Proceeding of the 2004 American Control Conference
Boston, Massachusetts June 30 - July 2, 2004

Classifying disc defects in optical disc drives by using time-series clustering

J. van Helvoirt*, G. A. L. Leenknegt‡, M. Steinbuch*, H. J. Goossens§

*Technische Universiteit Eindhoven
Dep. Mechanical Engineering
Control Systems Technology
P.O. Box 513, 5600 MB
Eindhoven, The Netherlands

‡Philips Optical Storage
ETS Development
P.O. Box 80002, 5600 JB
Eindhoven, The Netherlands

§Philips Research East Asia
Optical Storage & Digital TV
200070, Shanghai
China

***Abstract*— Optical disc drives are subject to various disturbances and faults. A special type of fault is the so-called disc defect. In this paper we present an approach for disc defect classification. It is based on hierarchical clustering of measured signals that are affected by disc defects. The time-series are mapped into a feature space after which the feature vectors are clustered in a hierarchical fashion. Finally, signals are fitted onto the clusters to obtain single representations for each fault class. The resulting class descriptions can then be used for (on-line) classification of new disc defects. The approach is evaluated by applying it to a set of test data.**

I. Introduction

For most dynamic systems, there are external stimuli that influence the output of the system. When their influence is unwanted we call the stimuli disturbances. One of the goals of control is to achieve rejection of these disturbances.

A special type of disturbance that seriously upsets a system and endangers its proper functioning is a so-called fault. Faults are usually associated with system malfunctions or sudden external disturbances that occur randomly and infrequently in time [1]. The random nature and severe impact of faults distinguish them from other disturbances that have a continuous but more moderate influence on the system.

For optical disc drives (ODD) a special type of faults exist. Despite the relative immunity of optical discs (e.g. CD, CD-ROM, DVDROM and DVD+RW) to damage or pollution of the surface, the disc can introduce faults into the optical drive system. One can think of scratches, dirt spots and fingerprints that arise on the polycarbonate surface or the impurities that are included in the substrate layer [2], [3], [4].

All the mentioned types of damage are called disc defects according to the following definition. *Disc defects* are those features locally present on or in an optical disc, which result in different behavior of servo signals than what can be expected from the geometry of the information track and the dimensions or shape of the disc [5].

Corresponding author is J. van Helvoirt, E-mail: J.v.Helvoirt@tue.nl

Disc defects distort the optical beam that is used to read out the data. Hence disc defects can result in erroneous data. The optical beam is also used to control the position of its own focusing point. Therefore disc defects can also result in deterministic measurement errors. In general, disc defects form a significant threat to the ability of an ODD to play back discs without generating errors that are noticeable to the user. Due to the random occurrence of disc defects and their severe impact on the drive's performance, we can regard disc defects as faults.

In general faults are dealt with in three steps. The first step is detecting the occurrence of a fault. After detection it is often required to classify the nature of the fault. Finally, when the type of fault is classified, it is possible to select proper countermeasures to prevent the fault from influencing the system.

Many work has been done on fault detection [6] and classification [7], and [8], [9] specifically address fault detection in optical disc drives. Fault classification for ODDs however, has received little attention in literature. In [4] a defect classification is proposed for the quality inspection during optical disc production, and in [3] a rough distinction is made between fingerprints and scratches on an optical disc.

This paper addresses the fault classification problem for ODDs. An approach for disc defect classification is presented that combines a time-series mapping with a well-known clustering technique. The fault classes are then described by signals that are fitted onto the found clusters. The resulting class descriptions can then be used for (on-line) fault classification. The approach is evaluated through experiments on an ODD with a set of optical discs containing various defects.

The paper is organized as follows. Section II introduces the proposed classification approach. In Section III we present the results of the experimental evaluation. Finally, we summarize and discuss the main contributions in Section IV.

