http://homepages.inf.ed.ac.uk/s0896251/pubs/main.pdf

Multispectral Image Fusion

Alex E. Hayes and Graham D. Finlayson

Colour Group
{Alex.Hayes,G.Finlayson}@uea.ac.uk

Image fusion involves combining information from two or more images of the same object or scene into a single composite image that contains more information or is otherwise more useful than any one of the individual original images. The focus of this research is multispectral image fusion, which involves combining images taken of the visible spectrum with those taken in the infrared or other spectral areas.

The pyramid of Gaussians (PoG) approach convolves the source images with Gaussian filters, which act as low-pass filters, only letting low frequency image components through. This resulting image is effectively blurred, and a pyramid is made of increasingly blurred images through repeated Gaussian convolutions. Then the ratios between the pixel intensities at each stage of the pyramid are taken, which represent the local contrast. The image fusion step is to take the maximum ratio at each pixel, at each stage, across all the pyramids for each input image, then invert the process to produce an output image[1].

The Socolinsky-Wolff method convolves the source images with horizontal and vertical gradient filters, then uses a 2x2 tensor matrix to combine these gradients through all the images. The eigenvector of this matrix represents the direction of maximal contrast at that pixel through all the images, and the square root of the eigenvalue is the magnitude of that contrast. The intensity image of the original colour visible spectrum image is then fit to this vector gradient field. This is accomplished by minimizing the following function:

$$\int\int_{\Omega} \bigtriangledown f - V^2 dxdy \tag{1}$$

Where Ω is the original multiband image, f is the output image, and V is the vector gradient field. This method is iterative, and converges to find a least-squares fit of the vector gradient field to the image[2].

The plan for this research is to investigate and implement the most common image fusion techniques, compare them using error metrics and subjective human testing, and work on new image fusion methods. Another area of action is to create more multispectral images, register and sort them, to create a multispectral data set.

References

[1] A. Toet, "Hierarchical image fusion," *Machine Vision and Applications*, vol. 3, no. 1, pp. 1–11, 1990.

[2] D. A. Socolinsky and L. B. Wolff, "Multispectral image visualization through first-order fusion," *Image Processing, IEEE Transactions on*, vol. 11, no. 8, pp. 923–931, 2002.

Shape-based Analysis of Time-series Data

Jon Hills and Tony Bagnall

Machine Learning and Statistics Group, School of Computing Sciences, University of East Anglia
{j.hills,anthony.bagnall}@uea.ac.uk

Shape-based analysis of time series focuses on subsequences of series, rather than whole series. A *shapelet* is a time-series subsequence that appears in members of one particular class, and does not appear in members of other classes (Fig.1 left). Shapelets are used for time-series classification. Our algorithm [1] performs a single pass through the training data to find a set of shapelets, and uses the shapelets to create a transform. The transformed data is represented as a set of distances between the shapelets and the series. For certain problems, transforming the data into shapelet space can make classification more accurate [1].

We also use shape-based analysis on unlabelled time series in an unsupervised process to discover *motifs*. Motifs are subsequences that recur in a longer series (Fig.1 right). The discovered motifs can be used to make predictions about the behaviour of the time series, or as primitives for other data-mining tasks.

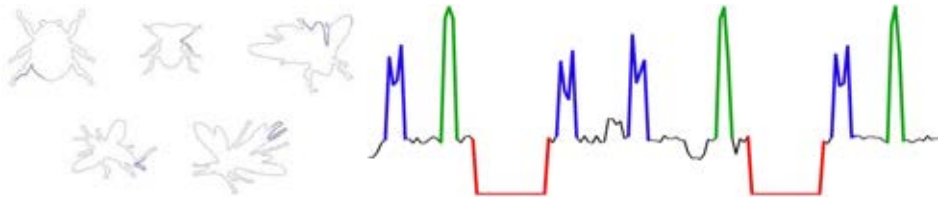


Figure 1: The best five shapelets for the Beetle/Fly dataset, superimposed over the image outlines (left). A simulated electricity-usage profile with motifs highlighted (right).

Clustering is an important subroutine in shape-based analysis of time series. For motif detection, clustering is essential to group the subsequences that represent repeated instances of the same pattern. The sequences will rarely be exact matches in realistic data, so an algorithm is necessary to determine if a given set of subsequences should be considered to match. Clustering is also applicable to the shapelet approach. The shapelet transform algorithm finds a large number of shapelets (by default, we find $10n$ shapelets, where n is the size of the training set). Often, many of the shapelets are matches of one another; they are different instances of the same pattern, a pattern that recurs among members of one class. Clustering such subsequence together offers a number of advantages. The dimensionality of the data is reduced, increasing the speed of classification, and improving the accuracy of certain classifiers by removing the highly correlated shapelets that represent the same pattern. Clustering shapelets also increases the interpretability of the data by removing duplicate shapelets.

Rule discovery and shape-based analysis can be combined to provide highly interpretable models. Such models can be used with motif discovery to predict the occurrence of patterns in a time series, or with shapelets to classify time series. For classification, the strength of rule-based systems is in classifying minority classes, as they can be used for partial classification, detecting only the class of interest [2, 3]. To make use of existing rule-discovery algorithms, and to improve the interpretability of the data, we discretise the shapelet-transformed data into binary data, where each attribute is 0 where the shapelet distance is greater than the information gain splitting point and 1 otherwise. Algorithms such as *Apriori* can then be applied straightforwardly to the transformed data.

References

[1] J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall, "Classification of time series by shapelet transformation," *Data Mining and Knowledge Discovery*, pp. 1–31, 2013.

[2] J. Hills, L. M. Davis, and A. Bagnall, "Interestingness measures for fixed consequent rules," in *Proceedings of the 13th international conference on Intelligent Data Engineering and Automated Learning*. Springer-Verlag, 2012, pp. 68–75.

[3] J. Hills, A. Bagnall, B. de la Iglesia, and G. Richards, "Brutesuppression: a size reduction method for apriori rule sets," *Journal of intelligent information systems*, vol. 40, no. 3, pp. 431–454, 2013.

Confusion Modelling for Automated Lip-Reading

Dominic Howell and Stephen Cox

Audio, Speech and Language Processing Laboratory
{dominic.howell, s.j.cox}@uea.ac.uk

Automated lip-reading involves recognising speech from only the visual signal. Recent studies have shown that automatic lip-reading performs significantly worse than audio speech recognition [1]. These poor results are most likely due to the lack of speech information available in a visual signal (for example, the position of some articulators cannot be seen, and there is no way to tell whether a sound is voiced or unvoiced). In addition, the purpose of speech is to produce distinctive sounds to convey a message, and the particular mouth-shapes used to produce these sounds are (usually) of no concern to the speaker: it is quite possible to produce a perfectly intelligible audio signal from mouth-shapes that are not distinct, something that is verified by human lip-readers who report that some people are much more "readable" than others. Furthermore, mouth shapes are severely affected by co-articulation [2]. Because of these limitations, human lip-readers make heavy use of pragmatics and contextual information to understand what is being spoken [3].

