Fuzzy-Pattern-Classifier Based Sensor Fusion for Machine Conditioning

Volker Lohweg and Uwe Mönks
Ostwestfalen-Lippe University of Applied Sciences, inIT – Institute Industrial IT, Lemgo
Germany

1. Introduction

Sensor and Information fusion is recently a major topic not only in traffic management, military, avionics, robotics, image processing, and e.g. medical applications, but becomes more and more important in machine diagnosis and conditioning for complex production machines and process engineering. Several approaches for multi-sensor systems exist in the literature (e.g. Hall, 2001; Bossé, 2007).
In this chapter an approach for a Fuzzy-Pattern-Classifier Sensor Fusion Model based on a general framework (e.g. Bocklisch, 1986; Eichhorn, 2000; Schlegel, 2004; Lohweg, 2004; Lohweg, 2006; Hempel, 2008; Herbst 2008; Mönks, 2009; Hempel 2010) is described. An application of the fusion method is shown for printing machines. An application on quality inspection and machine conditioning in the area of banknote production is highlighted.
The inspection of banknotes is a high labour intensive process, where traditionally every note on every sheet is inspected manually. Machines for the automatic inspection and authentication of banknotes have been on the market for the past 10 to 12 years, but recent developments in technology have enabled a new generation of detectors and machines to be developed. However, as more and more print techniques and new security features are established, total quality, security in banknote printing as well as proper machine conditions must be assured (Brown, 2004). Therefore, this factor necessitates amplification of a sensorial concept in general. Such systems can be used to enhance the stability of inspection and condition results for user convenience while improving machine reliability.
During printed product manufacturing, measures are typically taken to ensure a certain level of printing quality. This is particularly true in the field of security printing, where the quality standards, which must be reached by the end-products, i.e. banknotes, security documents and the like, are very high. Quality inspection of printed products is conventionally limited to the optical inspection of the printed product. Such optical inspection can be performed as an off-line process, i.e. after the printed product has been processed in the printing press, or, more frequently, as an in-line process, i.e. on the printing press, where the printing operation is carried out. Usually only the existence or appearance of colours and their textures are checked by an optical inspection system.

In general, those uni-modal systems have difficulties in detection of low degradation errors over time (Ross 2006; Lohweg, 2006). Experienced printing press operators may be capable of identifying degradation or deviation in the printing press behaviour, which could lead to the occurrence of printing errors, for instance characteristic noise produced by the printing press. This ability is however highly dependent on the actual experience, know-how and attentiveness of the technical personnel operating the printing press. Furthermore, the ability to detect such changes in the printing press behaviour is intrinsically dependent on personnel fluctuations, such as staff reorganisation, departure or retirement of key personnel, etc. Moreover, as this technical expertise is human-based, there is a high risk that this knowledge is lost over time. The only available remedy is to organize secure storage of the relevant technical knowledge in one form or another and appropriate training of the technical personnel.

Obviously, there is need for an improved inspection system which is not merely restricted to the optical inspection of the printed end-product, but which can take other factors into account than optical quality criteria. A general aim is to improve the known inspection techniques and propose an inspection methodology that can ensure a comprehensive quality control of the printed substrates processed by printing presses, especially printing presses which are designed to process substrates used in the course of the production of banknotes, security documents and such like.

Additionally, a second aim is to propose a method, which is suited to be implemented as an expert system designed to facilitate operation of the printing press. In this context, it is particularly desired to propose a methodology, which is implemented in an expert system adapted to predict the occurrence of printing errors and machine condition and provide an explanation of the likely cause of errors, should these occur. An adaptive learning model, for both, conditioning and inspection methods based on sensor fusion and fuzzy interpretation of data measures is presented here.

2. Data Analysis and Knowledge Generation

In this section some general ideas for sensor and information fusion are presented for clarity. The basic concept of fused information relies on the fact that the lack of information which is supplied by sensors should be completed by a fusion process. It is assumed that, for example, two sensory information sources S_1 and S_2 with different active physical principles (e.g. pressure and temperature) are connected in a certain way. Then symbolically the union of information is described as follows (Luo, 1989):

$$Perf(S_1 \cup S_2) > Perf(S_1) + Perf(S_2)\,. \tag{1}$$

The performance *Perf* of a system should be higher than the performance of the two mono-sensory systems, or at least, it should be ensured that:

$$Perf(S_1 \cup S_2) > \max\left(Perf(S_1), Perf(S_2)\right). \tag{2}$$

The fusion process incorporates performance, effectiveness and benefit. With fusion of different sources the perceptual capacity and plausibility of a combined result should be increased. It should be pointed out the above mentioned terms are not strictly defined as such. Moreover, they depend on a specific application as pointed out by Wald (Wald, 1999):

"Information fusion expresses the means and the tools for the alliance of data origination from different sources; it aims to obtain information of greater quality, the exact definition of greater quality will depend on the application."

The World Model (Luo, 1989) describes the fusion process in terms of a changing environment (cf. Fig. 1). The environment reacts on the system which controls (weighting factors A_i) a local fusion process based on different sensors S_i. On the basis of sensor models and the behaviour state of the sensors it is possible to predicate the statistical characteristics of the environment. Based on the World Model the environment stands for a general (printing) production machine. The fusion process generates in a best-case-scenario plausible and confident information which is necessary and sufficient for a stable decision.

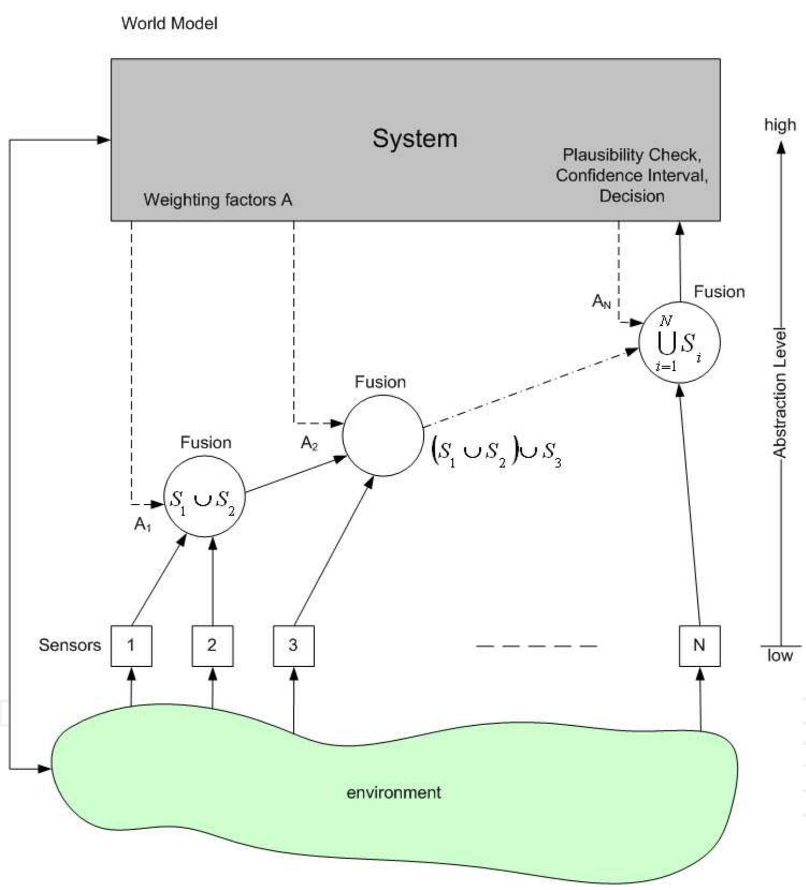


Fig. 1. World Model flow chart for multi-sensor information fusion (Luo, 1989)

2.1 Pitfalls in Sensor Fusion

In today's production world we are able to generate a huge amount of data from analogue or digital sensors, PLCs, middleware components, control PCs and if necessary from ERP systems. However, creating reliable knowledge about a machine process is a challenge because it is a known fact that $\text{Data} \neq \text{Information} \neq \text{Knowledge}$.

Insofar, a fusion process must create a low amount of data which creates reliable knowledge. Usually the main problems in sensor fusion can be described as follows: Too much data, poor models, bad features or too many features, and applications are not analysed properly. One major misbelieve is that machine diagnosis can be handled only based on the generated data – knowledge about the technical, physical, chemical, or other processes are indispensable for modeling a multi-sensor system.
Over the last decade many researchers and practitioners worked on effective multi-sensor fusion systems in many different areas. However, it has to be emphasized that some “Golden Rules” were formed which should be considered when a multi-sensor fusion system is researched and developed. One of the first who suggested rules (dirty secrets) in military applications were Hall and Steinberg (Hall, 2001a). According to their “Dirty Secrets” list, ten rules for automation systems should be mentioned here as general statements.