0-7803-8335-4/04/$17.00 ©2004 AACC

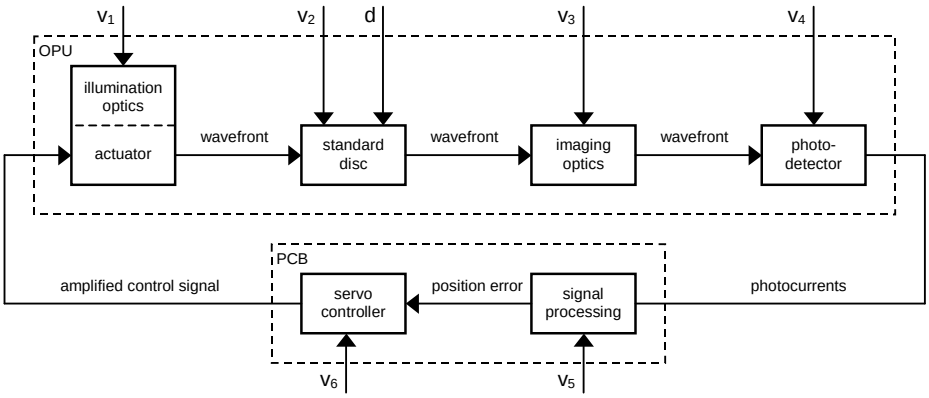


Fig. 1. Disc defects as an external stimulus (fault) for the optical drive system; OPU = optical pickup unit, PCB = printed circuit board.

II. Fault classification for ODD

In this section we introduce a fault classification approach for optical disc drives. The method is based on the clustering of measured signals that are affected by occurring faults. A novel set of features is used to describe the shape of these fault signals. The time-series are mapped into the multi-dimensional feature space, which enables us to apply a straightforward hierarchical clustering algorithm to them. Finally, we discuss how we can obtain a limited set of functions that describe the fault classes and that we can use to classify new faults.

A. Measuring fault influence

As mentioned we can regard disc defects as faults, despite the fact that they are present on the optical disc, which itself forms an integral part of the system [5]. Fig. 1 clarifies this. By introducing the concept of an ideal or standard disc we can model disc defects as faults, denoted by d, which represents the influence of the defects on the amplitude and phase of the outgoing optical wave front. The signals $v_1 \dots v_6$ in Fig. 1 represent all other disturbances that enter the system. In the remainder we assume that these disturbances are absent or sufficiently suppressed.

The presented notion of regarding disc defects as faults is not directly usable since a description of d in the optical domain is hard to obtain in practice. Furthermore, expressing the impact of the faults on the ideal optical signal is far from trivial. A way to circumvent this is by monitoring the effect of faults on a measurable electronic signal in the ODD. The effect of faults on such a signal can be regarded as a simple addition to an ideal zero-mean signal.

A signal particularly useful in our approach is the normalized mirror (MIRN) signal. This signal is proportional to the total amount of laser light received by the photodetector. Previous research shows that this signal has the most direct relation with the faults of interest [5] and it is locally constant when there are no disturbances or faults present. In the following section we propose a mapping for this signal that provides the input for the clustering algorithm.

B. Time-series mapping

Let $\mathcal{I}$ denote the set of all possible time-series of the MIRN signal from which the DC-components are removed. The subset $\mathcal{A} \subset \mathcal{I}$ is formed by all signals of length N that are distorted by the disc defects of interest only. Note that N is not fixed and depends on the physical size of the fault. Further we remark that in theory the size of the set $\mathcal{A}$ can be infinite, but in practice it will be limited in size. The label i is used to uniquely identify the corresponding element $\mathbf{y}_i \in \mathcal{A}$. Now let $p_1, p_2, \dots, p_m$ span a space $\mathcal{B}$ in $\mathbb{R}^m$, where p_j with $j = 1, 2, \dots, m$ are m different descriptive features for the elements in $\mathcal{A}$. Then $\mathcal{F} : \mathcal{A} \subset \mathcal{I} \rightarrow \mathcal{B}$ can be defined that maps the signals of interest (domain) to a set of descriptive signal features or attributes (range). This mapping is further specified by the functions $f_j : \mathcal{A} \subset \mathcal{I} \rightarrow p_j$ that are chosen as

$$f_1(\mathbf{y}_i) = W_1 \cdot \frac{1}{N} \sum_{k=1}^{N} y_i(k) \quad (1)$$

$$f_2(\mathbf{y}_i) = W_2 \cdot N \quad (2)$$

$$f_3(\mathbf{y}_i) = W_3 \cdot \max |y_i(k)| \quad (3)$$

$$f_4(\mathbf{y}_i) = W_4 \cdot \sum_{k=1}^{N} \begin{cases} 1, & |y_i(k)| < \Delta y \\ 0, & |y_i(k)| \geq \Delta y \end{cases} \quad (4)$$

$$\vdots$$

$$f_m(\mathbf{y}_i) = W_m \cdot \sum_{k=1}^{N} \begin{cases} 1, & |y_i(k)| \geq y_{max} \\ 0, & |y_i(k)| < y_{max} \end{cases} \quad (5)$$

where k is the running variable representing the sample instant and $W_1, W_2, \dots, W_m$ are the weighting factors that can be used to tune the signal mapping. See also Fig. 2.