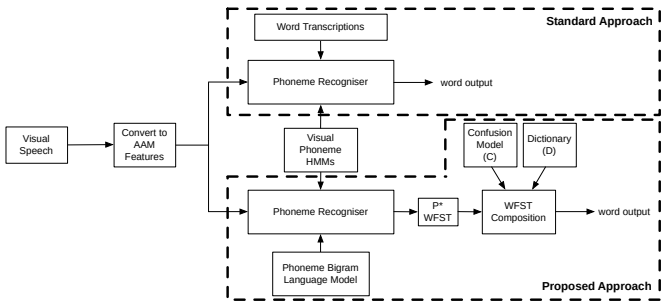


Figure 1: A comparison of the standard approach and proposed approaches to lip-reading

Visual speech has an interesting relationship to dysarthric speech. Dysarthric speakers have poor control over their articulators due to medical conditions that affect their motor functions (such as cystic fibrosis, stroke etc.). This leads to a phonemic repertoire that is both reduced and distorted, and hence to speech that has low intelligibility, and is difficult to recognise - an obvious parallel with visual speech, where certain sounds cannot be distinguished visually. In previous work on dysarthric speech recognition, we learnt patterns of phonemic confusions from a talker, and when these confusions were compensated at recognition time, recognition accuracy increased [4]. We take a similar approach to lip-reading: we model visual speech as a speech signal produced by a speaker who has a limited phonemic repertoire, and learn the patterns of confusion between the ground-truth phoneme sequences and the recognised sequences. At recognition time, we find the most likely interpretation of a reduced/distorted phoneme output sequence in the light of these patterns, as was successfully explored in [4].

Figure 1 illustrates the "standard" approach and our proposed approach. In this study, we use utterances of isolated words. Both approaches begin by converting the visual signal into a sequence of feature vectors. Hidden Markov models (HMMs) of each phoneme are then built. In the standard approach, the input feature vector sequence is decoded by forming a network of phoneme models such that any path through the network represents the transcription of one of the words in the vocabulary. The most probable route through this network is found using the Viterbi algorithm, and the word associated with this route is the recognised word. In our proposed approach, the recogniser first decodes a set of n-best phoneme sequences under the influence of a phone bigram language model and represents this set as a transducer, P. These sequences are passed to a second transducer, C, which is a model of confusions made in this decoding, built by passing a hold-out set through the recogniser. C then expands P into a much larger set of hypotheses, together with their associated probabilities. A third transducer D allows only hypotheses that represent vocabulary words to be decoded. The word associated with the most likely path through the transducer cascade is selected as the recognised word.

References

[1] S. Cox, R. Harvey, Y. Lan, J. L. Newman, and B.-J. Theobald, "The challenge of multispeaker lip-reading," *In Proceedings of the International Conference on Auditory-Visual Speech Processing*, pp. 179 – 184, 2008.

[2] P. Jackson, "The theoretical minimal unit for visual speech perception: Visemes and coarticulation." *The Volta Review*, 1988.

[3] P. Hansen and M. Coleman, "Speechreading skill and visual movement sensitivity are related in deaf speechreaders," *Perception*, vol. 34, pp. 205–216, 2005.

[4] S. O. C. Morales, "Error modelling techniques to improve automatic recognition of dysarthric speech," Ph.D. dissertation, UEA, May 2009.

Robust Speech Enhancement by Reconstruction with Model-Based Acoustic Features

Akihiro Kato and Ben Milner

Audio, Speech and Language Processing Laboratory
{Akihiro.Kato, B.Milner}@uea.ac.uk

Traditional methods of speech enhancement apply filters to remove an estimate of the noise, therefore, it is inevitable to have some deterioration in original speech such as musical noise in particular severe SNR condition. This research will move away from this traditional approach to filter the noise, instead, reconstructs a clean speech signal with a set of acoustic features generated based on the original signal and trained acoustic models. The basic sequence of the proposed method to realise this purpose is illustrated in Figure 1. In this research, some techniques in STRAIGHT[1], which is a high quality channel VOCODER, are studied whether it can be utilised as the speech model of the proposed method, and some techniques for the feature extraction in STRAIGHT are improved to build the speech model with high intelligibility.

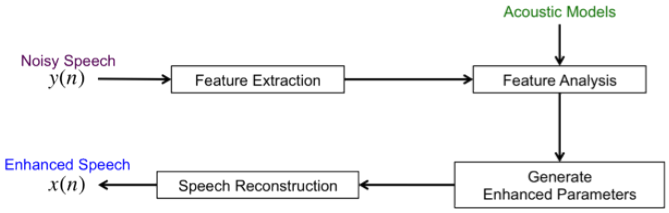


Figure 1: Basic sequence of the proposed method

Feature extraction process in STRAIGHT has some non-linearity because of smoothing operations in the both frequency and time domain and an operation of over-smoothing compensation[1][2]. In some cases, this non-linearity could cause less intelligibility for the system, therefore, some feature extraction methods are considered in order to improve the linearity.

The improved feature extraction can be achieved with pitch-adaptive windows compensated by the second order cardinal B-spline function. This method reduces the non-linear processes used in STRAIGHT, and it can gives enough intelligibility in the speech reconstruction in spite of no smoothing in the frequency domain. The improvement of the linearity is confirmed by an experiment of spectral subtraction with reference noise. The result is shown in Figure 2.

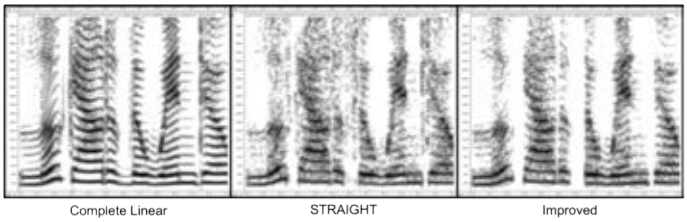


Figure 2: Improvement in the linearity.

Adopting the improved speech model, the Enhanced part should be built in the next stage. Key studies for the stage of the research are to detect SNR of input speech by using some techniques of ASR with sub-band HMMs, and to apply missing feature theory depending on the detected SNR in order to generate enhanced feature parameters as a combination of the features from the original and trained acoustic models.

References

[1] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigne, "Restructing speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, no. 3-4, pp. 187–207, April 1999.

[2] H. Kawahara, "Speech representation and transformation using adaptive interpolation of weighted spectrum: Voceder revisited," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1303–1306, April 1997.

Synthesis of an intelligible audio speech signal using information solely derived from visual speech features

Thomas Lee Le Cornu and Ben Milner

Audio, Speech and Language Processing Laboratory
{t.le-cornu,b.milner}@uea.ac.uk

The aim of the project is to build a system that is capable of synthesising an intelligible audio speech signal using only visual features extracted from the face of a speaker. Such a system would have application in surveillance scenarios when traditional audio recording methods were unusable.

The system requires the development of a speech model that can generate intelligible speech from a highly reduced set of acoustic speech features, and the discovery of robust methods of extracting acoustic speech features from visual speech features.

The acoustic feature information contained within a visual speech signal is considerably minimal. The only information that may be available is the spectral envelope from the lips of the speaker, and the fundamental frequency, phase, and voicing classification are missing. Therefore, estimations need to be made for these missing parameters. Research has been undertaken on the sinusoidal model of speech [1], and methods for the estimation of fundamental frequency and extraction of spectral envelope [2].

The sinusoidal model states that a speech signal is the sum of a set of k sinusoids of varying amplitude, frequency, and phase. During voiced frames of speech the sinusoids have a harmonic relationship, with those of higher frequency being integer multiples of the fundamental frequency, f_0. By exploiting this harmonic relationship the sinusoidal model can be reduced to

$$s(n) = \sum_{k=1}^{K} A(k)(2\pi k f_0 n + \phi_k). \tag{1}$$

where $s(n)$ is the output audio speech signal, K is the largest real number that satisfies the inequality $kf_0 < \frac{f_s}{2}$ (where f_s is the sampling frequency), f_0 is the fundamental frequency, $A(k)$ is the spectral envelope, and ϕ_k is the phase. The necessary input parameters for equation 1 are the fundamental frequency and spectral envelope; the phase of each sinusoid is interpolated to ensure that it is smooth across frame boundaries.

