Advances in Knowledge-Based Technologies

Proceedings of the
Master and PhD Seminar
Winter term 2007/08, part 1

Softwarepark Hagenberg
SCCH, Room 0/10
November 30, 2007

Software Competence Center Hagenberg
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 800
Fax +43 7236 3343 888
www.scch.at

Fuzzy Logic Laboratorium Linz
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 431
Fax +43 7236 3343 434
www.flll.jku.at

Program

9:00–10:00 Session 1 ***(Chair: Roland Richter)***

9:00 Edwin Lughofer:
On Human-Machine Interaction During On-line Image Classifier Training

9:30 Stefan Raiser:
Segmentation of binary images with clustering methods

10:00 Coffee Break

10:15–11:45 Session 2 ***(Chair: Bernhard Moser)***

10:15 Lorenzo Rosasco:
Elastic Net Regularization for Sparse and Table Recovery

10:45 Maarten Grachten:
Towards phrase structure reconstruction from expressive music performance data

11:15 Jan Petr:
Parallel Magnetic Resonance Imaging Reconstruction

On Human–Machine Interaction During On-line Image Classifier Training

Edwin Lughofer, James Smith, Davy Sannen, Christian Eitzinger

Abstract. This paper considers on a number of issues that arise when a trainable machine vision system learns directly from humans, rather than from a "cleaned" data set, i.e. data which is perfectly labelled with complete accuracy. This is done within the context of a generic system for the visual surface inspection of manufactured parts, however, the issues treated are relevant not only to wider computer vision applications such as medical image screening, but also to classification more generally. Some of these issues arise from the nature of humans themselves: they will be not only internally inconsistent, but will often not be completely confident about their decisions, especially if they are making decisions rapidly. People will also often differ systematically from each other in the decisions they make. Other issues may arise from the nature of the process, which may require the machine learning to have the capacity for real-time, online adaptation in response to users' input. It may be that the users cannot always provide input to a consistent level of detail. We describe how all of these issues may be tackled within a coherent methodology. Using a range of classifiers trained on real data sets from a CD imprint production process, we will present results which show that most of these issues may actually lead to improved performance.

Key words: Image classification, on-line adaptation and evolution, resolving contradictory inputs, variable input levels, partial input confidence

1 Introduction

In many machine vision applications, such as inspection tasks for quality control, an automatic system tries to reproduce human cognitive abilities. The most efficient and flexible way to achieve this is to learn the task from a human expert [1], either by supervised data or by knowledge acquisition from the human operators in form of rule bases. Typically Machine Learning systems are trained in supervised batch mode from a set of example data items each of which has a unique label. Although there may be inconsistencies or noise within the data, these are generally considered to be random in nature, and each point is considered to be labelled with complete accuracy.

However, as Machine Learning technology moves from research laboratories to practical applications such as Machine Vision, a range of issues arise concerning how humans relate to, and interact with such systems [2] [3]. Not only does this question the feasibility, or even relevance of considering "cleaned" data sets, there is an increasing demand for systems to operate in situations where off-line batch-mode processing is not appropriate [4]. This can occur if data is hard, time-consuming or costly to obtain, or if the underlying processes change fairly rapidly, requiring re-configuration. Both of these cases lead to the need for an element of incremental on-line training [5], which prompts a renewed interest in the nature of the human interaction with adaptive ML systems [6] [7].

In this paper we focus on a number of issues relating to human-machine interaction in the context of a generic system for the visual surface inspection of manufactured parts. Section 2 describes the basic architecture of our generic system, the data sets used in this work and the experimental framework. Part 1 of the talk deals with the issues arising when the nature of the application demands real-time on-line learning after an initial batch-mode phase. Part 2 of the talk deals with the fact that different users will often differ systematically from each other, and considers how best to incorporate this diversity of information. Other issues may arise from the fact that humans cannot always work as fast as the underlying applications. For example, Part 3 considers how demand for rapid user responses may reduce the level of detail in the feedback they can produce, and suggests some alternative ways for dealing with this. In In Part 4 of the talk we consider that for a number of reasons, the operator(s) may not be completely confident in their decisions and show how a suitable change in the human-machine interface used for online labelling can be exploited to capture this information and lead to performance improvements. We end by drawing some conclusions from this work, and highlighting areas that require further research attention (Section 3).

2 Architecture and Data-sets

The whole framework is shown in Figure 1. Starting from the original image (left) a so-called "contrast image" is calculated, where the gray value of each pixel correlates to the degree of deviation from the normal appearance of the

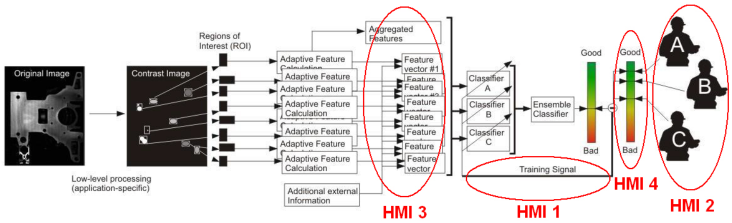

Fig. 1. Classification framework for classifying images into good and bad, the four major HMI issues marked with red ellipsoids.

surface. This contrast image just serves as an interface to the subsequent processing steps in order to remove the application-dependent elements. From the contrast image regions of interest (ROI) are extracted, each of which contains a single object which may or may not be a fault. From the segmented ROIs a large number of object features are calculated such as area, brightness, homogeneity or roundness of objects characterizing their shape, size etc. These are complemented by aggregate features characterizing images as a whole. The feature vectors are then processed by a trained classifier system that generates a final good/bad decision for the whole image. For off-line training the classifiers we exploited basically four different methods, namely: the decision tree-based classifiers *CART* [8], and *C4.5* [9]; k-Nearest Neighbours (*kNN*) [10]; and two incremental learning algorithms *eVQ-Class* [11] and *FLEXFIS-Class* [12]. When applying these classification algorithms on the standard aggregated feature sets (containing 17 pre-defined features) to real-world data from an on-line CD imprint production process, we achieved accuracies between 87% and 93% as estimated by 10-fold cross-validation. Even though the accuracies lie in a reasonable range, they fall short of the original goals for a very high-performance and robust system.