The feature f_1 is the *mean value* of the signal. This feature is particularly useful to distinguish between faults that result in a positive and negative signal level. The features f_2 and f_3 are the *duration*, expressed through the number of samples, and the *peak value* of the absolute signal respectively. They describe the width and depth of the distorted signal. In order to describe the shape of the signal

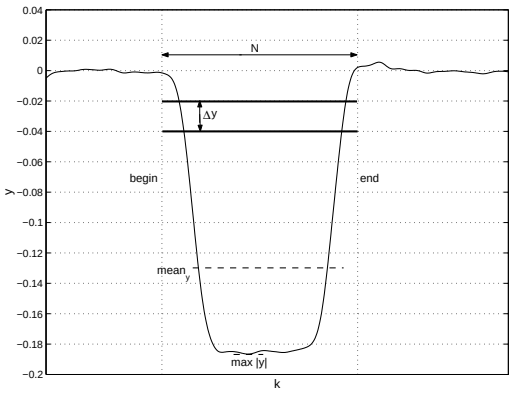


Fig. 2. Descriptive features for time-series. The vertical dashed lines mark the begin and end of the disturbed signal part; $\Delta y = 0.02$, $y_{max} = 0.2$, hence the number of amplitude bands is equal to 11 in this case.

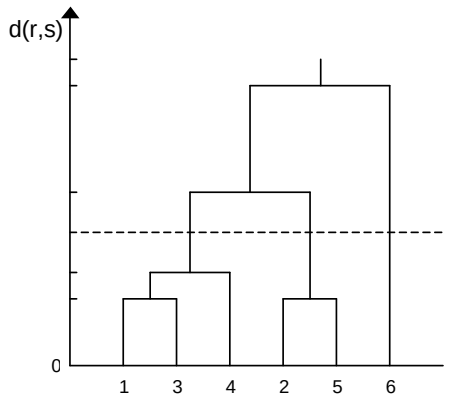


Fig. 3. Graphical representation (dendrogram) of a hierarchical cluster tree.

in more detail, we divide the absolute signal into a fixed number of *amplitude bands*. The features f_4 to f_m now represent the number of samples that fall within each band. The size of each amplitude band is given by Δy and the number of amplitude bands is limited by an upper amplitude limit y_{max}.

Finally, we remark that the set of features is heuristic and not scale invariant. By choosing appropriate weighting factors we can obtain feature values of the same order of magnitude. The set of features forms the input for the hierarchical clustering algorithm. In the following section we discuss this algorithm.

C. Hierarchical clustering

For each element in $\mathcal{A}$ we can write the feature set as a single vector. These vectors can be interpreted as points in $\mathcal{B}$, representing the corresponding MIRN signals in $\mathbb{R}^m$. In general, clustering or unsupervised classification is the process of distinguishing separate groups of similar data. With respect to the signal mapping result, clustering can be seen as the identification of different groups of closely spaced data points.

A clustering algorithm that directly uses this geometric interpretation of similarity is agglomerative hierarchical clustering [7]. The input for the clustering algorithm is a dissimilarity entity-to-entity matrix D. This matrix can easily be derived from the mapped data points (singletons) by calculating the Euclidean distance between every pair of elements in the data set. This distance is usually denoted as $\|\mathbf{p}_s - \mathbf{p}_r\|$ where r and s are the labels of the corresponding vectors. Note that D is symmetric and the elements of its main diagonal are zero.