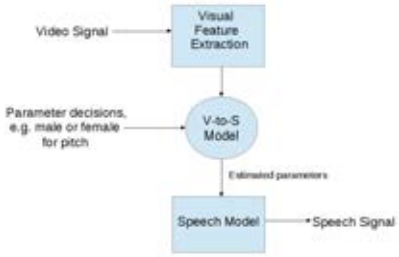


Figure 1: Components of the proposed system to generate intelligible audio speech from an input visual signal of a speaker's face.

To accurately estimate acoustic parameters from visual speech features, research will be undertaken on statistical estimation methods of spectral envelope from visual features. Such methods have been shown to be effective for creating a visually-derived Wiener filter for speech enhancement [3]. The sparseness of acoustic information contained in the visual features requires such statistical methods to rely on significant amounts of prior information and examination will be made into obtaining such information.

Further research on speech models and speech production is being continued, and research of visual speech feature extraction methods will begin in due course.

References

[1] W. B. Kleijn and K. K. Paliwal, *Speech coding and synthesis*. Elsevier Science Inc., 1995.

[2] H. Kawahara and H. Katayose, "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity." *EuroSpeech*, vol. 99, no. 6, pp. 2781–2784, 1999.

[3] I. Almajai and B. Milner, "Visually derived wiener filters for speech enhancement," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1642–1651, 2011.

Identification of gene fusions in cancer using exon microarrays

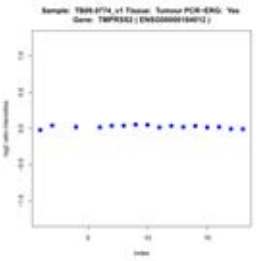
Bogdan Luca, Dan Brewer, Colin Cooper, Vincent Moulton

Computational Biology Lab
b.luca@uea.ac.uk

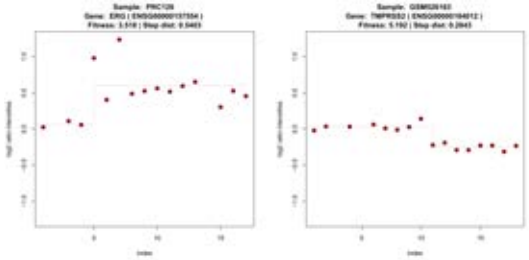
Cancer is a disease caused by a cascade of DNA alterations leading to abnormal cell division and cell transformation. Some of the alterations, such as *gene fusions* generated by chromosomal translocations, are recurrent in cancer and can be used as biomarkers for a better understanding and management of the disease (e.g. *TMPRSS2-ERG* fusions appears in approximately 50% of prostate cancers and has been associated with poorer survival outcome [1]).

Exon microarrays can be used to detect genes involved in fusions [2]. We developed a novel technique able to detect such fusion candidate genes (see examples in Figure 1). Our method normalises the data to a reference sample, which reduces the bias induced by alternative splicing, fits a custom step function, to detect the possible fusions and ranks the candidate fusion genes. The step function uses a special fitness function we created to assign a score and statistical significance to each detected fusion breakpoint.

The approach can be generally applied to different types of cancer. We successfully applied it to prostate, breast, gastric, colorectal and lung cancer datasets, detecting known fusions such as *ERG/ETV1-TMPRSS* in prostate cancer, or *VTI1A-TCF7L2* in colorectal cancer. We also discovered new candidate fusions, which might give us a better understanding of the disease.



(a) normal gene



(b) fusion pattern 1 (c) fusion pattern 2

Figure 1: Examples of normal and fused genes patterns.

References

[1] J. Clark and C. Cooper, "Ets gene fusions in prostate cancer," *Nature Reviews Urology*, vol. 6, no. 8, pp. 429–439, 2009.

[2] S. Jhavar, A. Reid, J. Clark, Z. Kote-Jarai, T. Christmas, A. Thompson, C. Woodhouse, C. Ogden, C. Fisher, C. Corbishley *et al.*, "Detection of tmprss2-erg translocations in human prostate cancer by expression profiling using genechip human exon 1.0 st arrays," *The Journal of Molecular Diagnostics*, vol. 10, no. 1, pp. 50–57, 2008.

Simulation and Mathematical Modelling of Microtubule Organization

Alex Mace and Wenjia Wang

Computational Biology Laboratory, School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK
{alex.mace,wenjia.wang}@uea.ac.uk

Microtubules are a part of the cell cytoskeleton. There are long cylindrical tubes composed of alpha and beta tubulin heterodimers with a diameter of approximately 25 nm. Microtubules undergo dynamic instability, switching from a growing to shrinking state (catastrophe) and the reverse (rescue). The roles of microtubules within a cell are many and diverse including the positioning of organelles, intra-cellular transport, cell motility, cell polarisation, and cell replication [1]. To better understand how microtubules are able to regulate their stability and positioning in order to perform these roles, computational modelling techniques are used.

In animal cells, microtubules are typically nucleated at a microtubule organizing centre (MTOC) which contributes to their structure. However, in Madin-Darby Canine Kidney cells (MDCK) microtubules are observed predominantly in an array running from the apex to the base of the cell and not on a MTOC. It is hypothesised that microtubules originally nucleate from a MTOC but are captured and then pulled onto the cell membrane [2] without microtubules interacting with each other. We modelled this hypothesis using a combination of both direct simulation and representing microtubules as partial differential equations. A large scale parameter search on the catastrophe and rescue rates of the MTOC and the apico-basal populations revealed that the hypothesis could work under biologically reasonable parameters assuming the apico-basal array had an enhanced stability relative to the MTOC array.

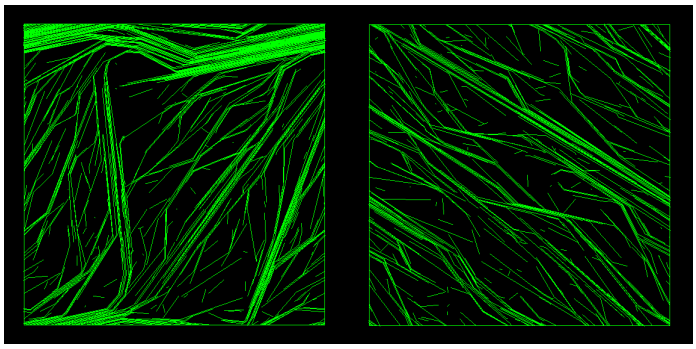


Figure 1: The ordering of cortical microtubules after a simulated 120 minutes with a wild-type catastrophe rate (right) and a fifth of the wild-type rate (left) which causes the cell to have two ordered regions.

In cortical plant cells, microtubules form highly ordered arrays but do this without the aid of MTOCs. Instead the interactions between microtubules, such as 'zippering' where a collision between two microtubules can cause the incident microtubule to grow parallel alongside the other, have a strong role in the ordering. Such interactions between microtubules make simulation studies more convenient than differential equation models. Currently, using a simulation model, we have investigated the effect of the spontaneous catastrophe rate. Increasing the spontaneous catastrophe rate above the wild-type value causes both the size and order of the array to decrease as microtubules do not live long enough to interact. Reducing the spontaneous catastrophe rate increases the size of the array but reduces the overall order as it increases the lifetime of microtubules that are not orientated in the dominant direction (Figure 1).

References

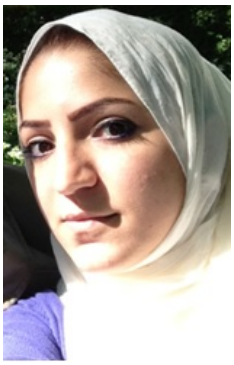
[1] A. Desai and T. J. Mitchison, "Microtubule polymerization dynamics," *Annual review of cell and developmental biology*, vol. 13, no. 1, pp. 83–117, 1997.

[2] G. Bellett, J. M. Carter, J. Keynton, D. Goldspink, C. James, D. K. Moss, and M. M. Mogensen, "Microtubule plus-end and minus-end capture at adherens junctions is involved in the assembly of apico-basal arrays in polarised epithelial cells," *Cell motility and the cytoskeleton*, vol. 66, no. 10, pp. 893–908, 2009.