1. The system designers have to understand the production machine, automation system, etc. regarding its specific behaviour. Furthermore, the physical, chemical, biological and other effects must be conceived in detail.
2. Before designing a fusion system, the technical data in a machine must be measured to clarify which kind of sensor must be applied.
3. A human expert who can interpret measurement results is a must.
4. There is no substitute for an excellent or at least a good sensor. No amount of data from a not understood or not reliable data source can substitute a single accurate sensor that measures the effect that is to be observed.
5. Upstream sins still cannot be absolved by downstream processing. Data fusion processing cannot correct for errors in the pre-processing (or a wrong applied sensor) of individual data. “Soft” sensors are only useful if the data is known as reliable.
6. Not only may the fused result be worse than the best sensor – but failure to address pedigree, information overload, and uncertainty may show a worst result.
7. There is no such thing as a magic fusion algorithm. Despite claims of the contrary, no algorithm is optimal under all conditions. Even with the use of agent systems, ontologies, Dempster-Shafer and neuro-fuzzy approaches – just to name a few – the perfect algorithm is not invented yet. At the very end the application decides which algorithms are necessary.
8. The data are never perfectly de-correlated. Sources are in most cases statistically dependent.
9. There will never be enough training data available in a production machine. Therefore, hybrid methods based on models and training data should be used to apply Machine Learning and Pattern Recognition.
10. Data fusion is not a static process. Fusion algorithms must be designed so that the time aspect has to be considered.

2.2 Single-sensor vs. Multi-sensor Systems

Many detection systems are based on one main sensory apparatus. They rely on the evidence of a single source of information (e.g. photo-diode scanners in vending machines, greyscale-cameras in inspection systems, etc.). These systems, called unimodal systems, have to contend with a variety of general difficulties and have usually high false error rates in classification. The problems can be listed as follows; we refer to (Ross, 2006):

1. *Raw data noise*: Noisy data results from not sufficiently mounted or improperly maintained sensors. Also illumination units which are not properly maintained can cause trouble. Also, in general, machine drives and motors can couple different kinds of noise into the system.
2. *Intraclass variations*: These variations are typically caused by changing the sensory units in a maintenance process or by ageing of illuminations and sensors over a period of time.
3. *Interclass variations*: In a system which has to handle a variety of different production states over a period of time, there may be interclass similarities in the feature space of multiple flaws.
4. *Nonuniversality*: A system may not be able to create expedient and stable data or features from a subset of produced material.

Some of the above mentioned limitations can be overcome by including multiple information sources. Such systems, known as multimodal systems, are expected to be more reliable, due to the presence of multiple, partly signal-decorrelated, sensors. They address the problems of nonuniversality, and in combination with meaningful interconnection of signals (fusion), the problem of interclass variations. At least, they can inform the user about problems with intraclass variations and noise.

A generic multi-sensor system consists of four important units: a) the sensor unit which captures raw data from different measurement modules resp. sensors; b) the feature extraction unit which extracts an appropriate feature set as a representation for the machine to be checked; c) the classification unit which compares the actual data with their corresponding machine data stored in a database; d) the decision unit which uses the classification results to determine whether the obtained results represent e.g. a good printed or valid banknote. In multimodal systems information fusion can occur in any of the units. Generally three fusion types, depending on the abstraction level, are possible. The higher the abstraction level, the more efficient is the fusion. However, the high abstraction level fusion is not necessarily more effective due to the fact that data reduction methods are used. Therefore, information loss will occur (Beyerer, 2006).

1. *Signal level fusion – Sensor Association Principle*. At signal level all sensor signals are combined. It is necessary that the signals are comparable in a sense of data amount resp. sampling rate (adaption), registration, and time synchronisation.
2. *Feature level fusion – Feature Association Principle*. At feature level all signal descriptors (features) are combined. This is necessary if the signals are not comparable or complementary in a sense of data amount resp. sampling rate (adaption), registration, and time synchronisation. Usually this is the case if images and 1D-sensors are in use. There is no spatio-temporal coherence between the sensor signals.
3. *Symbol level fusion – Symbol Association Principle*. At symbol level all classification results are combined. In this case the reasoning (the decision) is based e.g. on probability or fuzzy membership functions (possibility functions). This is necessary if the signals are not comparable or complementary in a sense of data amount resp. sampling rate (adaption), registration, synchronisation and expert's know-how has to be considered.

Table 1 summarises the above mentioned fusion association principles.

It is stated (Ross, 2006) that generic multimodal sensor systems which integrate information by fusion at an early processing stage are usually more efficient than those systems which perform fusion at a later stage. Since input signals or features contain more information about the physical data than score values at the output of classifiers, fusion at signal or feature level is expected to provide better results. In general, fusion at feature level is critical under practical considerations, because the dimensionality of different feature sets may not be compatible. Therefore, the classifiers have the task to adapt the different dimensionalities onto a common feature space. Fusion in the decision unit is considered to be rigid, due to the availability of limited information and dimensionality.

Fusion Level	*Signal Level*	*Feature Level*	*Symbol Level*
Type of Fusion	Signals, Measurement Data	Signal Descriptors, Numerical Features	Symbols, Objects, Classes, Decisions
Objectives	Signal and Parameter Estimation	Feature Estimation, Descriptor Estimation	Classification, Pattern Recognition
Abstraction Level	low	middle	high
Applicable Data Models	Random Variables, Random Processes	Feature Vectors, Random Variable Vectors	Probability Distributions, Membership Functions
Fusion Conditions (spatio-temporal)	Registration / Synchronisation (*Alignment*)	Feature Allocation (*Association*)	Symbol Allocation (*Association*)
Complexity	high	middle	low

Table 1. Fusion levels and their allocation methods (Beyerer, 2006)

3. General Approach for Security Printing Machines

Under practical considerations, many situations in real applications can occur where information is not precise enough. This behaviour can be divided into two parts. The first part describes the fact that the information itself is uncertain. In general, the rules and the patterns describe a system in a vague way. This is because the system behaviour is too complex to construct an exact model, e.g. of a dynamic banknote model. The second part describes the fact that in real systems and applications many problems can occur, such as signal distortions and optical distortions. The practice shows that decisions are taken even on vague information and model imperfectness. Therefore, fuzzy methods are valuable for system analysis.

3.1 Detection Principles for Securities

In the general approach, different methods of machine conditioning and print flaw detection are combined, which can be used for vending or sorting machines as well as for printing machines.

3.1.1 Visible Light-based Optical Inspection

Analysis of the behaviour of the printing press is preferably performed by modelling characteristic behaviours of the printing press using appropriately located sensors to sense operational parameters of the functional components of the printing press which are exploited as representative parameters of the characteristic behaviours. These characteristic behaviours comprise of:

1. faulty or abnormal behaviour of the printing press, which leads to or is likely to lead to the occurrence of printing errors; and/or
2. defined behaviours (or normal behaviours) of the printing press, which leads to or is likely to lead to good printing quality.

Further, characteristic behaviours of the printing press can be modelled with a view to reduce false errors or pseudo-errors, i.e. errors that are falsely detected by the optical inspection system as mentioned above, and optimise the so-called alpha and beta errors. Alpha error is understood to be the probability to find bad sheets in a pile of good sheets, while beta error is understood to be the probability to find good printed sheets in a pile of bad printed sheets. According to (Lohweg, 2006), the use of a multi-sensor arrangement (i.e. a sensing system with multiple measurement channels) efficiently allows reducing the alpha and beta errors.

3.1.2 Detector-based Inspection

We have not exclusively used optical printing inspection methods, but also acoustical and other measurements like temperature and pressure of printing machines. For the latter cepstrum methods are implemented (Bogert, 1963). According to (Lohweg, 2006), the inherent defects of optical inspection are overcome by performing an in-line analysis of the behaviour of the printing press during the processing of the printed sheets. The monitored machine is provided with multiple sensors which are mounted on functional components of the printing press. As these sensors are intended to monitor the behaviour of the printing press during processing of the printed substrates, the sensors must be selected appropriately and be mounted on adequate functional machine components. The actual selection of sensors and location thereof depend on the configuration of the printing press, for which the behaviour is to be monitored. These will not be the same, for instance, for an intaglio printing press, an offset printing press, a vending machine or a sorting machine as the behaviours of these machines are not identical. It is not, strictly speaking, necessary to provide sensors on each and every functional component of the machine. But also the sensors must be chosen and located in such a way that sensing of operational parameters of selected functional machine components is possible. This permits a sufficient, precise and representative description of the various behaviours of the machine. Preferably, the sensors should be selected and positioned in such a way as to sense and monitor operational parameters which are virtually de-correlated. For instance, monitoring the respective rotational speeds of two cylinders which are driven by a common motor is not being very useful as the two parameters are directly linked to one another. In contrast, monitoring the current, drawn by an electric motor used as a drive and the contact pressure between two cylinders of the machine provides a better description of the behaviour of the printing press. Furthermore, the selection and location of the sensors should be made in view of the actual set of behaviour patterns one desires to monitor and of the classes of printing errors one wishes to detect. As a general rule, it is appreciated that sensors might be provided on the printing press in order to sense any combination of the following operational parameters:

1. processing speed of the printing press, i.e. the speed at which the printing press processes the printed substrates;
2. rotational speed of a cylinder or roller of the printing press;
3. current drawn by an electric motor driving cylinders of the printing unit of the printing press;

4. temperature of a cylinder or roller of the printing press;
5. pressure between two cylinders or rollers of the printing press;
6. constraints on bearings of a cylinder or roller of the printing press;
7. consumption of inks or fluids in the printing press; and/or
8. position or presence of the processed substrates in the printing press (this latter information is particularly useful in the context of printing presses comprising of several printing plates and/or printing blankets as the printing behaviour changes from one printing plate or blanket to the next).