With D available the actual clustering goes as follows. First the two elements with the smallest distance are combined into a cluster. The two rows and columns for the combined points in D are then replaced by a single row and column, which correspond to the newly formed cluster. The values on this new row and column in D express the inter-cluster distance or dissimilarity between the new cluster and all the other points and/or clusters. This dissimilarity is expressed by

$$d(r,\, s) = l_r l_s \frac{d_c^2(r,\, s)}{l_r + l_s} \tag{6}$$

where l denotes the number of elements in a cluster and $d_c^2(r,\, s)$ is the squared Centroid distance between clusters (or points) $\mathcal{S}_r$ and $\mathcal{S}_s$ according to

$$d_c(r,\, s) = \|\overline{\mathbf{p}}_s - \overline{\mathbf{p}}_r\| \tag{7}$$

with $\overline{\mathbf{p}}_r$ and $\overline{\mathbf{p}}_s$ given by

$$\overline{\mathbf{p}}_r = \frac{1}{l_r} \sum_{i=1}^{l_r} \mathbf{p}_{ri} \tag{8}$$

and $\overline{\mathbf{p}}_s$ is defined similarly. These steps are then repeated until only two clusters remain.

The results of the clustering algorithm can be represented by a convenient and compact graphical tool called dendrogram. See Fig. 3 for an example. The numbers at the horizontal axis are the labels of the original points and they are called leaf nodes. The connecting horizontal lines or interior nodes, indicate the merging of two particular clusters and their heights indicate the dissimilarity between the combined clusters.

From such a dendrogram we can select an arbitrary number of clusters by drawing a horizontal line. All the elements—identified by their label i—that are connected below this line, belong to one particular cluster. The selection of clusters hence implies a trade-off between the number of clusters and their size and mutual distinctiveness.

D. Classification

With the clusters of grouped signals available, a final step remains. The clusters themselves are nothing more than distinctive groups of similar signals. In order to obtain a fault classification, each of these groups must be replaced by a single cluster representation. This results in a limited set of representations that preserve information on all elements of the data set, without the need to actually store them all individually. In the first place these representations

attempt to provide an unambiguous class description. Secondly they lead to a desirable—especially for an on-line implementation—reduction of the amount of data.

The most suitable class description is that of a time-series or model from which such a time-series can be derived. Not only is it easy to compare graphs of different signals qualitatively, but a time-series can also be used directly in mathematical computations. A class description in words for example, lacks this possibility.

When such class descriptions are available, the classification of new faults boils down to a similar clustering problem as discussed above. The task is then to compare the new signal with the class descriptions and find the one that is most similar in some sense. This process is called supervised classification. The similarity can be expressed via the Euclidean distance, after the new time-series and those of the class descriptions are mapped into the presented feature space.

To obtain descriptive signals for each disc defect cluster we propose the well-known least squares (LS) polynomial fitting method. The LS method has many advantages, the most important one being that the global minimum can be found efficiently and unambiguously (no local minima other than the global ones exist). Another benefit of this method in the context of disc defect classification is the ability of the method to deal with numerous signals simultaneously.

We choose polynomials to approximate the defect signals in the first place because we lack a parametric function structure that is based on a disc defect model. This makes it also difficult to choose a suitable model identification technique. The polynomial function structure provides a simple and easily computed alternative to obtain single representations for each cluster of MIRN signals.

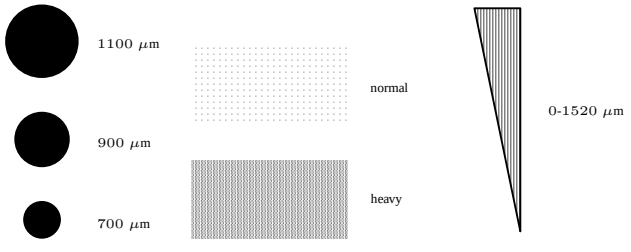
III. Evaluation of fault classification approach

We now apply the described approach to a set of MIRN signals that are affected by various faults. The goal is to evaluate whether the approach can lead to a usable disc defect classification. After the generation of the signal data base, we apply the clustering method to this data set. In order to obtain a single representation for each class, we then fit polynomials through all the signals in each cluster. Finally we classify some well-known disc defects to verify that the classification of these defects is done properly.

A. Collecting fault data

The test equipment consists of a commercially available DVD+RW drive, an additional electronic module to gain access to the drive's internal signals, a digital oscilloscope, and a personal computer that we use to control the drive and to store all measurement data.