A clustering analysis framework for heterogeneous datasets

Aalaa Mojahed, Beatriz de la Iglesia and Wenjia Wang

Machine Learning and Statistics Laboratory
{a.mojahed, B.Iglesia, Wenjia.Wang}@uea.ac.uk



Datasets in real-world applications are often heterogeneous where they comprise several data categories such as structured data, images, time-series, videos and others. Data heterogeneity, in this sense, is one of the biggest challenges that face pattern analysis tasks such as classification, clustering, rule induction... etc. Few research efforts are devoted to mining heterogeneous data sets so our research aims to make some significant contribution to this area. We start by trying to draw a framework for clustering large heterogeneous data sets. This research aims to define sinificant contribution by filling this gap; it attempts to draw a framework for clustering datasets with the curse of heterogeneity.

Clustering is an unsupervised classification technique where a set of patterns are organized into related groups based on some similarity or dissimilarity measures, without using predefined classes [1]. Distance measures play a critical role in clustering analysis. Applying appropriate measures results in more accurate configurations of the data [2]. Several similarity measures have been proposed and tested in the literature. These measures range from simple approaches which reflect the dissimilarity between two patterns based on their attribute values to others like those that categorize the conceptual similarity [3].Table 1 indicates some of the most reliable and widely used distance measures that produce accurate results in different scenarios. These are categorized by the data type which they deal with as different data types rely on different similarity measures. Most of the available similarity measures can only be applied to one type of data, however, some measures may work with quantitative (can be binary, discrete, continuous... etc) and qualitative attributes (can be represented as binary or discrete)[4, 5]. In this context, it is important to construct an appropriate similarity measure for comparing complex objects which described by components from different data categories.

Table 1: Similarity measures for different data type

data type	similarity measures
binary	Russell/Rao [6]
	Jaccard coefficient [7]
	Rand statistic [8]
	Dice's cofficient [9]
	Fowlkes and Mallows index [10]
discrete and continuous	Minkowski [11]
	Salton's Cosine [12]
	Pearson Correlation [13]
	Averaged Kullback-Leibler Divergence [14]

Six key objectives structure the road-map of this research, and these are:

1. State the problems and challenges of mining complex data and review all related studies.
2. Identify a unified data representation method for complex data that is capable of denoting structured data along with unstructured data types. The representation scheme should be extendable to fit other unstructured types, and initially it should be well-matched to existing clustering algorithms.
3. Determine a precise distance measurements for heterogeneous objects, either from existing techniques or proposing an original calculations that take into account the characteristics of the data.
4. Gather the available clustering algorithms that are capable of mining heterogeneous data. Then, try to combine or edit them to satisfy the research aim. Alternatively, develop an original clustering algorithm.
5. Apply the deliverables from the above study objectives on some real data.
6. Interpret the outcomes by evaluating the clustering results using validation methods.

At this point of the research, the first three objectives are delivered and as a result, an averaged weighted similarity formula using existing similarity metrics has been proposed to measure the similarity among heterogeneous objects. This approach has not yet been experimentally tested. The initial plan is to begin with a limited number of data types but make the approach extensible to others. The heterogeneous objects initially contain three types of data: structured data, free text and time-series. In order to develop this combination, a synthetic heterogeneous dataset has been created. Three individual data sources have been involved in generating this desired heterogeneity.

References

[1] J. Han and M. Kamber, *Data Mining Concepts And Techniques*, 2nd ed. Morgan Kaufmann, San Francisco, 2006.

[2] M. Steinbach, L. Ertoz, and V. Kumar, "The challenges of clustering high dimensional data," in *New Direction of Statistical Physics: Applications In Econophysics, Bioinformatics, And Pattern Recognition*, D. J. Hand, N. M. Adams, and R. J. Bolton, Eds. Springer-Verlag, 2003, pp. 237–302.

[3] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, USA, 1992.

[4] Y. Bashon, D. Neagu, and M. J. Ridley, "A framework for comparing heterogeneous objects: On the similarity measurements for fuzzy, numerical and categorical attributes," *Soft Computing*, vol. 17, no. 9, pp. 1595–1615, 2013.

[5] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, no. 1, pp. 1–34, 1997.

[6] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *Journal of the Royal Statistical Society, Series B*, vol. 10, pp. 159–193, 1948.

[7] S. Jaccard, "Nouvelles researches sur la distribution florale," *Bull. Soc. Vaud. Sci. Nat*, vol. 44, pp. 223–270, 1908.

[8] W. M. Rand, "Objective criteria foe the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1958.

[9] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297–302, 1945.

[10] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clustering," *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 553–569, 1983.

[11] E. F. Krause, *Taxicab Geometry: An Aadventure in Non-Euclidean Geometry*, 1987.

[12] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, New York, NY, USA, 1987.

[13] J. L. Rodgers and W. A. Nicewander, "The relation between Pearson's correlation coefficient and Salton's cosine measure," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 5, pp. 1027–1036, 1988.

[14] S. Kullback, *Information Theory and Statistics*, New York, NY, USA, 1959.

Constructing phylogenetic networks from noisy data using trinets

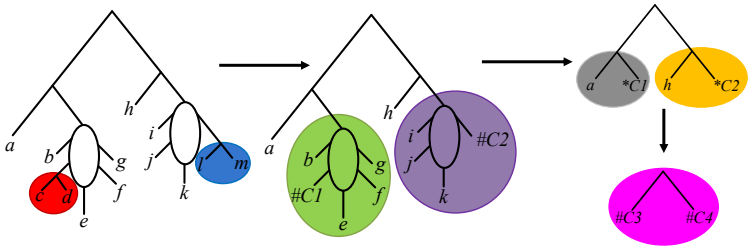
James Oldman, Vincent Moulton and Taoyang Wu

School of Computing Sciences, University of East Anglia
{J.Oldman,V.Moulton,Taoyang.Wu}@uea.ac.uk

The motivation of phylogenetic analysis is to discover the evolutionary relationships between a set of species and to understand the evolution of life. These relationships are often visualised and represented using a phylogenetic tree. Even though phylogenetic trees are interesting structures in their own right, in recent years there has been significant interest in representing evolution that is not best visualised on a phylogenetic tree. As a natural generalisation of phylogenetic trees, phylogenetic networks are used in biology to represent evolutionary histories that contain reticulate, or non-treelike events such as recombination, hybridisation and horizontal gene transfer. Such reticulate events are believed to occur in organisms such as bacteria, plants, viruses and certain groups of fishes and frogs [1].

The reconstruction of explicit phylogenetic networks from biological data is currently an active area of phylogenetics. Here we consider the problem of constructing such networks from trinets, that is, phylogenetic networks on three leaves. The core motivation of our work is to develop an algorithm to construct level-1 rooted phylogenetic networks from real data. The work here extends that of [2] by presenting an algorithm that will, given a dense set of trinets as input, always construct a level-1 rooted phylogenetic network. The algorithm has been designed to address potentially noisy input data and still return a level-1 network.

Loosely speaking, we adopt a bottom-up approach. That is, given a dense trinet set $\mathcal{T}$ on a leaf set X, we first identify an appropriate subset Y of X and construct a level-1 network N_Y on Y using the restriction of $\mathcal{T}$ on Y. Next, we compute the trinet set $\mathcal{T}^*$ induced by $\mathcal{T}$ on the set X^* formed by replacing the subset Y in X with a new element y^*. For the implementation of the algorithm, these subsets are recorded into lists of the cherry and cactus structures identified during the bottom-up phase of the algorithm. These recorded structures are then used to reconstruct the network from the root down to the leaves.