Depending on the particular configuration of the printing press, it might be useful to monitor other operational parameters. For example, in the case of an intaglio printing press, monitoring key components of the so called wiping unit (Lohweg, 2006) has shown to be particularly useful in order to derive a representative model of the behaviour of the printing press, as many printing problems in intaglio printing presses are due to a faulty or abnormal behaviour of the wiping unit.

In general, multiple sensors are combined and mounted on a production machine. One assumption which is made in such applications is that the sensor signals should be de-correlated at least in a weak sense. Although this strategy is conclusive, the main drawback is based on the fact that even experts have only vague information about sensory cross correlation effects in machines or production systems. Furthermore, many measurements which are taken traditionally result in ineffective data simply because the measurement methods are suboptimal.

Therefore, our concept is based on a prefixed data analysis before classifying data. The classifier's learning is controlled by the data analysis results. The general concept is based on the fact that multi-sensory information can be fused with the help of a Fuzzy-Pattern-Classifier chain, which is described in section 5.

4. Fuzzy Multi-sensor Fusion

It can hardly be said that information fusion is a brand new concept. As a matter of fact, it has already been used by humans and animals intuitively. Techniques required for information fusion include various subjects, including artificial intelligence (AI), control theory, fuzzy logic, and numerical methods and so on. More areas are expected to join in along with consecutive successful applications invented both in defensive and civilian fields.

Multi-sensor fusion is the combination of sensory data or data derived from sensory data and from disparate sources such that the resulting information is in some sense better than for the case that the sources are used individually, assuming the sensors are combined in a good way. The term 'better' in that case can mean more accurate, more complete, or more reliable. The fusion procedure can be obtained from direct or indirect fusion. *Direct fusion* is the fusion of sensor data from some homogeneous sensors, such as acoustical sensors; *indirect fusion* means the fused knowledge from prior information, which could come from human inputs. As pointed out above, multi-sensor fusion serves as a very good tool to obtain better and more reliable outputs, which can facilitate industrial applications and compensate specialised industrial sub-systems to a large extent.

The primary objective of multivariate data analysis in fusion is to summarise large amounts of data by means of relatively few parameters. The underlying theme behind many

multivariate techniques is reduction of features. One of these techniques is the Principal Components Analysis (PCA), which is also known as the Karhunen-Loéve transform (KLT) (Jolliffe, 2002).
Fuzzy-Pattern-Classification in particular is an effective way to describe and classify the printing press behaviours into a limited number of classes. It typically partitions the input space (in the present instance the variables – or operational parameters – sensed by the multiple sensors provided on functional components of the printing press) into categories or pattern classes and assigns a given pattern to one of those categories. If a pattern does not fit directly within a given category, a so-called "goodness of fit" is reported. By employing fuzzy sets as pattern classes, it is possible to describe the degree to which a pattern belongs to one class or to another. By viewing each category as a fuzzy set and identifying a set of fuzzy "if-then" rules as assignment operators, a direct relationship between the fuzzy set and pattern classification is realized. Figure 2 is a schematic sketch of the architecture of a fuzzy fusion and classification system for implementing the machine behaviour analysis. The operational parameters P_1 to P_n sensed by the multi-sensor arrangement are optionally preprocessed prior to feeding into the pattern classifier. Such preprocessing may in particular include a spectral transformation of some of the signals output by the sensors. Such spectral transformation will in particular be envisaged for processing the signal's representative of vibrations or noise produced by the printing press, such as the characteristic noise or vibration patterns of intaglio printing presses.

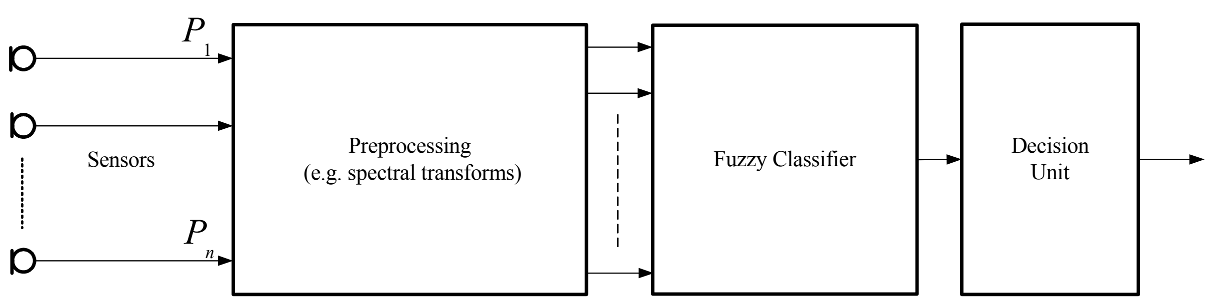


Fig. 2. Multi-sensor fusion approach based on Fuzzy-Pattern-Classifier modelling

5. Modelling by Fuzzy-Pattern-Classification

Fuzzy set theory, introduced first by Zadeh (Zadeh, 1965), is a framework which adds uncertainty as an additional feature to aggregation and classification of data. Accepting vagueness as a key idea in signal measurement and human information processing, fuzzy membership functions are a suitable basis for modelling information fusion and classification. An advantage in a fuzzy set approach is that class memberships can be trained by measured information while simultaneously expert's know-how can be taken into account (Bocklisch, 1986).
Fuzzy-Pattern-Classification techniques are used in order to implement the machine behaviour analysis. In other words, sets of fuzzy-logic rules are applied to characterize the behaviours of the printing press and model the various classes of printing errors which are likely to appear on the printing press. Once these fuzzy-logic rules have been defined, they can be applied to monitor the behaviour of the printing press and identify a possible correspondence with any machine behaviour which leads or is likely to lead to the

occurrence of printing errors. Broadly speaking, Fuzzy-Pattern-Classification is a known technique that concerns the description or classification of measurements. The idea behind Fuzzy-Pattern-Classification is to define the common features or properties among a set of patterns (in this case the various behaviours a printing press can exhibit) and classify them into different predetermined classes according to a determined classification model. Classic modelling techniques usually try to avoid vague, imprecise or uncertain descriptive rules. Fuzzy systems deliberately make use of such descriptive rules. Rather than following a binary approach wherein patterns are defined by "right" or "wrong" rules, fuzzy systems use relative "if-then" rules of the type "if *parameter alpha* is equal to (greater than, …less than) *value beta*, then *event A* always (often, sometimes, never) happens". Descriptors "always", "often", "sometimes", "never" in the above exemplary rule are typically designated as "linguistic modifiers" and are used to model the desired pattern in a sense of gradual truth (Zadeh, 1965; Bezdek, 2005). This leads to simpler, more suitable models which are easier to handle and more familiar to human thinking. In the next sections we will highlight some Fuzzy-Pattern-Classification approaches which are suitable for sensor fusion applications.

5.1 Modified-Fuzzy-Pattern-Classification

The Modified-Fuzzy-Pattern-Classifier (MFPC) is a hardware optimized derivate of Bocklisch's Fuzzy-Pattern-Classifier (FPC) (Bocklisch, 1986). It should be worth mentioning here that Hempel and Bocklisch (Hempel, 2010) showed that even non-convex classes can be modelled within the framework of Fuzzy-Pattern-Classification. The ongoing research on FPC for non-convex classes make the framework attractive for Support Vector Machine (SVM) advocates.

Inspired from Eichhorn (Eichhorn, 2000), Lohweg et al. examined both, the FPC and the MFPC, in detail (Lohweg, 2004). MFPC's general concept of simultaneously calculating a number of membership values and aggregating these can be valuably utilised in many approaches. The author's intention, which yields to the MFPC in the form of an optimized structure, was to create a pattern recognition system on a Field Programmable Gate Array (FPGA) which can be applied in high-speed industrial environments (Lohweg, 2009). As MFPC is well-suited for industrial implementations, it was already applied in many applications (Lohweg, 2006; Lohweg, 2006a; Lohweg, 2009; Mönks, 2009; Niederhöfer, 2009).