In order to obtain a representative data base, various standardized test discs are gathered. These discs contain artificial disc defects of various sizes such as black dots that simulate dirt spots, white dots that represent areas with an abnormal high reflectivity, scratches and fingerprints. See also Fig. 4. Next to these well-defined artificial defects, several abused discs with realistic defects are added.



(a) Black dot (b) Fingerprint (c) Scratch

Fig. 4. Artificial disc defects.

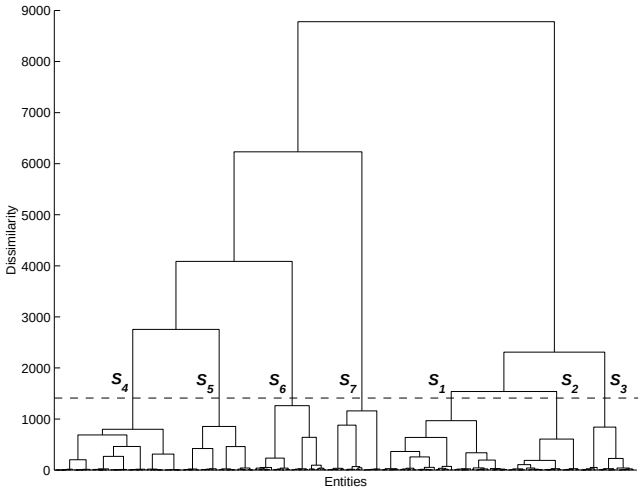


Fig. 5. Dendrogram of the disc defect clustering; Euclidean distance, Ward linkage, $W_1 = 1 \cdot 10^4$, $W_2 = 5$, $W_3, \ldots, W_m = 1$.

After measuring the MIRN signal during these faults, noise is removed via zero-phase filtering and the disturbed signal part is selected manually. Finally, the DC-offsets of the measured signals are removed.

The resulting data base consists of 320 measurements on 32 different defects. This limited data base is not large enough to provide a realistic disc defect classification. However, for the evaluation of the suggested approach it is considered to be sufficiently varied and large enough. We discuss the outcomes of the algorithm in the next section.

B. Disc defect clustering

During various trial runs the weighting parameters of the signal mapping are tuned. This is done in such a way that the clustering method makes a correct distinction between defects that are considered to be different. The final clustering results from the algorithm are depicted graphically in Fig. 5.

Table I summarizes which disc defects are grouped together in the different clusters that are selected in the dendrogram. The obtained clustering result is according

TABLE I

ELEMENTS IN DISC DEFECT CLUSTERS

Cluster	Elements
$\mathcal{S}_1$	black dots 700–900 μm[a] scratches 420–820 μm
$\mathcal{S}_2$	black dots 1100 μm[a] scratches 920–1120 μm
$\mathcal{S}_3$	scratches 1320–1520 μm
$\mathcal{S}_4$	black dots 700–1100 μm[b] realistic scratches[c]
$\mathcal{S}_5$	scratches 320 μm realistic scratches[d]
$\mathcal{S}_6$	white dots
$\mathcal{S}_7$	fingerprints

[a]Measured at the center of the black dots.
[b]Measured at the edge and at one quarter of the black dots.
[c]Measured at the smallest realistic scratch.
[d]Measured at all other realistic scratches.

TABLE II

COMBINING DISC DEFECT CLUSTERS

Number of clusters	Combined clusters
6	$\mathcal{S}_1 \cup \mathcal{S}_2$
5	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$
4	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ $\mathcal{S}_4 \cup \mathcal{S}_5$
3	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ $\mathcal{S}_4 \cup \mathcal{S}_5 \cup \mathcal{S}_6$
2	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ $\mathcal{S}_4 \cup \mathcal{S}_5 \cup \mathcal{S}_6 \cup \mathcal{S}_7$

to our expectations, based on knowledge of the different defects and the resulting MIRN signals.

As mentioned, we can alter the number of clusters by shifting the bisecting line up or down in the dendrogram. Table II shows which clusters are combined when a smaller number of clusters is selected or in other words, when the bisecting line is placed higher.

Both from Table II and the dendrogram in Fig. 5 the existing hierarchy in the data set becomes clear. Actually two major clusters can be identified. One with all centered black dots and artificial scratches and the other holding all the other defects. In both these large groups a further subdivision of disc defect types can be made.

In the remainder we consider the results for a selection of six clusters. The derivation of single class descriptions for these clusters of MIRN signals will be discussed next.