Hence, one goal of the enhanced human-machine interaction issues discussed in this paper is to guide the classifiers towards 99% accuracy. Another goal is to widen the applicability and usability of the whole system. The specific issues for human-machine interaction are highlighted in Figure 1, where the labels HMI 1-4 refer to the following issues:

1. HMI 1: the incorporation of operator's feedback on classifier(s) decisions during on-line mode, which lead to a refinement and improvement of the classifiers accuracy, especially when changing operating conditions or system behaviors arise during on-line mode. A fixed kept static classifier would be out-dated and deliver wrong classification results after a while.
2. HMI 2: feedback or labels from different experts may be contradictory, hence this should be resolved by ensembling mechanisms of different classification statements.
3. HMI 3: labels may be provided at different levels of detail (on images or single objects) according to the speed of on-line processing and according to the effort for labelling off-line.

4. HMI 4: the operator(s) may provide an additional information in form of a confidence level as not being completely confident in its (their) decision(s) due to low experience or occurring similar patterns between faulty and non-faulty regions.

All these four topics will be treated and underlined how they impact the classification performance of the whole system on both data sets mentioned above. Classification performance will be measured in terms of 10-fold CV error, except for the incremental on-line training issue, where the accuracy on a separate test set is calculated (as CV is an off-line procedure).

3 Conclusion and Outlook

As machine learning systems move out of the laboratory and into real-world applications such as vision and image processing, it is valuable to reconsider some of the assumptions that have been made about how such systems can best learn from users. In this paper we have discussed some of the more important of these issues, and suggested how they might be handled. Experiments conducted with 'real' data within the context of a generic image processing system show that when properly handled, the human factors can represent an additional form of information to these systems for improving performance and may widen the applicability and usability, rather than to be a disagreeable source of noise. Key issues of these factors include on-line guidance and feedback, a diversity of user skills, uncertainties as well as different levels of know-how and detail in users' input. The improvements are made possible by recent advances in the speed with which graphical user interfaces can operate. The next generation of user-interaction devices offers the potential to build on this research, creating much richer human-machine learning interaction.

Acknowledgements

This work was supported by the European Commission (project Contract No. STRP016429, acronym DynaVis). This publication reflects only the authors' views.

References

1. Castillo, E., Alvarez, E.: Expert Systems: Uncertainty and Learning. Springer Verlag New York Inc., New York, USA (2007)
2. Jaimes, A., Sebe, N.: Multimodal human computer interaction: A survey. In: Proceedings of the International Workshop on Human-Computer Interaction, HCI/ICCV 2005, Springer Verlag (2005) 1–15
3. Cipolla, R., (editors), A.P.: Computer Vision for Human-Machine Interaction. Cambridge University Press, Cambridge, UK (1998)

4. Gayubo, F., Gonzalez, J., Fuente, E.D.L., Miguel, F., Peran, J.: On-line machine vision system for detect split defects in sheet-metal forming processes. In: Proc. of the 18th International Conference on Pattern Recognition, IEEE Comput. Soc., Los Alamitos, CA, USA (2006)
5. Kasabov, N.: Evolving Connectionist Models - Methods and Applications in Bioinformatics, Brain Study and Intelligent Machines. Springer Verlag, London (2002)
6. Caleb-Solly, P., Smith, J.: Adaptive surface inspection via interactive evolution. Image and Vision Computing **25**(7) (2007) 1058–1072
7. Bekel, H., Bax, I., Heidemann, G., Ritter, H.: Adaptive computer vision: online learning for object recognition. In: Proceedings of the 26th DAGM Symposium on Pattern Recognition, Springer-Verlag Berlin (2004) 447–454
8. Breiman, L., Friedman, J., Stone, C., Olshen, R.: Classification and Regression Trees. Chapman and Hall, Boca Raton (1993)
9. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc, U.S.A. (1993)
10. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Verlag, New York, Berlin, Heidelberg, Germany (2001)
11. Lughofer, E.: Extensions of vector quantization for incremental clustering. Pattern Recognition **41**(3) (2008) 995–1011
12. Lughofer, E., Angelov, P., Zhou, X.: Evolving single- and multi-model fuzzy classifiers with flexfis-class. In: Proceedings of FUZZ-IEEE 2007, London, UK (2007) 363–368
13. Woods, K., Kegelmeyer, W.P., Bowyer, K.: Combination of multiple classifiers using local accuracy estimates. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 405–410
14. Kuncheva, L.I.: Combining pattern classifiers: Methods and algorithms. Wiley (2004)
15. Sannen, D., Van Brussel, H., Nuttin, M.: Classifier fusion using discounted dempster-shafer combination. In: Poster Proc. of International Conferernce on Machine Learning and Data Mining (2007) 216–230
16. Caleb-Solly, P., Steuer, M.: Classification of surface defects on hot rolled steel using adaptive learning methods. In: Proc. of the IEEE Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies,vol. 1. 103–108