Based on membership functions $\mu(m,\mathbf{p})$, MFPC is employed as a useful approach to modelling complex systems and classifying noisy data. The originally proposed unimodal MFPC fuzzy membership function $\mu(m,\mathbf{p})$ can be described in a graph as:

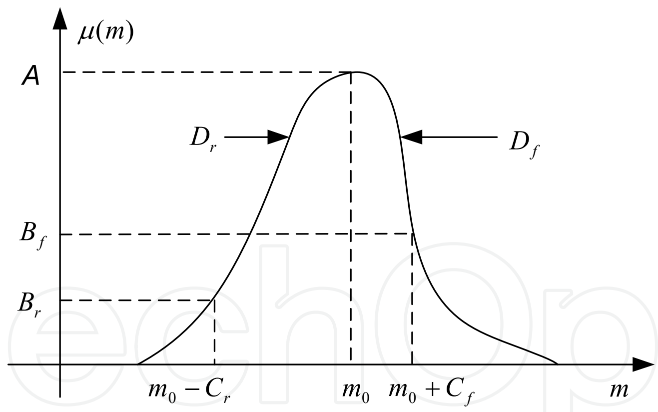


Fig. 3. Prototype of a unimodal membership function

The prototype of a one-dimensional potential function $\mu(m,\mathbf{p})$ can be expressed as follows (Eichhorn, 2000; Lohweg, 2004):

$$\mu(m,\mathbf{p}) = A \cdot 2^{-d(m,\mathbf{p})}, \tag{3}$$

with the difference measure

$$d(m,\mathbf{p}) = \begin{cases} \left(\dfrac{1}{B_r} - 1\right) \cdot \left(\dfrac{|m - m_0|}{C_r}\right)^{D_r}, \forall m < m_0 \\ \left(\dfrac{1}{B_f} - 1\right) \cdot \left(\dfrac{|m - m_0|}{C_f}\right)^{D_f}, \forall m \geq m_0 \end{cases}. \tag{4}$$

As for Fig. 3, the potential function $\mu(m,\mathbf{p})$ is a function concerning parameters A and the parameter vector $\mathbf{p}$ containing coefficients m_0, B_r, B_f, C_r, C_f, D_r, and D_f. A is denoted as the amplitude of this function, and in hardware design usually set $A = 1$. The coefficient m_0 is featured as center of gravity. The parameters B_r and B_f determine the value of the membership function on the boundaries $m_0 - C_r$ and $m_0 + C_f$ correspondingly. In addition, rising and falling edges of this function are described by $\mu(m_0 - C_r, \mathbf{p}) = B_r$ and $\mu(m_0 + C_f, \mathbf{p}) = B_f$. The distance from the center of gravity is interpreted by C_r and C_f. The parameters D_r and D_f depict the decrease in membership with the increase of the distance from the center of gravity m_0. Suppose there are M features considered, then Eq. 3 can be reformulated as:

$$\mu(\mathbf{m},\mathbf{p}) = 2^{-\frac{1}{M}\sum_{i=0}^{M-1} d_i(m_i,\mathbf{p}_i)}. \tag{5}$$

With a special definition ($A = 1$, $B_r = B_f = 0.5$, $C_r = C_f$, $D_r = D_f$) Modified-Fuzzy-Pattern Classification (Lohweg, 2004; Lohweg 2006; Lohweg 2006a) can be derived as:

$$\mu_{MFPC}(\mathbf{m},\mathbf{p}) = 2^{-\frac{1}{M}\sum_{i=0}^{M-1} d_i(m_i,\mathbf{p}_i)}, \tag{6}$$

where

$$d_i(m_i,\mathbf{p}_i) = \left(\frac{|m_i - m_{0,i}|}{C_i}\right)^D, \quad m_{0,i} = \frac{1}{2}(m_{\max_i} + m_{\min_i}), \quad C_i = (1 + 2 \cdot P_{CE}) \cdot (\frac{m_{\max_i} - m_{\min_i}}{2}). \tag{7}$$

The parameters $m_{\max}$ and $m_{\min}$ are the maximum and minimum values of a feature in the training set. The parameter m_i is the input feature which is supposed to be classified. Admittedly, the same objects should have similar feature values that are close to each other. In such a sense, the resulting value of $m_i - m_{0,i}$ ought to fall into a small interval, representing their similarity. The value P_{CE} is called elementary fuzziness ranging from zero to one and can be tuned by experts' know-how. The same implies to D = (2, 4, 8, …). The aggregation is performed by a fuzzy averaging operation with a subsequent normalization procedure.

As an instance of FPC, MFPC was addressed and successfully hardware-implemented on banknote sheet inspection machines. MFPC utilizes the concept of membership functions in fuzzy set theory and is capable of classifying different objects (data) according to their features, and the outputs of the membership functions behave as evidence for decision makers to make judgments. In industrial applications, much attention is paid on the costs and some other practical issues, thus MFPC is of great importance, particularly because of its capability to model complex systems and hardware implementability on FPGAs.

5.2 Adaptive Learning Model for Modified-Fuzzy-Pattern-Classification

In this section we present an adaptive learning model for fuzzy classification and sensor fusion, which on one hand adapts itself to varying data and on the other hand fuses sensory information to one score value. The approach is based on the following facts:

1. The sensory data are in general correlated or
2. Tend to correlate due to material changes in a machine.
3. The measurement data are *time-variant*, e.g., in a production process many parameters are varying imperceptively.
4. The definition of "good" production is always human-centric. Therefore, a committed quality standard is defined at the beginning of a production run.
5. Even if the machine parameters change in a certain range the quality could be in order.

The underlying scheme is based on membership functions (local classifiers) $\mu_i(m_i,\mathbf{p}_i)$, which are tuned by a learning (training) process. Furthermore, each membership function is weighted with an attractor value A_i, which is proportional to the eigenvalue of the corresponding feature m_i. This strategy leads to the fact that the local classifiers are trained based on committed quality and weighted by their attraction specified by a Principal Component Analysis' (PCA) (Jolliffe, 2002) eigenvalues. The aggregation is again performed by a fuzzy averaging operation with a subsequent normalization procedure.

5.2.1 Review on PCA

The Principal Components Analysis (PCA) is effective, if the amount of data is high while the feature quantity is small (< 30 features). PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data are hard to find in data of high dimensions, where the graphical representation is not available, PCA is a powerful tool for analysing data. The other main advantage of PCA is that once patterns in the data are found, it is possible to compress the data by reducing the number of dimensions without much loss of information. The main task of the PCA is to project input data into a new (sub-)space, wherein the different input signals are de-correlated. The PCA is used to find weightings of signal importance in the measurement's data set.

PCA involves a mathematical procedure which transforms a set of correlated response variables into a smaller set of uncorrelated variables called principal components. More formally it is a linear transformation which chooses a new coordinate system for the data set such that the greatest variance by any projection of the set is on the first axis, which is also called the first principal component. The second greatest variance is on the second axis, and so on. Those created principal component variables are useful for a variety of things including data screening, assumption checking and cluster verification. There are two possibilities to perform PCA: first applying PCA to a covariance matrix and second applying PCA to a correlation matrix. When variables are not normalised, it is necessary to choose the second approach: Applying PCA to raw data will lead to a false estimation, because variables with the largest variance will dominate the first principal component. Therefore in this work the second method in applying PCA to standardized data (correlation matrix) is presented (Jolliffe, 2002).

In the following the function steps of applying PCA to a correlation matrix is reviewed concisely. If there are M data vectors $\mathbf{x}_{1N}^T \dots \mathbf{x}_{MN}^T$ each of length N, the projection of the data into a subspace is executed by using the Karhunen-Loéve transform (KLT) and their inverse, defined as:

$$\mathbf{Y} = \mathbf{W}^T \cdot \mathbf{X} \text{ and } \mathbf{X} = \mathbf{W} \cdot \mathbf{Y}, \tag{8}$$

where $\mathbf{Y}$ is the output matrix, $\mathbf{W}$ is the KLT transform matrix followed by the data (input) matrix:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{pmatrix}. \tag{9}$$

Furthermore, the expectation value $E(\bullet)$ (average $\overline{\boldsymbol{x}}$) of the data vectors is necessary:

$$\overline{\mathbf{x}} = E(X) = \begin{pmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_M) \end{pmatrix} = \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_M \end{pmatrix}, \text{ where } \overline{x}_i = \frac{1}{N}\sum_{i=1}^{N} x_i . \tag{10}$$

With the help of the data covariance matrix

$$\mathbf{C} = E\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})^T\right] = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1M} \\ c_{21} & c_{22} & \cdots & c_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ c_{M1} & {}_{M2} & \cdots & c_{MM} \end{pmatrix}, \tag{11}$$

the correlation matrix $\boldsymbol{R}$ is calculated by:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & 1 \end{pmatrix}, \text{ where } \rho_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}. \tag{12}$$

The variables c_{ii} are called variances; the variables c_{ij} are called covariances of a data set. The correlation coefficients are described as ρ_{ij}. Correlation is a measure of the relation between two or more variables. Correlation coefficients can range from -1 to +1. The value of -1 represents a perfect negative correlation while a value of +1 represents a perfect positive correlation. A value of 0 represents no correlation. In the next step the eigenvalues λ_i and the eigenvectors $\mathbf{V}$ of the correlation matrix are computed by Eq. 13, where $\operatorname{diag}(\lambda)$ is the diagonal matrix of eigenvalues of **C:**

$$\operatorname{diag}(\lambda) = \mathbf{V}^{-1} \cdot \mathbf{R} \cdot \mathbf{V}. \tag{13}$$