C. Classification

In order to accurately approximate the shape of the MIRN signals, a fifteenth order polynomial is fitted onto each cluster. The clusters and the resulting polynomials are

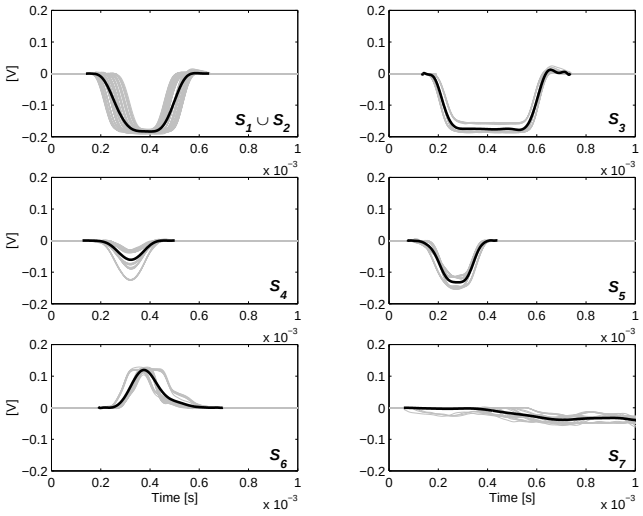


Fig. 6. Fitted 15th order polynomials for clustered disc defect signals.

shown in Fig. 6. We conclude that the signals, obtained from the fitted polynomials, describe the disc defect classes reasonably well. However, due to the flexibility of the six fitted polynomials, we observe some small oscillations in the resulting signals that are not present in the original signals. Especially at the start and end of the defect signal these deviations can become significant. This is undesirable, since it makes the fast and correct classification of new faults more difficult. Applying a more accurate fitting could resolve this problem.

D. Evaluation

With the descriptions for the disc defect classes available, we now evaluate the proposed supervised classification for new defects. This is done by applying the mapping to the fitted class signals and doing the same for some new defect measurements on several of our test discs. We then calculate the Euclidean distance between the feature vectors of the test measurements and those of the six defect classes. The class to which a test measurement belongs, should yield the smallest distance value.

An extra check is performed by calculating the correlation between the test signal and the fitted class signals. The correct class signal should result in the highest correlation coefficient.

The signals that we use for testing are shown in Fig. 7 and the resulting distances and correlation coefficients are summarized in Tables III and IV. By comparing the results for our test signals with Table I, we can conclude that the method assigns the test defects to the correct classes. Again we note that these tests are only for evaluation purposes. We want to know whether the approach could be usable for (on-line) classification of disc defects. However a much larger data base and extensive testing would be required, before a classification algorithm could be implemented in a commercial ODD.

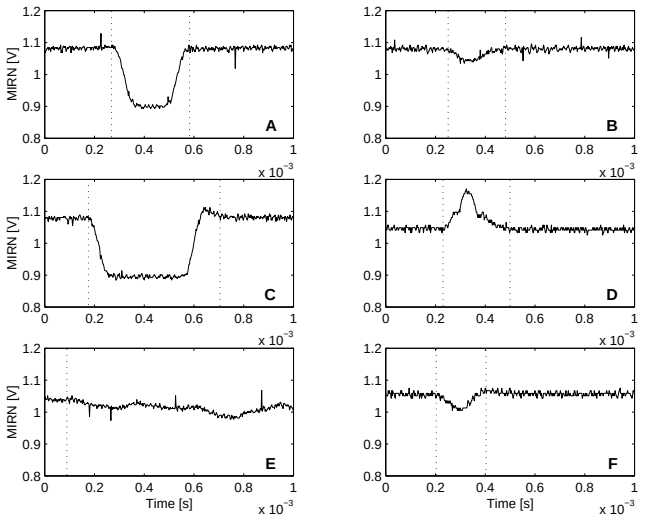


Fig. 7. Various test measurements used for class validation; A: center of 700 μm black dot, B: edge of 1100 μm black dot, C: artificial scratch 1420 μm, D: center of white dot 500 μm, E: realistic fingerprint, F: small realistic scratch.