Technical Report
FLLL–TR–2007-01

FUZZY.LOGIC.LAB.LINZ

Segmentation of binary images with clustering methods

Stefan Raiser
Fuzzy Logic Laboratorium Linz-Hagenberg
e-mail `stefan.raiser@jku.at`

Institut für Wissensbasierte Mathematische Systeme
Fuzzy Logic Laboratorium Linz

Johannes Kepler Universität Linz
Altenberger Str. 69
A-4040 Linz
Austria

softwarepark hagenberg

Softwarepark Hagenberg
Softwarepark 21
A-4232 Hagenberg
Austria

Abstract

In this paper clustering methods are used to perform segmentation of binary images originated from an industrial inspection process. Two clustering algorithms for detecting arbitrary and non-connected shapes of "white" pixels - a hierarchical and a density-based one - are presented. Also a method for prior checking the existence of clustering structures in an image is discussed.

Keywords: image segmentation, grouping, connected components, Hopkins index, hierarchical clustering, density-based clustering

1 Introduction

In many industrial inspection systems images of a "master" part are compared with images from the ongoing production process in order to detect faulty units. The differences between these two images are located, classified and - based on the output of the classifier - a decision, whether the part has to be rejected, is made.

Successful operation of such a quality control application is based on a reliable algorithm for recognizing "suspicious" objects in the current image. In the next chapters clustering methods, originally from the field of data-mining, are evaluated, if they are feasible to fulfil this task. Artificially created image datasets and real images from a CD print process are used for this evaluation.

2 Segmentation / grouping problem and the clustering approach

Starting with a "master" and the current "test" image, the inspection system first creates a difference image by taking the absolute difference between them. In order to find all deviations a thresholding, which sets every non-zero pixel to "white", is performed. The resulting binary difference image is the starting point for the subsequent algorithms.

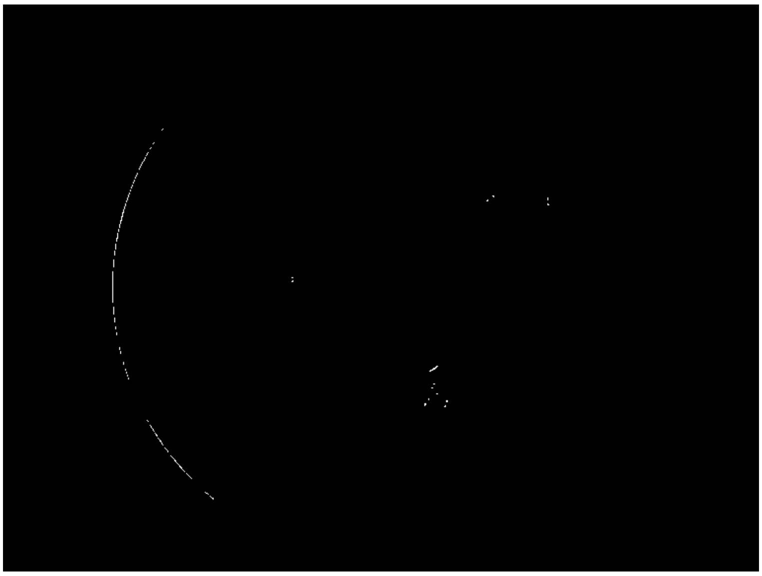

Figure 1: Binary version of a CD print difference image

The task now is to find all the objects in the binary image by grouping all "white" pixels belonging to the same object together. This segmentation process is easy as long as objects consist of connected pixels (here a connected component algorithm can do the job). But in many applications objects are kind of "widespread" (i.e. ink splashes) or discontinuous (i.e. scratches of varying depth). Also generally no information about their size, shape and number is available.

The basic idea now is to use clustering methods for the object recognition task. The coordinates of the "white" pixels represent the datapoints to be clustered. The question "Which pixels belong together / to the same object?" is now addressed by the clustering algorithm and on how it partitions the datapoints.

3 Description of three clustering methods

In the next section a measure of cluster tendency is described, followed by a review of the single linkage hierarchical algorithm plus some ideas to find out the "right" number of clusters. Finally a density-based clustering approach is investigated.

3.1 Cluster tendency - The Hopkins Index

Because most of the clustering algorithms cannot deal with less than two clusters, a measure of cluster tendency has to be calculated before trying to find groups in the data. In [1] the so called Hopkins Index is described, which can be calculated by the following algorithm:

Given a dataset

$$X = \{x_1, .., x_n\} \in \mathbb{R}^p$$

1. Choose randomly m points

$$S = \{s_1, .., s_m\} \in X$$

2. Choose randomly m points

$$R = \{r_1, .., r_m\} \in H_{convex}(X)$$

3. Calculate the distances d_{s_i} (d_{r_i}) from each s_i (r_i) to its nearest neigbour in X

4. Calculate h according to formula

$$h = \frac{\sum\limits_{i=1}^{m} d_{r_i}^p}{\sum\limits_{i=1}^{m} d_{r_i}^p + \sum\limits_{i=1}^{m} d_{s_i}^p} \in [0, 1]$$

5. Perform algorithm multiple times and take the mean of h.

If h $\approx$ 0.5 then the datapoints are likely to have a random pattern (only a single cluster).
If h $\approx$ 0 then the datapoints are located in a regular grid (as many clusters as datapoints).
If h $\approx$ 1 then a clustering structure exists (number of clusters between 2 and the number of datapoints).