The eigenvectors generate the KLT matrix and the eigenvalues represent the distribution of the source data's energy among each of the eigenvectors. The cumulative energy content for the *p*th eigenvector is the sum of the energy content across all of the eigenvectors from 1 through *p*. The eigenvalues have to be sorted in decreasing order:

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_M \end{pmatrix}, \text{ where } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M. \tag{14}$$

The corresponding vectors $\mathbf{v}_i$ of the matrix $\mathbf{V}$ have also to be sorted in decreasing order like the eigenvalues, where $\mathbf{v}_1$ is the first column of matrix $\mathbf{V}$, $\mathbf{v}_2$ the second and $\mathbf{v}_M$ is the last column of matrix $\mathbf{V}$. The eigenvector $\mathbf{v}_1$ corresponds to eigenvalue λ_1, eigenvector $\mathbf{v}_2$ to eigenvalue λ_2 and so forth. The matrix **W** represents a subset of the column eigenvectors as basis vectors. The subset is preferably as small as possible (two eigenvectors). The energy distribution is a good indicator for choosing the number of eigenvectors. The cumulated energy should map approx. 90 % on a low number of eigenvectors. The matrix $\mathbf{Y}$ (cf. Eq. 8) then represents the Karhunen-Loéve transformed data (KLT) of matrix $\mathbf{X}$ (Lohweg, 2006a).

5.2.2 Modified Adaptive-Fuzzy-Pattern-Classifier

The adaptive Fuzzy-Pattern-Classifier core based on the world model (Luo, 1989) consists of M local classifiers (MFPC), one for each feature. It can be defined as

$$AFPC = diag\left(\mu_i\right) = \begin{bmatrix} \mu_1\left(m_1, \mathbf{p}_1\right) & 0 & 0 & 0 \\ 0 & \mu_2\left(m_2, \mathbf{p}_2\right) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mu_M\left(m_M, \mathbf{p}_M\right) \end{bmatrix}. \tag{15}$$

The adaptive fuzzy inference system (AFIS), is then described with a length M unit vector $\mathbf{u} = \left(1, \ldots, 1\right)^T$ and the attractor vector $\mathbf{A} = \left(A_1, A_2, \ldots, A_M\right)^T$ as

$$\mu_{AFIS} = \frac{1}{\mathbf{A}^T \cdot \mathbf{u}} \mathbf{A}^T \cdot diag\left(\mu_i\right) \cdot \mathbf{u}, \tag{16}$$

which can be written as

$$\mu_{AFIS} = \frac{1}{\sum_{i=1}^{M} A_i} \cdot \sum_{i=1}^{M} A_i \cdot 2^{-d_i} . \tag{17}$$

The adaptive Fuzzy-Pattern-Classifier model output μ_{AFIS} can be interpreted as a score value in the range of $\{0 \ldots 1\}$. If $\mu_{AFIS} = 1$, a perfect match is reached, which can be assumed as a measure for a "good" system state, based on an amount of sensor signals. The score value $\mu_{AFIS} = 0$ represents the overall "bad" measure decision for a certain trained model. As it will be explained in section 6 the weight values of each parameter are taken as the weighted components of eigenvector one (PC1) times the square roots of the corresponding eigenvalues:

$$A_i = \left|v_{1i} \cdot \sqrt{\lambda_1}\right| . \tag{18}$$

With Eq. 17 the Modified-Adaptive-Fuzzy-Pattern-Classifier (MAFPC) results then in

$$\mu_{MAFPC} = \frac{1}{\sum_{i=1}^{M} \left|v_{1i} \cdot \sqrt{\lambda_1}\right|} \cdot \sum_{i=1}^{M} \left|v_{1i} \cdot \sqrt{\lambda_1}\right| \cdot 2^{-d_i} . \tag{19}$$

In section 6.1 an application with MAFPC will be highlighted.

5.3 Probabilistic Modified-Fuzzy-Pattern-Classifier

In many knowledge-based industrial applications there is a necessity to train using a small data set. It is typical that there are less than ten up to some tens of training examples. Having only such a small data set, the description of the underlying universal set, from which these examples are taken, is very vague and connected to a high degree of uncertainty. The heuristic parameterisation methods for the MFPC presented in section 5.1 leave a high degree of freedom to the user which makes it hard to find optimal parameter values. In this section we suggest an automatic method of learning the fuzzy membership functions by estimating the data set's probability distribution and deriving the function's parameters automatically from it. The resulting *Probabilistic MFPC (PMFPC)* membership function is based on the MFPC approach, but leaves only one degree of freedom leading to a shorter learning time for obtaining stable and robust classification results (Mönks, 2010).

Before obtaining the different PMFPC formulation, it is reminded that the membership functions are aggregated using a fuzzy averaging operator in the MFPC approach. Consequently, on the one hand the PMFPC membership functions can substitute the MFPC membership function. On the other hand the fuzzy averaging operator used in the MFPC can be substituted by any other operator. Actually, it is also possible to substitute both parts of the MFPC at the same time (Mönks, 2010), and in all cases the application around the classifier remains unchanged. To achieve the possibility of exchanging the MFPC's core parts, its formulation of Eq. 6 is rewritten to

$$\mu_{MFPC}(\mathbf{m},\mathbf{p}) = 2^{-\frac{1}{M}\sum_{i=0}^{M-1} d_i(m_i,\mathbf{p}_i)} = \left(\prod_{i=1}^{M} 2^{-d_i(m_i,\mathbf{p}_i)}\right)^{\frac{1}{M}}, \tag{20}$$

revealing that the MFPC incorporates the geometric mean as its fuzzy averaging operator. Also, the unimodal membership function, as introduced in Eq. 3 with $A=1$, is isolated clearly, which shall be replaced by the PMFPC membership function described in the following section.

5.3.1 Probabilistic MFPC Membership Function

The PMFPC approach is based on a slightly modified MFPC membership function

$$\mu(m,\mathbf{p}) = 2^{-ld\left(\frac{1}{B}\right)d(m,\mathbf{p})} \in [0,1]. \tag{21}$$

D and *B* are automatically parameterised in the PMFPC approach. P_{CE} is yet not automated to preserve the possibility of adjusting the membership function slightly without needing to learn the membership functions from scratch. The algorithms presented here for automatically parameterising parameters *D* and *B* are inspired by former approaches: Bocklisch as well as Eichhorn developed algorithms which allow obtaining a value for the (MFPC) potential function's parameter *D* automatically, based on the used training data set. Bocklisch also proposed an algorithm for the determination of *B*. For details we refer to (Bocklisch, 1987) and (Eichhorn, 2000). However, these algorithms yield parameters that do not fulfil the constraints connected with them in all practical cases (cf. (Mönks, 2010)). Hence, we propose a probability theory-based alternative described in the following.

Bocklisch's and Eichhorn's algorithms adjust *D* after comparing the actual distribution of objects to a perfect uniform distribution. However, the algorithms tend to change *D* for every (small) difference between the actual distribution and a perfect uniform distribution. This explains why both algorithms do not fulfil the constraints when applied to random uniform distributions.

We actually stick to the idea of adjusting *D* with respect to the similarity of the actual distribution compared to an artificial, ideal uniform distribution, but we use probability theoretical concepts. Our algorithm basically works as follows: At first, the empirical cumulative distribution function (ECDF) of the data set under investigation is determined. Then, the ECDF of an artificial perfect uniform distribution in the range of the actual distribution is determined, too. The similarity between both ECDFs is expressed by its correlation factor which is subsequently mapped to *D* by a parameterisable function.