Cherry Name	Leaf One	Leaf Two
#C1	c	d
#C2	l	m
#C3	a	*C1
#C4	h	*C2
#C5	#C3	#C4

Cactus Name	Bottom Leaf	Left Leaves	Right Leaves
*C1	e	b, #C1	g, f
*C2	k	i, j	#C2

Figure 1: An example of the bottom-up phase of the network construction algorithm

References

[1] C. Than, D. Ruths, and L. Nakhleh, "Phylonet: a software package for analyzing and reconstructing reticulate evolutionary relationships," *BMC Bioinformatics*, vol. 9, no. 1, p. 322, 2008. [Online]. Available: http://www.biomedcentral.com/1471-2105/9/322

[2] K. Huber and V. Moulton, "Encoding and constructing 1-nested phylogenetic networks with trinets," *Algorithmica*, vol. 66, no. 3, pp. 714–738, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00453-012-9659-x

Free surface flows over submerged obstructions

C. Page, [1] *S. Laycock,* [1] *E.I. Părău* [2] *and S. Grandison*

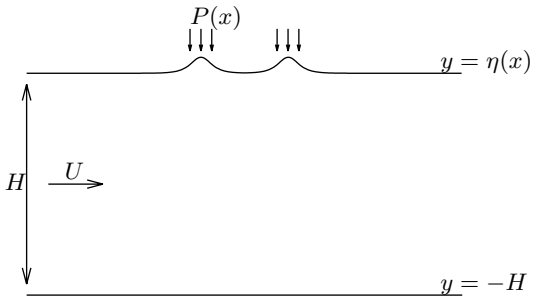
[1]School of Computing Sciences, [2] School of Mathematics
{charlotte.page,s.laycock,e.parau}@uea.ac.uk

We investigate the classical problem of the irrotational flow of an ideal fluid, passed various disturbances in a channel, in the form of either localised applied pressure distributions on the free surface or submerged obstructions on the bottom of the channel. There exists a large amount of literature on weakly nonlinear solutions to this classical problem, especially in the case of pure-gravity flows where the effects of surface tension on the free surface have been neglected. There are fewer fully nonlinear results. We seek steady gravity-capillary solutions to this classical problem, using fully nonlinear boundary integral equation techniques based on the Cauchy integral formula. The integrodifferential equations are solved iteratively using Newton's method. The behaviour of the gravity-capillary fully nonlinear solutions is characterised by the dimensionless Froude number F and the dimensionless Bond number τ

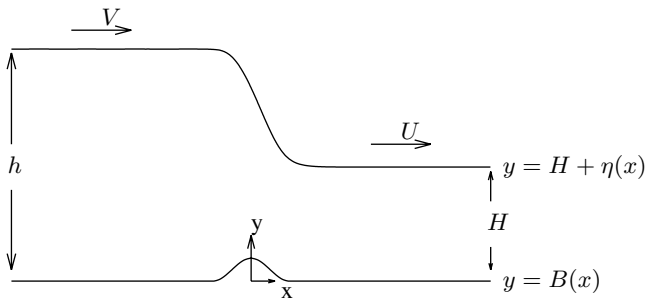
$$F = \frac{U}{\sqrt{gH}} \qquad \text{and} \qquad \tau = \frac{\sigma}{\rho g H}, \tag{1}$$

as well as by the orientation of the underlying forcing.

Maleewong, Asavanant and Grimshaw (2005) studied fully nonlinear gravity-capillary solutions passed a single disturbance in the channel. Binder, Dias and Vanden-Broeck (2005) studied fully nonlinear pure-gravity flows over multiple submerged obstructions. We combine and extend this work by considering subcritical ($F < 1$) gravity-capillary flows, with strong surface tension ($\tau > 1/3$) passed two localised pressure distributions. [1]. We further consider critical (hydraulic fall) solutions, where the flow upstream is subcritical and downstream is supercritical ($F > 1$) [2]. By utilising multiple obstructions on the bottom of the channel, we look for 'trapped wave' solutions where a train of waves appear, trapped between two obstructions. The addition of a further obstacle upstream of the hydraulic fall in the pure gravity case, results in a train of trapped waves appearing upstream between the obstacles (Dias and Vanden-Broeck 2004). We extend this work by considering the effects of surface tension on the free-surface. Critical flow solutions over a single submerged obstruction are studied, and we show that trapped waves are found, in this case, by including an additional obstruction downstream of the hydraulic fall. Furthermore, provided that the surface tension is very small, a train of trapped waves may also be found between the hydraulic fall and an upstream obstruction, as in the pure-gravity case.



(a) Non-critical flow configuration passed two pressure distributions.



(b) Critical flow configuration over an arbitrary obstacle.

We go on to consider the effect of a thin continuous ice sheet covering the free-surface [3]. Hydraulic falls and trapped wave solutions are sought as in [2], by modelling the ice sheet as a thin elastic shell using the special Cosserat theory of hyperelastic shells.

The aim now is to consider time dependent gravity-capillary hydraulic falls, where the obstacle on the bottom of the channel slowly moves to the left.

References

[1] C. Page, E. Părău, and S. Grandison, "Gravity-capillary water waves generated by multiple pressure distributions," *The Journal of Nonlinear Science and Applications*, vol. 6, pp. 137–144, 2013.

[2] C. Page, S. Grandison, and E. Părău, "The influence of surface tension upon trapped waves and hydraulic falls," *European Journal of Mechanics - B/Fluids*, vol. 43, pp. 191 – 201, 2014.

[3] C. Page and E. Părău, "Hydraulic falls under a floating ice plate due to submerged obstructions," *Journal of Fluid Mechanics*, in review.

miRNA detection and analysis from high-throughput small RNA sequencing data

Claudia Paicu, Tamas Dalmay, Simon Moxon, Vincent Moulton

School of Computing Sciences, University of East Anglia
{c.paicu, t.dalmay}@uea.ac.uk, simon.moxon@tgac.ac.uk, v.moulton@uea.ac.uk

Recent developments in high-throughput sequencing technologies have led to large datasets containing tens of millions of small RNA sequences. This has led to new challenges for the tools used to analyse such data. MicroRNAs (miRNAs), a well-studied classes of sRNAs, need to be identified and analysed because of their important cellular functions in gene regulation. These tiny, 22-nucleotide RNAs control many pathways including developmental timing, hematopoiesis, organogenesis and development, apoptosis, cell proliferation and possibly even tumorigenesis [1].

There are several tools for miRNA prediction ([2],[3]) from high-throughput sequencing datasets each suffering from both false-positive and false-negative predictions. These tools predict different miRNAs from the same input suggesting that further improvements to such software are required.

The goal of this project is to improve miRCat [3] in order to get a better sensitivity and specificity as well as reducing memory requirements and decreasing run time to keep pace with the growing datasets produced in high-throughput sequencing experiments. In order to achieve that two approaches will be considered: firstly, current tools will be analysed and compared; secondly, we will make use of mutant data to develop a new algorithm.

References

[1] V. Narry Ki, "MicroRNA biogenesis: Coordinated cropping and dicing," *Molecular Cell Biology*, vol. 6, pp. 376–382, 2005.

[2] M. R. Friedlnder, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, and N. Rajewsky, "Discovering microRNAs from deep sequencing data using miRDeep," *Nature Biotechnology*, vol. 26, pp. 407–414, 2008.

[3] M. B. Stocks, S. Moxon, D. Mapleson, H. C. Woolfenden, I. Mohorianu, L. Folkes, F. Schwach, T. Dalmay, and V. Moulton, "The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets," in *Bioinformatics*, 2012, pp. 2059–2061.