In the objects recognition context the Hopkins Index can be used to detect whether there are (more than one) objects in a binary image or not. If it is smaller than a certain threshold all "white" pixels belong to the same object. Tests with an artificially created image dataset, consisting of about 20000 images (128x128px, no noise), show the following (promising) results:

```
Number of images used : 19224
Number of images with a single cluster : 1213
Number of single clusters found : 1205
Number of single clusters not found : 8
Number of single clusters overdetected : 47
```

3.2 Hierarchical clustering algorithm - Single Linkage

In [2] hierarchical clustering for object recognition is discussed in detail. For binary image segmentation it has two big advantages: the shape of the clusters can be arbitrary and - when using the single linkage distance measure - elongated clusters are properly detected ("chaining effect"). Due to its nature, a hierarchy of possible partitions is generated and a criterion has to be found in order to find the right "cutoff" of the hierarchy. To solve this problem in literature [3] cluster validation indices like the RMSSDT, SPR or RS are applied to each step of the hierarchical clustering. When they reach extremal values, the best partition of the datapoints is assumed to be found. But most of these indices (all the before mentioned) cannot handle arbitrary shaped clusters, so they are not feasible for the binary image segmentation / grouping task. Another way to find an appropriate "cutoff" is by examining the graph of the merge-distance (the distance of the closest clusters at a certain level of the hierarchy) versus the number of clusters. The point where the largest magnitude difference, the largest ratio difference or the largest second derivative occurs, is supposed to give the right hint for finding the "cutoff". In [4] the L-method, based on fitting a pair of straight lines to the merge-distance graph and minimizing some measure derived from the previous fit, is shown as a possible approach. Till now a fixed threshold for the "cutoff" is implemented. The clustering algorithm stops, if the closest clusters are more far away than the threshold. This saves computational time, but lacks of "adaptiveness" to the content of the current image.

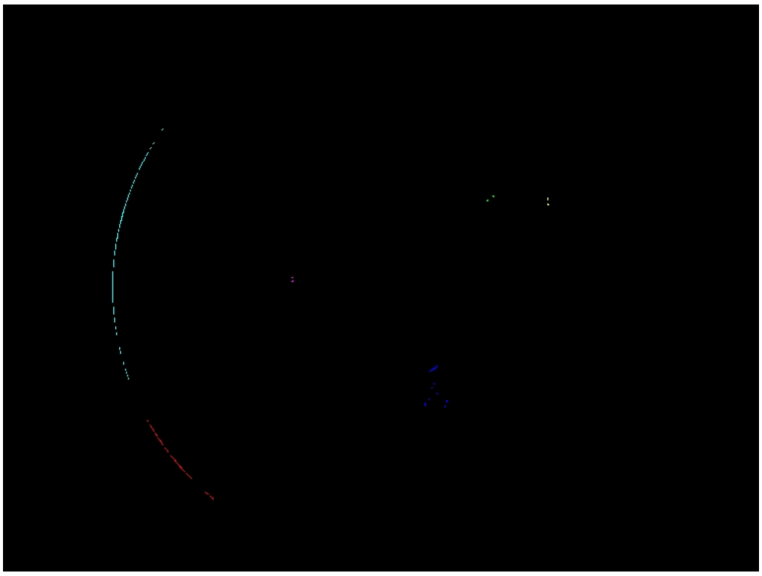

Figure 2: Hierarchical clustering on a CD print image

Tests on real images (CD prints) have shown that hierarchical clustering performs quiet well as long as the right "cutoff" is found.

3.3 Density-based clustering algorithm - DBSCAN

The basic idea of density-based clustering algorithms is that clusters are dense regions in the data space, separated by regions of lower object density (local cluster criterion). A detailed description of the density-based algorithm DBSCAN can be found in [5]. Again clusters of arbitrary shape can be identified. Two parameters, namely the maximum radius ϵ of the neighbourhood and the minimum number of points $MinPts$ in an ϵ-neighbourhood of a datapoint, have to be set by the user. It turned out that the selection of the right values for these two parameters is very tricky and only slight variations can influence the clustering result dramatically. In [5] a heuristic method using the so-called $MinPts$-distance plot is described to estimate ϵ, while $MinPts$ is set to 4 for two-dimensional data. In the implementation of [6] an estimation formula is applied to automatically set the value of ϵ ($MinPts$ is set to 4 as default).

One interesting feature of DBSCAN is that it can handle noise, i.e. detect outliers. The question arises how should outliers be treated in the context of image segmentation / grouping. At the moment they are simply ignored (with the risk to miss some important objects).