5.3.1.1 Determining the Distributions' Similarity

Consider a sorted vector of *n* feature values $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ with $m_1 \leq m_2 \leq \ldots \leq m_n$, thus $m_{\min} = m_1$ and $m_{\max} = m_n$. The corresponding empirical cumulative distribution function $P_m(x)$ is determined by $P_m(x) = \frac{|\tilde{\mathbf{m}}|}{n}$ with $\tilde{\mathbf{m}} = (m_i | m_i \leq x \; \forall i \in \mathbb{N}_n)$, where $|\mathbf{x}|$ denotes the number of elements in vector $\mathbf{x}$ and $\mathbb{N}_n = [1, 2, \ldots, n]$. The artificial uniform distribution is created by equidistantly distributing *n* values u_i, hence $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, with $u_i = m_1 + (i-1) \cdot \frac{m_n - m_1}{n-1}$. Its ECDF $P_u(x)$ is determined analogously by substituting $\mathbf{m}$ with $\mathbf{u}$. In the next step, the similarity between both distribution functions is computed by calculating the *correlation factor* (Polyanin, 2007)

$$c = \frac{\sum_{i=1}^{k} \left(P_m[x_i] - \bar{P}_m\right)\left(P_u[x_i] - \bar{P}_u\right)}{\sqrt{\sum_{i=1}^{k} \left(P_m[x_i] - \bar{P}_m\right)^2 \sum_{i=1}^{k} \left(P_u[x_i] - \bar{P}_u\right)^2}}, \tag{22}$$

where $\bar{P}_a$ is the mean value of $P_a(x)$, computed as $\bar{P}_a = \frac{1}{k}\sum_{i=1}^{k} P_a[x_i]$. The correlation factor must now be mapped to *D* while fulfilling Bocklisch's constraints on *D* (Bocklisch, 1987). Therefore, the average influence $\bar{\alpha}(D)$ of the parameter *D* on the MFPC membership function, which is the base for PMFPC membership function, is investigated to derive a mapping based on it. First $\alpha_D(x)$ is determined by taking $\frac{\partial}{\partial D}\mu(x, D)$ with $x = \frac{m - m_0}{C}$, $x > 0$:

$$\alpha_D(x) = \frac{\partial}{\partial D}\mu(x, D) = \frac{\partial}{\partial D} 2^{-x^D} = \ln(2)\left(-2^{-x^D}\right) x^D \ln(x). \tag{23}$$

The locations *x* represent the distance to the membership function's mean value m_0, hence $x = 0$ is the mean value itself, $x = 1$ is the class boundary $m_0 + C$, $x = 2$ twice the class boundary and so on. The average influence of *D* on the membership function $\bar{\alpha}(D) = \frac{1}{x_r - x_l}\int_{x_l}^{x_r} \alpha_D(x)\, dx$ is evaluated for $-1 \leq x \leq 1$: This interval bears the most valuable information since all feature values of the objects in the training data set are included in this interval, and additionally those of the class members are expected here during the classification process, except from only a typically neglectable number of outliers. The mapping of $D : c \to [2, 20]$, which is derived in the following, must take *D*'s average influence into consideration, which turns out to be exponentially decreasing (Mönks, 2010).

5.3.1.2 Mapping the Distributions' Similarity to the Edge's Steepness

In the general case, the correlation factor c can take values from the interval $[-1,1]$, but when evaluating distribution functions, the range of values is restricted to $c \in [0,1]$, which is because probability distribution functions are monotonically increasing. This holds for both distributions, $P_m(x)$ as well as $P_u(x)$. It follows $c \geq 0$. The interpretation of the correlation factor is straight forward. A high value of c means that the distribution $P_m(x)$ is close to a uniform distribution. If $P_m(x)$ actually was a uniform distribution, $c = 1$ since $P_m(x) = P_u(x)$. According to Bocklisch, D should take a high value here. The more $P_m(x)$ differs from a uniform distribution, the more $c \to 0$, the more $D \to 2$. Hence, the mapping function $D(c)$ must necessarily be an increasing function with taking the exponentially decreasing average influence of D on the membership function $\bar{\alpha}(D)$ into consideration (cf. (Mönks, 2010)). An appropriate mapping $D : c \to [2,20]$ is an exponentially increasing function which compensates the changes of the MFPC membership function with respect to changes of c. We suggest the following heuristically determined exponential function, which achieved promising results during experiments:

$$D(c) = 19^{c^{2q}} + 1 \Rightarrow D(c) \in [2,20], \tag{24}$$

where q is an adjustment parameter. This formulation guarantees that $D \in [2,20] \quad \forall c$ since $c \in [0,1]$. Using the adjustment parameter q, D is adjusted with respect to the aggregation operator used to fuse all n membership functions representing each of the n features. Each fuzzy aggregation operator behaves differently. For a fuzzy averaging operator $h(\mathbf{a})$, Dujmović introduced the objective measure of global andness ρ_g (for details cf. (Dujmović, 2007), (Mönks, 2009)). Assuming $q = 1$ in the following cases, it can be observed that, when using aggregation operators with a global andness $\rho_g^{h(\mathbf{a})} \to 0$, the aggregated single, n-dimensional membership function is more fuzzy than that one obtained when using an aggregation operator with $\rho_g^{h(\mathbf{a})} \to 1$, where the resulting function is sharp. This behaviour should be compensated by adjusting D in such a way, that the aggregated membership functions have comparable shapes: at some given correlation factor c, D must be increased if ρ_g is high and vice versa. This is achieved by mapping the aggregation operator's global andness to q, hence $q : \rho_g \to \mathbb{R}$. Our suggested solution is a direct mapping of the global andness to the adjustment parameter q, hence $q(\rho_g) = \rho_g \Rightarrow q \in [0,1]$. The mapping in Eq. 24 is now completely defined and consistent with Bocklisch's constraints and the observations regarding the aggregation operator's andness.

5.3.1.3 Determining the Class Boundary Membership Parameter

In addition to the determination of D, we present an algorithm to automatically parameterise the class boundary membership B. This parameter is a measure for the membership $\mu(m,\mathbf{p})$ at the locations $m \in \{m_0 + C, m_0 - C\}$. The algorithm for determining B is based on the algorithm Bocklisch developed, but was not adopted as it stands since it has

some disadvantages if this algorithm is applied to distributions with a high density especially on the class boundaries. For details cf. (Bocklisch, 1987).
When looking at the MFPC membership functions, the following two constraints on B can be derived: (i) The probability of occurrence is the same for every object in uniform distributions, also on the class boundary. Here, B should have a high value. (ii) For distributions where the density of objects decreases when going towards the class boundaries B should be assigned a small value, since the probability that an object occurs at the boundary is smaller than in the centre.
Hence, for sharp membership functions ($D \to 20$) a high value for B should be assigned, while for fuzzy membership functions ($D \to 2$) the value of B should be low. $B = f(D)$ must have similar properties like $\bar{\alpha}(D)$, meaning B changes quickly where $\bar{\alpha}(D)$ changes quickly and vice versa. We adopted Bocklisch's suitable equation for computing the class boundary membership (Bocklisch, 1987):

$$B = \frac{1}{1 + \left(\frac{1}{B_{\max}} - 1\right) \cdot \left(\frac{D_{\max}}{D}\right)^{1+\frac{1}{q}}}, \tag{25}$$

where $B_{\max} \in (0,1)$ stands for the maximum possible value of B with a proposed value of 0.9, $D_{\max} = 20$ is the maximum possible value of D and q is identical in its meaning and value to q as used in Eq. 24.

5.3.1.4 An Asymmetric PMFPC Membership Function Formulation

A data set may be represented better if the membership function was formulated asymmetrically instead of symmetrically as is the case with Eq. 21. This means

$$\mu(m, \mathbf{p}) = \begin{cases} 2^{-ld\left(\frac{1}{B_r}\right)\left(\frac{|m-m_0|}{C_r}\right)^{D_r}} & , m \le m_0 \\ 2^{-ld\left(\frac{1}{B_f}\right)\left(\frac{|m-m_0|}{C_f}\right)^{D_f}} & , m > m_0 \end{cases}, \tag{26}$$

where $m_0 = \frac{1}{M}\sum_{i=1}^{M} m_i$, $m_i \in \mathbf{m}$ is the arithmetic mean of all feature values. If m_0 was computed as introduced in Eq. 7, the resulting membership function would not describe the underlying feature vector $\mathbf{m}$ appropriately for asymmetrical feature distributions. A new computation method must therefore also be applied to $C_r = m_0 - m_{\min} + P_{CE} \cdot (m_{\max} - m_{\min})$ and $C_f = m_{\max} - m_0 + P_{CE} \cdot (m_{\max} - m_{\min})$ due to the change to the asymmetrical formulation. To compute the remaining parameters, the feature vector must be split into the left side feature vector $\mathbf{m}_r = (m_i | m_i \le m_0)$ and the one for the right side $\mathbf{m}_f = (m_i | m_i \ge m_0)$ for all $m_i \in \mathbf{m}$. They are determined following the algorithms presented in the preceding sections 5.3.1.2 and 5.3.1.3, but using only the feature vector for one side to compute this side's respective parameter.
Using Eq. 26 as membership function, the Probabilistic Modified-Fuzzy-Pattern-Classifier is defined as

$$\mu_{PMFPC}(\mathbf{m},\mathbf{p}) == \begin{cases} \left(\prod_{i=1}^{M} 2^{-ld\left(\frac{1}{B_r}\right)\left(\frac{|m-m_0|}{C_r}\right)^{D_r}} \right)^{\frac{1}{M}}, m \le m_0 \\ \left(\prod_{i=1}^{M} 2^{-ld\left(\frac{1}{B_f}\right)\left(\frac{|m-m_0|}{C_f}\right)^{D_f}} \right)^{\frac{1}{M}}, m > m_0 \end{cases}, \tag{27}$$

having in mind, that the geometric mean operator can be substituted by any other fuzzy averaging operator. An application is presented in section 6.2.