TABLE III

EUCLIDEAN DISTANCE BETWEEN MAPPED CLASS AND TEST SIGNALS

Class	Test A	Test B	Test C	Test D	Test E	Test F
$\mathcal{S}_1 \cup \mathcal{S}_2$	**370**	906	373	1280	1214	916
$\mathcal{S}_3$	604	1250	**214**	1592	1164	1259
$\mathcal{S}_4$	839	**258**	1108	548	1363	**244**
$\mathcal{S}_5$	503	439	842	877	1410	452
$\mathcal{S}_6$	1400	759	1509	**445**	1195	731
$\mathcal{S}_7$	1721	1732	1388	1689	**179**	1722

IV. CONCLUSION

In this paper we proposed an approach for classifying disc defects in ODDs. The classification is based on the mapping of MIRN signals onto a heuristic feature space. This mapping of the MIRN signals makes it possible to use a well-known algorithm to cluster the various time signals. In order to reduce the amount of data needed to describe the classes, single functions are fitted onto each cluster.

The test results make it plausible that the suggested approach can provide a usable (on-line) disc defect classification. However, the clustering algorithm requires some manual tuning for which a certain insight of the differences between faults is needed. Furthermore, the used fault data base is not large enough to draw conclusions about which and how many defect classes are needed in practice.

TABLE IV

CORRELATION BETWEEN CLASS AND TEST SIGNALS

Class	Test A	Test B	Test C	Test D	Test E	Test F
$\mathcal{S}_1 \cup \mathcal{S}_2$	**0,983**	0,850	0,854	0,274	0,332	0,738
$\mathcal{S}_3$	0,768	0,605	**0,999**	0,364	0,444	0,538
$\mathcal{S}_4$	0,896	**0,998**	0,603	0,192	0,202	**0,968**
$\mathcal{S}_5$	0,945	0,983	0,654	0,208	0,277	0,934
$\mathcal{S}_6$	0,279	0,220	0,387	**0,985**	0,388	0,304
$\mathcal{S}_7$	0,627	0,521	0,863	0,494	**0,815**	0,408

Further work should therefor focus on applying the approach to a large and representative set of disc defect measurements. Improvements of the signal mapping and the fitting of accurate class descriptions also requires attention. The final goal will be to implement an on-line detection and classification algorithm that makes it possible to initiate (different) countermeasures whenever faults occur in an ODD.

ACKNOWLEDGMENT

The authors would like to thank the Emerging Technologies and Systems Laboratory of Philips Optical Storage, Eindhoven for facilitating this research.

REFERENCES

[1] W. M. Goble, *Control Systems Safety Evaluation and Reliability*, 2nd ed. Research Triangle Park, NC: Instrument Society of America, 1998.

[2] F. G. Pavuza, G. Beszedics, and H. Pichler, "Surface evaluation of compact disks by electronic means," in *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, Edmonton, Canada, 1999.

[3] E. Vidal, P. Andersen, J. Stoustrup, and T. S. Pedersen, "A study on the surface defects on a compact disk," in *Proc. IEEE Int. Conference on Control Applications (CCA'01)*, Mexico City, Mexico, 2001.

[4] D. Toth, A. Condurache, and T. Aach, "A two-stage-classifier for defect classification in optical media inspection," in *Proc. 16th International Conference on Pattern Recognition*, 2002.

[5] J. van Helvoirt, "Disc defect handling in optical disc drives," Master's thesis, Eindhoven University of Technology, Eindhoven, 2002, DCT 2002.48.

[6] R. J. Patton, P. M. Frank, and R. N. Clark, Eds., *Issues of Fault Diagnosis for Dynamic Systems*. Berlin: Springer, 2000.

[7] B. Mirkin, *Mathematical Classification and Clustering*. Dordrecht: Kluwer Academic Publishers, 1996.

[8] R. L. Tousain, J.-C. Boissy, M. L. Norg, M. Steinbuch, and O. H. Bosgra, "Suppressing non-periodically repeating disturbances in mechanical servo systems," in *Proc. 37th Conference on Decision and Control*, Tampa, USA, 1998.

[9] E. Vidal, K. G. Hansen, P. Andersen, K. B. Poulsen, J. Stoustrup, P. Andersen, and T. S. Pedersen, "Linear quadratic controller with fault detection in compact disk players," in *Proc. IEEE Int. Conference on Control Applications (CCA'01)*, Mexico City, Mexico, 2001.