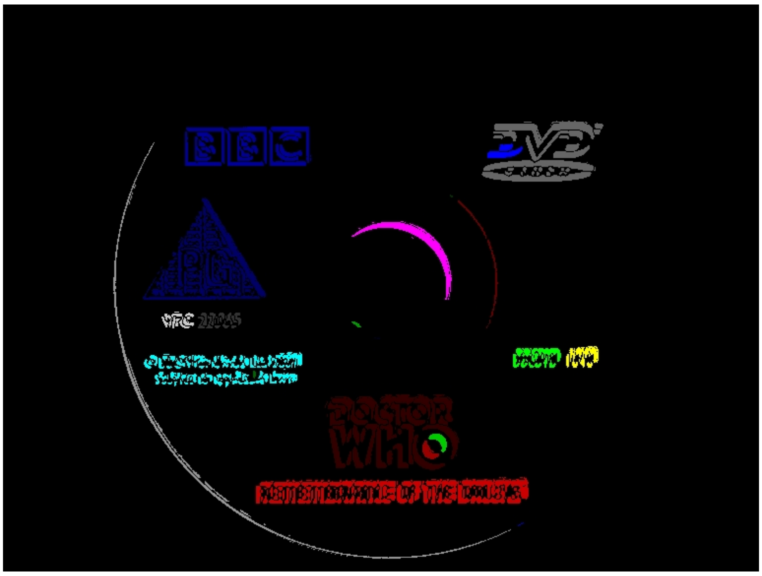

Figure 3: Density-based clustering on a CD print image

As mentioned before DBSCAN is very sensitive to the choice of its two parameters. The automatic setting via the estimation formula sometimes produces very unsatisfying results on the real CD print images but works pretty well for the artificial ones.

4 Conclusion and future work

Clustering methods seem to be a feasible solution for the binary image segmentation / grouping problem. Two major challenges in using them in an industrial inspection system are their high computational cost and the problem of automatically finding the right parameters (i.e. the number of clusters). In the future some more algorithms will be evaluated together with alternative ways of finding optimal parameter settings.

References

[1] T. A. Runkler, *Information Mining*. Vieweg, April 2000.

[2] E. Lughofer and R. Richter, "Object recognition in deviation images for fault detection – a comparison of methods," *Technical Report FLLL-TR-06-1*, 2006.

[3] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, no. 2/3, pp. 107–145, 2001.

[4] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *ICTAI 04*, 2004.

[5] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996.

[6] M. Daszykowski, "DBSCAN implementation in Matlab," Department of Chemometrics, Institute of Chemistry, University of Silesia, 2004.

Elastic Net Regularization for Sparse and Table Recovery

Lorenzo Rosasco
rosasco@disi.unige.it
DISI, Universita' di Genova
Via Dodecaneso 35, 16146 Genova, Italy

November 30, 2007

Abstract

In many applications of supervised learning a main goal, besides achieving good generalization properties, is to detect which features are meaningful to build an estimator. There at least two main difficulties in the solution of this type of problems:

1. the initial number of potentially relevant features is often much larger than the number of examples, and
2. it is often the case that many of the variables (also those that are relevant) are strongly dependent.

Both the above issues make the problem of variable selection ill-posed: in this work we explore the use of regularization theory techniques for restoring well-posedness and ensure generalization property. (Joint work with Ernesto de Vito and Christine de Mol)

Towards Phrase Structure Reconstruction from Expressive Performance Data

Maarten Grachten Gerhard Widmer
Department of Computational Perception
Johannes Kepler University, Linz, Austria
{maarten.grachten,gerhard.widmer}@jku.at

Abstract

Using a simple pattern finding approach, we investigate to what degree patterns found in the tempo and loudness curves measured from musical performances coincide with repeated musical structures in the score that was performed. We show that the high frequency content in such curves is more useful for finding repetitions of musical structures than low frequency content. In some cases removing low frequency content even improves the accuracy of pattern finding.

1 Introduction

It is commonly asserted that a primary function of musical expression is to clarify the structure of the music that is being played [Clarke, 1991, Palmer, 1997]. And indeed, countless studies of expressive music performance find that structural aspects of the musical score are in some way or another reflected in the expressive information that is extracted from performances. One such aspect is phrase structure, which is typically marked by a decrease of both tempo and loudness at phrase boundaries [Todd, 1989].

The observation that phrase structure is reflected in expressive tempo and loudness information as measured from performances, raises the question whether it would be possible to recognize the phrase structure by merely observing expressive information (and not, for example, pitch, or rhythmic information). This would form a complementary approach to studies that investigate regularities in expressive data in a score-driven way (for example [Repp, 1990]). From a practical point of view, performance-based phrase structure reconstruction could provide additional cues to systems that try to infer the structure of music pieces from, e.g., scores or MIDI files. Furthermore, applications such as score-following/automatic page turning could benefit from phrase-structure recognition.

An apparently discouraging argument against the endeavor of reconstructing phrase structure from expressive information is that a musician is by no means obliged to play repeated parts of the score in a similar way (a phenomenon termed 'consistency' in [Madsen and Widmer, 2006]). It can even be argued that playing repeated parts in different ways is one of the aspects that make human performances intriguing. In practice however, there is often considerable agreement between the performance of repeated parts [Repp, 1990].

In this paper we investigate to what degree the phrase structure of a piece is reflected in the tempo and loudness information measured from performances of the piece. We do this by measuring how well patterns found in the tempo and loudness curves coincide with the phrase structure, more specifically melodic gestures, relatively small musical constructs (typically containing less than ten notes). Rather than determining the precise beginnings and endings of phrases and melodic gestures, our goal is to identify patterns that are repeated throughout the performance. We measure accuracy in terms of how many of the instances of the pattern span repeated melodic gestures (precision), and how many repeated melodic gestures are identified as instances of the same pattern (recall). Obviously, a phrase structure reconstruction is not correct if the boundaries of the phrases/melodic gestures are not correct, but we believe that if repeated parts of the score are identified largely correctly, a useful step towards phrase structure reconstruction has been made.