Lassoing and corralling rooted phylogenetic trees

Andrei-Alin Popescu and Katharina T. Huber

Computational Biology Lab
{Andrei-Alin.Popescu,Katharina.Huber}@uea.ac.uk

For a set X of individuals, the construction of a dendogram on X is a key component of, for example, genome wide association studies. A dendogram is a rooted edge-weighted tree (T,ω) with no degree 2 vertices except possibly the root and leaf set X and edge-weighting $\omega : E(T) \longrightarrow \mathbb{R}$, such that the distance between the root and a leaf x is the same for all $x \in X$. See Fig. 1 for two examples of a dendogram with leaf set $\{a,b,c\}$. However even with modern sequencing technologies the distances on the individuals required for the construction of such a structure may not always be reliable making it tempting to exclude them from an analysis. This, in turn, results in an input set for dendogram construction that consists of only partial distance information which raises the following fundamental question. Suppose we have a dendogram T with leafset X. Then for what sets $\mathscr{L}$ of pairs of elements of X, is T uniquely determined (or lassoed) by the distances T induces on the elements of $\mathscr{L}$?

Depending on the extent to which $\mathscr{L}$ determines (T,ω), we can define three types of lassos. $\mathscr{L}$ is an *equidistant lasso* if it uniquely determines the edge-weights of (T,ω), a *topological lasso* if it uniquely determines the topology of T and a *weak lasso* if it only allows for *refinements* of T. T' is said to be a refinement of T if T can be obtained from T' by collapsing edges. See Fig. 1 for an example of an equidistant lasso and an illustration of refinement.

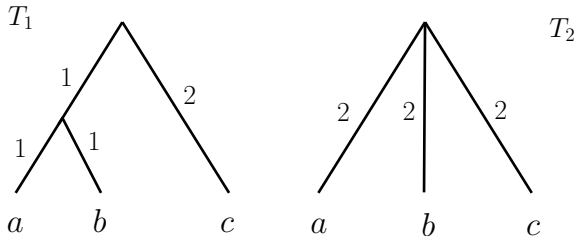


Figure 1: The dendograms T and T' with indicated edge-weights induce the same distance on the leaf pairs (a,c) and (b,c), but they are not the same. T is a refinement of T', while $\mathscr{L} = \{ab,ac\}$ is an equidistant lasso of T

In [2] we show that for a dendogram T, the above types of lassos can be characterised in terms of a special graph called the child-edge graph defined at an interior vertex v for some $\mathscr{L}$ and T. These characterisations span the spectrum of connectivity of this child-edge graph, with equidistant lassos and topological lassos occupying the extreme ends, and weak lassos occupying a location in-between. We refer the reader to [1] for discussions of the analogous problem for unrooted phylogenetic trees.

References

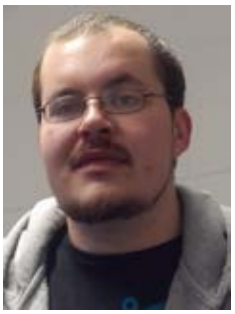
[1] A. W. M. Dress, K. T. Huber, and M. Steel. 'Lassoing' a phylogenetic tree I: basic properties, shellings and covers. *J Math Biol.*, 65:77–105, 2012.

[2] K.T. Huber and AA. Popescu. Lassoing and corralling rooted phylogenetic trees. *Bull Math Biol.*, 75:444–465, 2013.

Shape from colour: Depth without disruption

Chris Powell and Graham Finlayson

The Colour Group
{Christopher.Powell,G.Finlayson}@uea.ac.uk



Many techniques have been developed in computer vision for recovering three-dimensional shape from two-dimensional images; such as digital photographs. Broadly speaking, these techniques exploit information (e.g. radiance, irradiance, shading, colour, shape cues) from one or more images of the same object and produce an estimate of its shape in some format (e.g. a mesh, depth/height map, surface normals).

Shape recovery is a problem which is far from being solved. For example, does a darkened area of an object represent an indentation or is it a shadow cast from something else in the scene? Efforts to address such issues have involved various combinations of restrictions upon scene conditions, prior knowledge of subjects and/or assumptions about image contents.

In our work, we are particularly interested in a method known as "Shape from Color" [1]. It requires spectrally distinct light sources that are frontally positioned and ideally at a distance. In additions the spatial position of the light sources must not be coplanar.

However, restrictive lab conditions can feel rather intrusive or uncomfortable when applied to a human subject (imagine being sat in a dark room with three differently coloured lights shining directly at your face)! We are interested in discovering more natural conditions which allow for acceptable recovery of shape as well as a process that can go unnoticed.

References

[1] M. S. Drew, "Shape from color," Citeseer, Tech. Rep., 1992.

Real-Time Accumulation of Occlusion-Based Snow

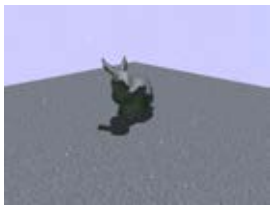
DT Reynolds, SD Laycock and AM Day

Graphics, Colour and Visualisation Laboratory
{daniel.reynolds,s.laycock,andy.day}@uea.ac.uk

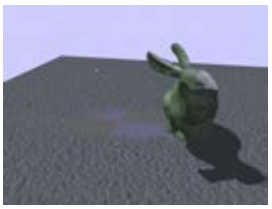
The effect of snowfall occlusion has been widely achieved by occlusion-based rendering techniques similar to those used in the process of projecting shadows. Current implementations which allow snow cover to be accumulated are most often performed in a manner which limits their use to completely static, un-movable scenes in real-time in which animation would lead to unrealistic discontinuity of the simulation. This abstract describes a real-time, occlusion-based accumulation technique which allows dynamic, moving scenes.

The initial stage of accumulating snow is determining which surfaces are visible from the direction of snowfall. The scene is rendered from above allowing the rendering process to depth test faces and produce a fully occluded picture. To achieve the visual effect of the randomness of snow, a simple uniform noise texture is introduced at the occlusion stage and is mapped across the viewable area. By altering the direction of the occlusion render, a differing snowfall direction can be introduced to incorporate the effects of a varying, uni-directional wind force and flake-flutter, as shown in Figure 1(c). To store snow accumulation, 2D buffers are uniquely mapped to the surface of each object in the scene. A render pass is performed containing a single quad for each pixel in the occlusion render, mapped to update the accumulation buffers, as shown in Figure 1(a) and Figure 1(b). Once snow has been accumulated onto a surface a Gaussian blur is performed on the accumulated height-maps, giving a smooth edge and gradual change to levels of snow approximating a curved falloff, as shown in Figure 1(d). During a second render pass the normal map shader compares the snow height on each given texel to its direct neighbours, generating surface normals. By generating high detail normal maps, per pixel lighting can be used to create very fine detail on smooth snow cover without the need for any additional geometry.

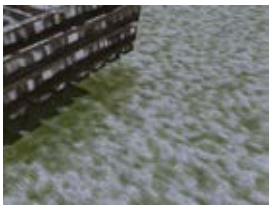
(a) Garden bench accumulating snow



(b) 'Stanford Bunny' accumulating snow



(c) Windswept snowfall occluded by 'Stanford Bunny'

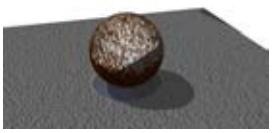


(d) Close-up of blurred snow accumulation

(e) Dynamic tessellation of 'Stanford Bunny'

(f) Tessellated surface offset by snow height

(g) Persistent accumulation under moving cart



(h) Rotating basketball loosing snow cover gradually

Figure 1: Real-Time Accumulation of Occlusion-Based Snow

For areas where added geometric detail is required to produce a realistic result, dynamic tessellation of the surface is performed. Faces are subdivided to a level dictated by a combination of their maximum snow support and their visible size, with larger faces or ones closer to the viewer requiring greater detail, as shown in Figure 1(e). The tessellated mesh is then offset by the values stored in the height-map to geometrically show rising levels in snow accumulation in addition to the visually striking colour and material difference, as shown in Figure 1(f). During the individual passes of each buffer, to allow dynamic detail, each surface is assessed to determine whether rotation has caused the surface to point downwards, and in such cases the deposited snow is removed. Removing snow is done over the course of several frames by subtracting in increments, giving the effect of snow sticking to the surface slightly in areas of high build-up. This allows the realistic simulation of rotating objects within a scene and persistent accumulation within surface buffers allows occlusion to be handled properly with moving objects above a snow supporting surface, as shown in Figure 1(g) and Figure 1(h). An important feature not possible using previous techniques.