6. Applications

6.1 Machine Condition Monitoring

The approach presented in section 4 and 5.1 was tested in particular with an intaglio printing machine in a production process. As an interesting fact print flaws were detected at an early stage by using multi-sensory measurements. It has to be noted that one of the most common type of print flaws (Lohweg, 2006) caused by the wiping unit was detected at a very early stage.

The following data are used for the model: machine speed - motor current - printing pressure side 1 (PPS1) - printing pressure side 2 (PPS2) - hydraulic pressure (drying blade) - wiping solution flow - drying blade side 1 (DBS1) - drying blade side 2 (DBS2) - acoustic signal (vertical side 1) - acoustic signal (horizontal side 1) - acoustic signal (vertical side 2) - acoustic signal (horizontal side 1).
It has been mentioned that it might be desirable to preprocess some of the signals output by the sensors which are used to monitor the behaviour of the machine. This is particularly true in connection with the sensing of noises and/or vibrations produced by the printing press, which signals a great number of frequency components. The classical approach to processing such signals is to perform a spectral transformation of the signals. The usual spectral transformation is the well-known Fourier transform (and derivatives thereof) which converts the signals from the time-domain into the frequency-domain. The processing of the signals is made simpler by working in the thus obtained spectrum as periodic signal components are readily identifiable in the frequency-domain as peaks in the spectrum. The drawbacks of the Fourier transform, however, reside in its inability to efficiently identify and isolate phase movements, shifts, drifts, echoes, noise, etc., in the signals. A more adequate “spectral” analysis is the so-called “cepstrum” analysis. “Cepstrum” is an anagram of “spectrum” and is the accepted terminology for the inverse Fourier transform of the logarithm of the spectrum of a signal. Cepstrum analysis is in particular used for analysing “sounds” instead of analysing frequencies (Bogert, 1963).
A test was performed by measuring twelve different parameters of the printing machine’s condition while the machine was running (data collection) (Dyck, 2006). During this test the wiping pressure was decreased little by little, as long as the machine was printing only error sheets. The test was performed at a speed of 6500 sheets per hour and a sample frequency of

7 kHz. During this test 797 sheets were printed, that means, the set of data contained more than three million values per signal. In the first step before calculating the KLT of the raw data, the mean value per sheet was calculated to reduce the amount of data to 797 values per signal. As already mentioned, 12 signals were measured; therefore the four acoustical signals were divided by cepstrum analysis in *six new parameters*, so that all in all 14 parameters built up the new input vectors of matrix $\mathbf{X}$. As described above, at first the correlation matrix of the input data was calculated. Some parameters are highly correlated, e.g. PPS1 and PPS2 with a correlation factor 0.9183, DBS1 and DBS2 with a correlation factor 0.9421, and so forth. This fact already leads to the assumption that implementing the KLT seems to be effective in reducing the dimensions of the input data. The classifier model is shown in Fig. 4.
The KLT matrix is given by calculating the eigenvectors and eigenvalues of the correlation matrix, because the eigenvectors build up the transformation matrix. In Fig. 5 the calculated eigenvalues are presented. On the ordinate the variance contribution of several eigenvalues in percentage are plotted versus the number of eigenvalues on the abscissa axis. The first principal component has already a contribution of almost 60 % of the total variance. Looking at the first seven principal components, which cover nearly 95 % of the total variance, shows that this transformation allows a reduction of important parameters for further use in classification without relevant loss of information. The following implementations focussed only on the first principal component, which represents the machine condition state best.

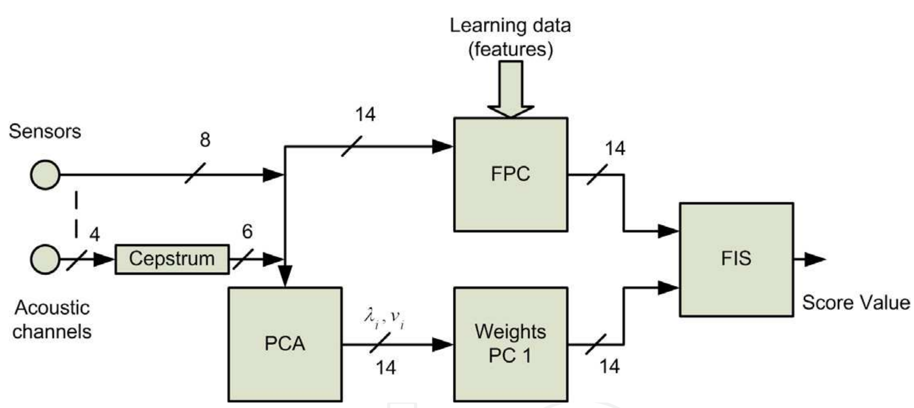


Fig. 4. The adaptive Fuzzy-Pattern-Classifier Model. The FPC is trained with 14 features, while the fuzzy inference system is adapted by the PCA output. Mainly the first principal component is applied.

PCA is not only a dimension-reducing technique, but also a technique for graphical representations of high-dimension data. Graphical representation of variables in a two-dimensional way shows which parameters are correlated. The coordinates of the parameter are calculated by weighting the components of the eigenvectors with the square root of the eigenvalues: the *i*th parameter is represented as the point ($\left|v_{1i}\sqrt{\lambda_1}\right|,\left|v_{2i}\sqrt{\lambda_2}\right|$). This weighting is executed for normalisation.

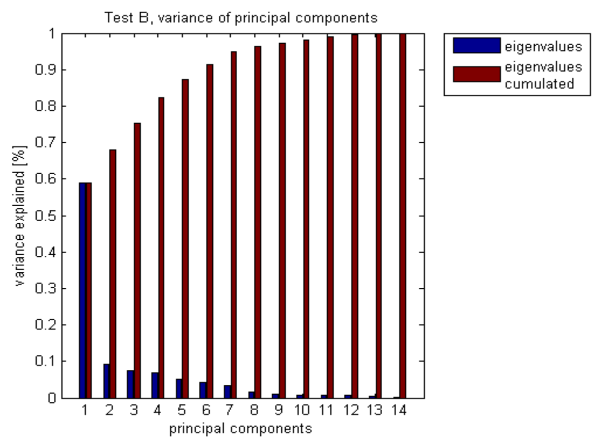


Fig. 5. Eigenvalues (blue) and cumulated eigenvalues (red). The first principal component has already a contribution of almost 60 % of the total normalized variance.

For the parameter "speed" of test B the coordinates are calculated as:

1. $(\left|v_{1,1}\sqrt{\lambda_1}\right|, \left|v_{2,1}\sqrt{\lambda_2}\right|) = (0.24\sqrt{7.8}, 0.14\sqrt{1.6}) = (0.67, 0.18)$, where

   $\mathbf{v}_1^T = (-0.24, -0.34, 0.19, 0.14, -0.02, -0.18, -0.34, ...)$, and
2. $\mathbf{v}_2^T = (0.14, -0.03, 0.65, 0.70, 0.10, 0.05, ...)$ and

   $\boldsymbol{\lambda}_i = \text{diag}(7.8, 1.6, 1.1, 0.96, 0.73, 0.57, ...)$.

All parameters calculated by this method are shown in Fig. 6. The figure shows different aspects of the input parameters. Parameters which are close to each other have high correlation coefficients. Parameters which build a right angle in dependence to the zero point have no correlation.

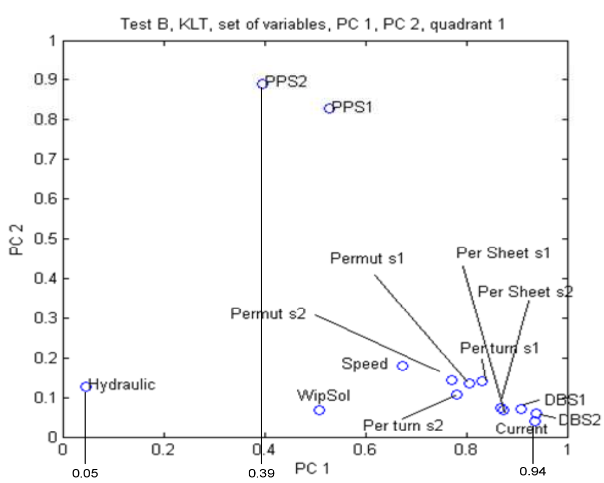


Fig. 6. Correlation dependency graph for PC1 and PC2.

The x-axis represents the first principal component (PC1) and the y-axis represents the second principal component (PC2). The values are always between zero and one. Zero means that the parameters' effect on the machine condition state is close to zero. On the other hand a value near one shows that the parameters have strong effects on the machine condition state. Therefore, a good choice for adaptation is the usage of normalized PC1 components.