From the experiments reported here it becomes clear that even the patterns found by a relatively simple pattern finding approach coincide to a considerable extent with repetitions of melodic gestures in the phrase structure. The results also indicate that the high frequency content of the tempo and loudness curves is more characteristic for melodic gestures than the low frequency content.

2 Related Work

It is undisputed that phrase structure is one of the factors that determine the expressive features of performance. Nevertheless, most of the work on automatic pattern finding and recognition of structure in music pieces has focused on score information. The pattern finding problem is often conceptually divided into a segmentation step, in which the boundaries of musical compounds are determined, and a clustering step, in which the delimited segments are grouped by identity, similarity or any other musically meaningful relation. Some approaches just deal with the segmentation problem [Temperley, 2001, Cambouropoulos, 2001], others deal with the clustering problem [Cambouropoulos and Widmer, 2000], or with both [Rolland, 1999].

The strategy we present in this paper, as said before, deals with performance information rather than score information. Furthermore, no prior segmentation of the data is used. Instead, our algorithm considers all non-overlapping pairs of equally long subsequences as possible instances of a single pattern. In this sense, our approach is related to that of [Madsen and Widmer, 2006], in which patterns found in expressive information are used to characterize the degree to which performers play repeated parts similarly.

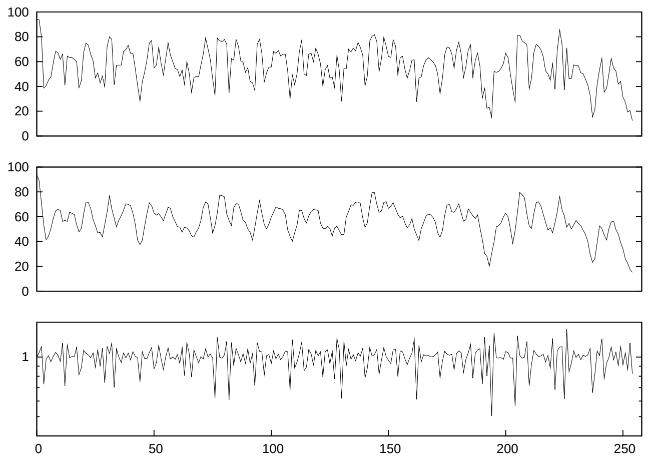

Figure 1: Decomposition of a tempo curve (from Schumann's Träumerei, performed by Horowitz, 1987) into slow and fast fluctuations. Top: Original tempo (in BPM); Middle: Smoothed Tempo (BPM); Bottom: Residual as a proportion of smoothed tempo (log scale)

3 Method

In this section we report the setup of an experiment in which we apply a pattern finding algorithm to expressive performance information in order to find repeated musical structures. We compare the results under three conditions: 1) using the original tempo and loudness curves, 2) using only the low-frequency content of the tempo and loudness curves, and 3) using only the high frequency content.

3.1 Data

The performance data used here stems from six performances of Schumann's piece "Träumerei" by renown pianists. The piece is played by Argerich, Kempff, Brendel, and Horowitz (of whom three different recordings of the piece are included). For each performance, instantaneous tempo and loudness information at half beat level is available (based on semi-automatic beat-tracking, cf. [Widmer, 2005]). Thus, the expressive performance information extracted from an audio recording is represented as a chronological sequence of pairs of tempo and loudness values, where each pair corresponds to a half beat position in the score. The total sequence consists of 255 pairs.

3.2 Decomposition of tempo and loudness curves

As stated in the introduction, typically both tempo and loudness curves convey phrase structure by a slowly evolving increase and decrease over the course of a musical phrase, roughly approximating a parabolic form. That is, the phrase is started relatively slow and soft, and after growing faster and louder towards the middle of the phrase, tempo and loudness decrease towards the end of the phrase. Although this might facilitate finding the beginnings and endings of phrases, it possibly makes distinguishing phrases more difficult, since the tempo and loudness curves of distinct phrases have their overall parabolic form in common.

Assuming that identifying distinct phrases in a piece is hindered by the parabolic component they have in common, an obvious solution is to fit a set of second order polynomials to the tempo and loudness curves on the interval of each phrase and subtract these from the original curves (as in [Tobudic and Widmer, 2003]). However, such an approach might introduce a bias towards the structure present in the score, a danger of the score-driven approach that we wish to avoid. As a simple non score-driven alternative, we apply a low-pass filter to the curves. The low-pass filtered curve contains only the lower frequency content. When this curve is subtracted from the original curve, the residual thus contains just the high frequency content. Most of the parabolic component will be contained in the low frequency curve. An example of a tempo curve and its low and high frequency components is shown in figure 1. A three point moving average filter is used as a low-pass filter both in the example and in the experiments. The peaks in the residual correspond to the sides of the parabolic forms, the points at which the original curve shows rapid changes.

3.3 Pattern Finding

We employ a simple pattern finding approach that is based on the correlation coefficient (r) between pairs of subsequences of tempo and loudness values. Tempo and loudness curves are treated in parallel, and we define the match score of a pair of subsequences as the average of r values for tempo and loudness (we will refer to this average as the r value of the match). After a subsequence length l and a threshold α for the r values have been fixed, the pattern finding algorithm returns a graph where the vertices are subsequences, and edges represent a match ($r \geq \alpha$) between two subsequences. We define the patterns to be the connected components in the graph.