Modelling and image processing of microtubule dynamics

Chris Rookyard, Mette Mogensen, Scott Grandison, Barry Theobald

Schools of Biological and Computing Sciences, UEA
c.rookyard@uea.ac.uk

Nearly every one of the cells that make up our bodies, and those of other animals, has its own skeleton, known as the cytoskeleton. The cytoskeleton is not like the skeletons with which we are more familiar; it is not rigid and made from bone, but flexible and made from a few different types of protein. However, it does fulfil similar roles: it provides structural rigidity for the cell, and also helps it move. I study one particular part of the cytoskeleton, known as the microtubule. Microtubules are made from polymerisation of a protein called tubulin. There are millions of tubulin units in the cell, and they can be assembled end-to-end to form filaments; it is these filaments that are microtubules. Microtubules have an interesting characteristic whereby they will change between growing (i.e. adding tubulin units) and shrinking (losing tubulin units) phases; this is known as dynamic instability. They may also pause, where no net addition or loss of tubulin units occurs. The frequencies with which microtubules transition between these phases, along with how quickly they grow and shrink, are known as dynamics parameters. Proper functionality must be realised by means of appropriate microtubule behaviour; be it through the inherent self-organisational properties of the microtubule array, differences in microtubule dynamics, number, orientation, or interactions with one another and cellular components. Furthermore, the properties of the microtubule network can be modulated by microtubule-associated proteins (MAPs) or post-translational modification (PTM). The result of these organisational processes is a cell-wide network, tuned in space and time to perform specific tasks. My research aims to elucidate the relationship(s) between microtubule dynamics and organisation, and how these change with the changing needs of the cell, for example, as cells make contacts and become polarised. I take a multi-faceted approach to these challenges, using a combination of computational modelling, microscopy and image processing in conjunction with wet lab-based experiments.

Microtubules are quite amenable to modelling since they have qualitatively well-defined states: these are the growth, shrinking and pausing phases. At the same time, the rates of growth and shrinking and the frequencies of transition between states, are yet to be resolved fully. One reason for this is that it is difficult to observe microtubules throughout the cell; the density of microtubules towards the cell centre precludes detection of individual microtubules. To circumvent this problem, I use what is essentially a reverse engineering approach to infer microtubule dynamics from their organisation. The modelling part of this method can be described as a generative model; through stochastic simulations, I produce model microtubule networks with different underlying dynamics. By considering higher-order, organisational features of these microtubule networks (defined with quantitative image processing see below), the model dynamics-organisation relationships can be defined. With prior knowledge of these relationships and similar processing of real cell images, I can estimate dynamics in real cells.

To quantify organisation of the microtubule network, I am working on creating tailored image processing algorithms to described features of the microtubule cytoskeleton. Such features include the length distribution, microtubule alignment and orientation, and microtubule flexibility (i.e., bendiness). These quantitative measures will provide a comprehensive description of microtubule organisation in a cell. One example of these algorithms is the use of the Fourier transform of microtubule images. The Fourier transform is used commonly in image processing and elsewhere. Here, I use it to determine the degree of alignment between microtubules; statistical descriptors of the power spectrum of the transform are used to quantify directional anisotropies in frequencies simultaneously indicating the degree of alignment and its length scale.

To inform the microtubule dynamics models described above, I use live cell imaging of GFP-labelled tubulin and other proteins that associate with microtubules. The aim of this approach is to measure microtubule dynamics throughout the cell; there are a number of proteins that associate just with the tip of growing microtubules, and by tracking these, we can measure microtubule dynamics. Note that, although this experiment provides information that the reverse-engineering modelling/image processing method aims to capture, it is relatively time-consuming and also interferes with the cell, thus justifying the other method.

I am using the modelling and image processing approaches just described to study microtubule re-organisation as cells make contacts and subsequently become polarised. We already know a little about the changes that accompany formation of cell-cell contacts and the role of microtubules, including transportation of N-cadherin to adherens junctions, or relocation of membrane-localised E-cadherin to the junction. Epithelial polarisation is also dependent upon relocation of proteins and other molecules, which are in turn dependent upon correct microtubule organisation. The timing of modulation of microtubule dynamics and spatial patterning is thus likely to be essential for successful polarisation. For example, recent work in the Mogensen lab has demonstrated the importance of the timing of the action of plus end-associated protein End-Binding 2 (EB2) in this process. However, we do not have a fully comprehensive picture of cell-cell contact-mediated microtubule network re-arrangements and the subsequent effects on the organisation of the cell; I aim to provide this by quantifying local and global changes in the microtubule network as cells make contacts and polarise, complementing this with mechanistic models to understand how and why these changes occur.

Active Learning Models for Machine Learning

Awat Saeed, Gavin Cawley, and Tony Bagnall

Comptational Biology Laboratory
{Awat.Saeed,G.Cawley,Anthony.Bagnall}@uea.ac.uk

Supervised machine learning methods, it must often be trained on hundreds (even thousands) of labeled instances to get good performance. However, often we have a data set with few labeled instances, and a large number of unlabeled ones. Regularly, for more sophisticated machine learning applications unlabelled data are available at little or no cost, but to obtain a small amount of labelled data, and labeling unlabelled instances is expensive, very difficult, and time-consuming [1]. In addition, so far many machine learning and data mining researchers have focused data collected to analysts, which are not always of good quality. These reasons have given rise of the applied active learning and it has used by researchers to close the gap between data achievement and model building[2, 3].
Active learning (in the statistics literature called optimal experimental design) is a method in machine learning and, more generally, artificial intelligence. It attempts to achieve better performance using as few labeled instances as possible by direct the queries in the form of unlabeled instances to be labeled by an oracle (e.g., a human annotator). Thus a good model can be trained with a relatively small number of queries, minimizing the cost of obtaining labeled data,Figure1 showed an active learning example[3].

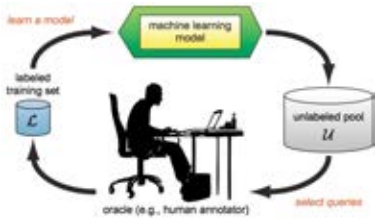


Figure 1: Active learning example(Pool-based active learning)[2].

Three main scenarios are illustrated in Figure2 which active learners may pose queries. From De novo query synthesis scenario the learner can select arbitrary unlabeled instance to request labels in the input space x use examples not drawn from $P(x)$. In the Pool-based active learning queries are selected unlabeled instance from large pool at the start of training. Finally, in Stream-based active learning scenario examples are made available continuously. There are also several different query strategies are exist to decide which instances are most informative, such as, uncertainty sampling and query-by-committee sampling[2, 1].

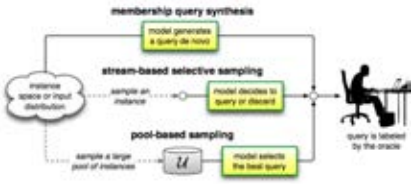


Figure 2: Three main active learning scenarios[2]

References

[1] G. C. Cawley, "Baseline methods for active learning." *Journal of Machine Learning Research-Proceedings Track*, vol. 16, pp. 47–57, 2011.