Although the instances of a single pattern do not overlap (overlapping subsequences are excluded from matching), the instances of different patterns may overlap. When the instances of two patterns overlap pairwise by a constant offset, the two patterns can be seen as parts of a larger pattern that covers both. In this case, the two patterns are fused, so that each instance of the new pattern spans an overlapping pair of instances of the old patterns. Especially for lower α values, this reduces the number of patterns considerably. Note that as a result of this fusing, patterns may have different lengths (although the instances of a single pattern of course *do* have the same length), and that the size l that was chosen acts a *minimum* size, rather than a fixed size.

3.4 Evaluation

The patterns that are found are compared to the phrase analysis of the piece in terms of melodic gestures (that was adopted from [Repp, 1995]). We focus on the MG's in the soprano voice. For this voice, the piece has eight distinct MG's, most of which occur several times throughout the phrases. We evaluate repeated patterns found in the performance data by measuring how well they coincide with repetitions of the MG's.

We define the precision of a pattern as the degree of MG agreement among the instances of the pattern at each position. To this end, we define an *MGid* for each position, that is, a pair of $(\mathit{MGlabel}, \mathit{offsetIntoMG})$. For example, the MGid of a position that is the first element of an instance of $\mathit{MG2}$ would be $(\mathit{MG2}, 0)$. We define A to be the set of MGid's of all positions in the performance. The precision of a set of patterns is simply the average of the precisions per pattern, which is in turn the average of the precisions per position in the pattern. The precision at a position, finally, is the fraction of pattern instances with MGid a at that position, for the $a \in A$ that maximizes this fraction:

$$\mathit{Prec} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{L_n}\sum_{i=1}^{L_n}\max_{a\in A}\frac{|\{\ k \mid s_i^{k,n} = a\ \}|}{K_n},\ _{1\leq k\leq K_n}$$

where N is the number of patterns, L_n is the length of the n-th pattern, K_n the number of instances the n-th pattern, and $s_i^{k,n}$ is the MGid corresponding to the i-th position of instance k of pattern n.

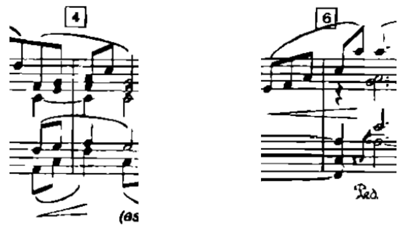

Figure 2: Two fragments from Schumann's Träumerei that were matched using tempo and loudness information

Given a set of patterns, recall is defined as the average recall over all positions. Informally speaking, the recall at a position measures the largest fraction of related positions (in terms of MGid's) that is covered by a single pattern. More precisely, let a_j be the MGid of the j-th position, and let $B_j = \{\ x \mid a_x = a_j\}$ be the set of all instances of a_j. Furthermore, let $pos(s_i^{k,n})$ denote the (global) position of the i-th element of the k-th instance of the n-th pattern. The recall is then defined as:

$$\mathit{Rec} = \frac{1}{Q}\sum_{j=1}^{Q}\max_{\substack{1\leq n\leq N,\\ 1\leq o\leq L_n}}\frac{|B_j \cap \{\ pos(s_i^{k,n}) \mid i = o\ \}|}{|B_j|},\ _{1\leq k\leq K_n}$$

where Q is the length of the sequence of tempo/loudness values. The interpretation of N, K_n, and L_n is as above.

Using the above definitions of precision and recall, we evaluate the overall accuracy of a set of patterns with the F-score $F = 2\cdot \mathit{Prec}\cdot \mathit{Rec}/(\mathit{Prec} + \mathit{Rec})$.

4 Results and Discussion

As an illustrative example, figure 2 displays two score fragments that were matched based on the tempo and loudness of the performance. Although the fragments are not instances of the same melodic gesture according to the phrase analysis, there are several interesting similarities. For example, the position in the metrical grid is the same, both fragments end in a chord, and are not immediately continued in the soprano voice, and the soprano voices in both cases are largely ascending. Also, both fragments contain a crescendo. This however was not a necessary nor a sufficient condition for the match, since other instances of the same pattern did not contain a crescendo, nor did the pattern contain all crescendos.

For each of the six recordings, we applied the pattern finding algorithm to the original tempo and loudness curves (OR), the low-frequency components (LF), and the high-frequency components (HF) respectively, using various segment sizes and r-threshold values. Figure 3 shows the F-scores (averaged over the six recordings) for each of the three curve types as a function of the r-threshold, for four different segment sizes.

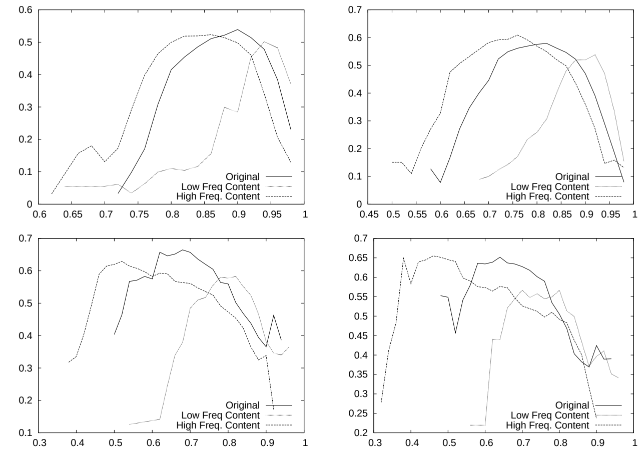

Figure 3: F-scores as a function of r-threshold, for segment sizes of 3 beats (upper left), 5 beats (upper right), 11 beats (lower left), and 15 beats (lower right)

Unsurprisingly, the F-scores for LF and OR peak at higher r-thresholds than HF (regardless of segment size). This is in accordance with our hypothesis that the low frequency components make it harder to discriminate different MG's, and thus need a higher r-threshold. A more interesting result is that pattern finding on LF gives systematically lower performance than on the others, implying that the high frequency components of the tempo and loudness curves contain essential information for telling MG's apart. Moreover, pattern finding on HF only gives results that are comparable to the results for OR, and in some cases (segment size 10) even better.

5 Conclusions and Future Work

In this paper, we have described a first step towards phrase reconstruction from expressive performance data, by detecting patterns in tempo and loudness curves. Although we have not addressed the question of finding the exact boundaries of musical phrases, we have found that repetitions of musical structures can be identified with modest success.

Moreover, our experiments show that removing the low frequency content from the tempo and dynamics curves hardly decreases, and sometimes even increases the ability to find expressive patterns that coincide with musical structures. It must be noted however that, even if six different performances were used, the current experiment covers just one musical piece. Further experiments are needed to investigate to what extent the results generalize to pieces that are performed at very regular tempos.

Lastly, although the F-score that was used for evaluation is a good indicator of the accuracy of the individual patterns that are found, it does not fully describe the accuracy of a set of patterns when interpreted as a hypothetical phrase structure. For example, it does not explicitly measure redundancy between patterns, nor incorrect phrase boundaries. More elaborate evaluation will be required to address such issues.

Acknowledgments

This work is funded by the Austrian National Science Fund (FWF) under project number P19349-N15. Furthermore, we are indebted to Sebastian Flossmann for comments on draft versions of this paper.

References

[Cambouropoulos, 2001] Cambouropoulos, E. (2001). The local boundary detection model (lbdm) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference (ICMC'2001)*, Havana, Cuba.

[Cambouropoulos and Widmer, 2000] Cambouropoulos, E. and Widmer, G. (2000). Melodic clustering: Motivic analysis of schumann's träumerei. In *Proceedings of the III Journeés d'Informatique Musicale*, Bordeaux, France.

[Clarke, 1991] Clarke, E. F. (1991). Expression and communication in musical performance. In Sundberg, J., Nord, L., and Carlson, R., editors, *Music, Language, Speech and Brain*. MacMillan Academic and Professional Ltd.

[Madsen and Widmer, 2006] Madsen, S. T. and Widmer, G. (2006). Exploring pianist performance styles with evolutionary string matching. *International Journal on Artificial Intelligence Tools*, 15(4):495–513. Special Issue on Artificial Intelligence in Music and Art.

[Palmer, 1997] Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–138.

[Repp, 1990] Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, 88:622–641.

[Repp, 1995] Repp, B. H. (1995). Diversity and commonality in music performance - An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, 92(5):2546–2568.

[Rolland, 1999] Rolland, P. (1999). Discovering patterns in musical sequences. *Journal of New Music Research*, 28 (4):334–350.

[Temperley, 2001] Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, Mass.

[Tobudic and Widmer, 2003] Tobudic, A. and Widmer, G. (2003). Playing Mozart phrase by phrase. In *Proceedings of the Fifth International Conference on Case-Based Reasoning (ICCBR-03)*, number 2689 in Lecture Notes in Artificial Intelligence, pages 552–566. Springer-Verlag.

[Todd, 1989] Todd, N. (1989). A computational model of rubato. *Contemporary Music Review*, 3 (1).

[Widmer, 2005] Widmer, G. (2005). Studying a creative act with computers: Music performance studies with automated discovery methods. *Musicae Scientiae*, IX(1):11–30.

Parallel Magnetic Resonance Imaging Reconstruction

Jan Petr
Jan.Petr@scch.at
Software Competence Center Hagenberg
Softwarepark 21, A-4232 Hagenberg, Austria

November 30, 2007

Abstract

The acquisition speed of magnetic resonance imaging (MRI) is an important issue. Increasing the acquisition speed shortens the total patient examination time, it reduces motion artifacts and increases the frame rate of dynamic MRI. Parallel MRI is a way to use multiple receiver coils with distinct spatial sensitivities to increase the MRI acquisition speed. The acquisition is speeded up by undersampling the k-space in the phase-encoding direction. The resulting data loss and consequent aliasing is compensated for by the use of additional information obtained from several receiver coils.

In my talk, I summarize the state-of-the-art in parallel MRI area. We also provide the theoretical background of MRI because full understanding of the principles behind parallel MRI is needed to understand its further extension. The main contribution of this talk is introduction of a novel parallel MRI method. Our method takes advantage of the smoothness of the reconstruction transformation in space. B-spline functions are used to approximate the reconstruction transformation. This reduces the number of the reconstruction parameters and makes the method more robust especially in areas with low signal-to-noise ratio. The B-spline coefficients are estimated by minimizing the total expected reconstruction error. We compare our new method theoretically and experimentally with two commercially available methods - SENSE and GRAPPA. The experiments were performed on simulated, phantom and in-vivo images. We show that our method outperforms the SENSE and GRAPPA reconstruction methods on a considerable number of input images and reaches the same quality on the rest.