[2] B. Settles, "Active learning literature survey," *University of Wisconsin, Madison*, 2010.

[3] I. Guyon, G. Cawley, G. Dror, and V. Lemaire, "Results of the active learning challenge," 2011.

Expressive Visual Speech Synthesis

Felix Shaw and Barry-John Theobald

Speech, Language and Graphics Lab
{f.shaw, b.theobald}@uea.ac.uk

Much work has gone into creating realistic computer generated humanoid characters. The video game and movie industries spend many millions on animating characters and CGI created scenes and every year. Two main techniques exist. In one, artists model key frames and 3D software packages interpolate the intermediate frames. The manual modelling of keyframes is a time consuming process. In the other, motion-capture is used, where movements of an actor are recorded, digitised and then retargeted to a computer model. These techniques work well but provide little flexibility. Changes and new scenes must always be recaptured. Automated approaches have been proposed over recent years, which can be used to synthesise entire segments of animation parametrically. These save on the tedious process of designing key frames, or capturing mo-cap data, but often at the cost of subtlety, nuance, expression and dynamics. It is thought that the addition of expression to neutral sequences of animation will increase realism dramatically. This project is about trying to find a way of automatically projecting emotional expression into model based video sequences. Ultimately the hope is that it will be possible to automate the synthesis of complex, realistic looking video without the need for manual creation of keyframes, mo-cap or the large databases of video required for concatenative synthesis.

In a recent paper [1], we presented a method for transforming neutral visual speech sequences into realistic expressive visual speech sequences. By applying Independent Component Analysis (ICA) to visual features extracted from time aligned neutral and equivalent expressive sequences, a model that separates speech from expression can be learned. Analysing the behaviour of different speaking styles in terms of this model provides both a means for identifying the component(s) responsible for expression, and for learning the correspondence between different speaking styles. Exploiting this correspondence to transform neutral visual speech into expressive visual speech creates sequences that have the same time varying expressive dynamics as the equivalent ground-truth sequences, and an objective analysis shows that the neutral ICA parameters are shifted into the appropriate ranges for expressive visual speech.

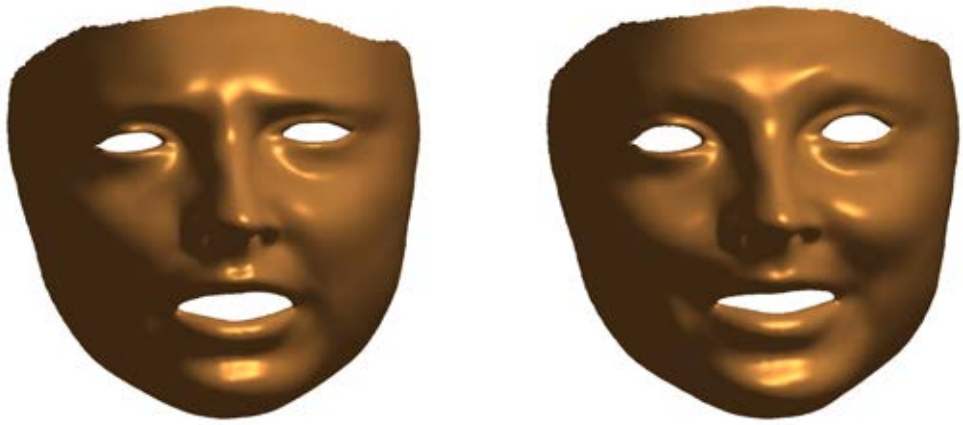


Figure 1: An example frame scaled from neutral (left) to expressive (right).

References

[1] F. Shaw and B. J. Theobald, "Transforming Neutral Visual Speech into Expressive Visual Speech," *Audio-Visual Speech Processing: . . .*, 2013.

Animation of jointly optimised audiovisual speech

Ausdang Thangthai, Barry-John Theobald and Stephen Cox

Speech, Language and Audio Processing Laboratory
{A.Thangthai,B.Theobald,S.J.Cox}@uea.ac.uk

This project will involve developing new methods for synthesising audiovisual speech. Most traditional audiovisual speech synthesis systems, first synthesise the acoustic speech, then synthesise the accompanying video sequence. The major drawback of separating the two systems is audiovisual speech lacks synchrony. The combination of two synthesised sequences that originate from different phonemes can cause the McGurk effect, which degrades intelligibility. Recently, there has been interest in synthesising both modalities jointly to propose the highest possible coherence between the audio and the visual modelling. It is believed that the joint model can improve the perceived quality of the synthesised audiovisual speech. Our preliminary work in audiovisual speech synthesis found that the output quality of a synthesiser when trained using the LIPS2008 audiovisual speech database [Theobald et al., 2008] is not acceptable. Additionally, [Mattheyses et al., 2013] reported that the LIPS2008 corpus does not contain enough data to analyse phoneme-to-viseme mapping techniques. Hence, this work will design a new database suitable for both acoustic and visual speech synthesis. Both objective tests (correlation and normalized root mean square error (RMSE)) and subjective test (mean opinion scores (MOS)) will be used to evaluate the performance.

References

[Mattheyses et al., 2013] Mattheyses, W., Latacz, L., and Verhelst, W. (2013). Comprehensive many-to-many phoneme-to-viseme mapping and its application for concatenative visual speech synthesis. *Speech Communication*, 55(78):857 – 876.

[Theobald et al., 2008] Theobald, B., Fagel, S., Elsei, F., and Bailly, G. (2008). Lips2008: Visual speech synthesis challenge. In *Proceedings of Interspeech*, pages 1875–1878.

Making the Invisible Visible

Roshanak Zakizadeh and Graham Finlayson

The Colour Group, Vision Lab, School of Computing Sciences
r.zakizadeh@uea.ac.uk

Silicon-based digital camera sensors show significant sensitivity beyond the visible spectrum and they capture wavelengths up to 1100 nm. Usually the information beyond the visible range is considered as noise and blocked by an IR filter (hot mirror). On the other hand, the NIR images appear sharper and contain more details in distant objects and could be useful to enhance images in some applications such as haze removal, realistic skin smoothing, etc.
Different techniques have been proposed to use information from both visible and near-IR parts of spectrum to enhance digital images and combine this information as a four-channel image that covers a spectrum of range 400-1100 nm. The NIR is considered as lightness or spatial information[1].
An example of such experiments can be observed in the figure 1 [2]. In this technique the lightness channel V in HSV colour space is replaced with the near-IR channel. Therefore the image is a combination of visible colours and near-IR intensity and spatial information. In the top row: on the left is the V channel of the visible image and on the right is the near-IR image. In the Bottom row: On the left, the original visible image and on the right, the coloured NIR image can be seen. As it can be observed the blue colour of the sky is lost in the coloured NIR image and its almost achromatic.

Figure 1: An example of fusion of visible colours (RGB image) and NIR image into one coloured NIR image.

Using the NIR information, we would like to make the details which are invisible to the naked eye, visible. The current techniques introduce artifacts into the images and some information might be lost.
In this project, we will look for methods to render all the spectral information, both visible and beyond visible in a single true colour image without artifacts. We aim to augment the visible colour image which is generally available with the contrast which is invisible. The information beyond the visible range of spectrum is sometimes represented in an image of lower resolution or with fewer samples available. Therefore, we would like our method to be able to produce the augmented colour image with the few information available in the NIR range. We will also study the possible applications of the NIR images which havent been studied yet.

References

[1] S. Süsstrunk and C. Fredembach, "Enhancing the visible with the invisible: Exploiting near-infrared to advance computational photography and computer vision," *SID International Symposium Digest*, 2010.

[2] C. Fredembach and S. Süsstrunk, "Colouring the near-infrared," *Color and Imaging Conference, Society for Imaging Science and Technology*, vol. 2008, no. 1, 2008.