The acoustical operational parameters sensed by the multiple-sensor arrangement are first analysed with the cepstrum analysis prior to doing the principal component analysis (PCA). The cepstrum analysis supplies the signal's representative of vibrations or noises produced by the printing press, such as the characteristic noises or vibrations patterns of intaglio printing presses. Thereafter the new acoustical parameters and the remaining operational parameters have to be fed into the PCA block to calculate corresponding eigenvalues and eigenvectors. As explained above, the weight-values of each parameter are taken as the weighted components of eigenvector one (PC1) times the square roots of the corresponding eigenvalues. Each weight-value is used for weighting the output of a rule in the fuzzy inference system (Fig. 4). E.g., the parameter "hydraulic pressure" receives the weight 0.05, the parameter "PPS2" receives the weight 0.39, the parameter "Current" receives the weight 0.94 and so forth (Fig. 6). The sum of all weights in this test is 9.87. All 14 weights are fed into the fuzzy inference system block (FIS).
Figure 7 shows the score value of test B. The threshold is set to 0.5, i.e. if the score value is equal to or larger than 0.5 the machine condition state is "good", otherwise the condition state of the machine is "bad" and it is predictable that error sheets will be printed. Figure 7 shows also that the score value passes the threshold earlier than the image signals. That means the machine runs in bad condition state before error sheets are printed.

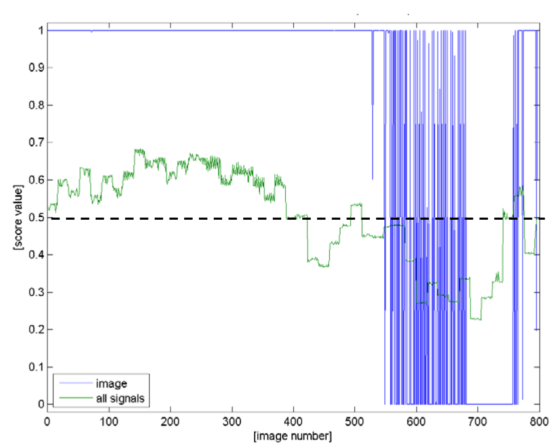


Fig. 7. Score value representation for 797 printed sheets. The green curve represents the classifier score value for wiping error detection, whilst the blue curve shows the results of an optical inspection system. The score value 0.5 defines the threshold between "good" and "bad" print.

6.2 Print Quality Check

As a second application example, an optical character recognition application is presented here. In an industrial production line, the correctness of dot-matrix printed digits are checked in real-time. This is done by recognizing the currently printed digit as a specific number and comparing it with what actually was to be printed. Therefore, an image is acquired from each digit, and 17 different features are extracted. Here, each feature can be interpreted as a single sensor, reacting on different characteristics (e. g., brightness, frequency content, etc.) of the signal (i. e. the image). Examples of the printed digits can be seen in Fig. 8. Actually, there exist also a slightly modified "4" and "7" in the application, thus twelve classes of digits must be distinguished.

Fig. 8. Examples of dot-matrix printed digits.

The incorporated classifier uses both the MFPC and PMFPC membership functions as introduced in section 5.3. Each membership function represents one of the 17 features obtained from the images. All membership functions are learned based on the dedicated training set consisting of 17 images per class. Their outputs, based on the respective feature values of each of the 746 objects which were investigated, are subsequently fused through aggregation using different averaging operators by using the classifier framework presented in (Mönks, 2009). Here, the incorporated aggregation operators are Yager's family of *Ordered Weighted Averaging (OWA)* (Yager, 1988) and Larsen's family of *Andness-directed Importance Weighting Averaging (AIWA)* (Larsen, 2003) operators (applied unweighted here)—which both can be adjusted in their andness degree—and additionally MFPC's original geometric mean (GM). We refer to (Yager, 1988) and (Larsen, 2003) for the definition of OWA and AIWA operators. As a reference, the data set is also classified using a *Support Vector Machine (SVM)* with a Gaussian radial basis function (RBF). Since SVMs are capable of distinguishing between only two classes, the classification procedure is adjusted to pairwise (or one-against-one) classification according to (Schölkopf, 2001). Our benchmarking measure is the classification rate $r_+ = \frac{n_+}{N}$, where n_+ is the number of correctly classified objects and N the total number of objects that were evaluated. The best classification rates at a given aggregation operator's andness ρ_g are summarised in the following Table 2, where the best classification rate per group is printed bold.

	Aggregation Operator	*PMFPC*		*MFPC*							
				$D=2$		$D=4$		$D=8$		$D=16$	
ρ_g		P_{CE}	r_+	P_{CE}	r_+	P_{CE}	r_+	P_{CE}	r_+	P_{CE}	r_+
0.5000	AIWA	**0.255**	**93.70 %**	0.370	84.58 %	0.355	87.67 %	0.310	92.36 %	0.290	92.90 %
	OWA	**0.255**	**93.70 %**	0.370	84.58 %	0.355	87.67 %	0.310	92.36 %	0.290	92.90 %
0.6000	AIWA	**0.255**	**93.16 %**	0.175	87.13 %	0.205	91.02 %	0.225	92.36 %	0.255	92.23 %
	OWA	**0.255**	**93.57 %**	0.355	84.58 %	0.365	88.47 %	0.320	92.63 %	0.275	92.76 %
0.6368	GM	**0.950**	**84.45 %**	0.155	81.77 %	0.445	82.17 %	0.755	82.44 %	1.000	82.44 %
	AIWA	**0.245**	**91.42 %**	0.135	85.52 %	0.185	90.08 %	0.270	89.81 %	0.315	89.95 %
	OWA	**0.255**	**93.57 %**	0.355	84.72 %	0.355	88.74 %	0.305	92.63 %	0.275	92.76 %
0.7000	AIWA	**1.000**	**83.65 %**	0.420	82.71 %	0.790	82.57 %	0.990	82.31 %	1.000	79.22 %
	OWA	**0.280**	**93.57 %**	0.280	84.85 %	0.310	89.01 %	0.315	92.76 %	0.275	92.63 %

Table 2. "OCR" classification rates r_+ for each aggregation operator at andness degrees ρ_g with regard to membership function parameters D and P_{CE}.

The best classification rates for the "OCR" data set are achieved when the PMFPC membership function is incorporated, which are more than 11 % better than the best using the original MFPC. The Support Vector Machine achieved a best classification rate of $r_+ = 95.04\%$ by parameterising its RBF kernel with $\sigma = 5.640$, which is 1.34 % or 10 objects better than the best PMFPC approach.

7. Conclusion and Outlook

In this chapter we have reviewed fuzzy set theory based multi-sensor fusion built on Fuzzy-Pattern-Classification. In particular we emphasized the fact that many traps can occur in multi-sensor fusion. Furthermore, a new inspection and conditioning approach for securities and banknote printing was presented, based on modified versions of the FPC, which results in a robust and reliable detection of flaws. In particular, it was shown that this approach leads to reliable fusion results. The system model "observes" the various machine parameters and decides, using a classifier model with manually tuned or learned parameters, whether the information is as expected or not. A machine condition monitoring system based on an adaptive learning was presented, where the PCA is used for estimating significance weights for each sensor signal. An advantage of the concept is that not only data sets can be classified, but also the influence of input signals can be traced back. This classification model was applied to different tests and some results were presented. In the future we will mainly focus on classifier training with a low amount of samples, which is essential for many industrial applications. Furthermore, the classification results should be improved by the application of classifier nets.

8. References

Beyerer, J.; Punte León, F.; Sommer, K.-D. Informationsfusion in der Mess- und Sensortechnik (Information Fusion in measurement and sensing), Universitätsverlag Karlsruhe, 978-3-86644-053-1, 2006

Bezdek, J.C.; Keller, J.; Krisnapuram, R.; Pal, N. (2005). Fuzzy Models and Algorithms for Pattern Recognition and Image Processing, The Handbook of Fuzzy Sets, Vo. 4, Springer, 0-387-24515-4, New York

Bocklisch, S. F. & Priber, U. (1986). A parametric fuzzy classification concept, Proc. International Workshop on Fuzzy Sets Applications, pp. 147–156, Akademie-Verlag, Eisenach, Germany

Bocklisch, S.F. (1987). Prozeßanalyse mit unscharfen Verfahren, Verlag Technik, Berlin, Germany

Bogert et al. (1963). The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, Cross-Cepstrum, and Saphe Cracking, Proc. Symposium Time Series Analysis, M. Rosenblatt (Ed.), pp. 209-243, Wiley and Sons, New York

Bossé, É.; Roy, J.; Wark, S. (2007). Concepts, models, and tools for information fusion, Artech House, 1596930810, London, UK, Norwood, USA

Brown, S. (2004). Latest Developments in On and Off-line Inspection of Bank-Notes during Production, Proceedings, IS&T/SPIE 16th Annual Symposium on Electronic Imaging, Vol. 5310, pp. 46-51, 0277-786X, San Jose Convention Centre, CA, January 2004, SPIE, Bellingham, USA

Dujmović, J.J. & Larsen, H.L. (2007). Generalized conjunction/disjunction, In: International Journal of Approximate Reasoning 46(3), pp. 423–446

Dyck, W. (2006). Principal Component Analysis for Printing Machines, Internal lab report, Lemgo, 2006, private communications, unpublished

Eichhorn, K. (2000). Entwurf und Anwendung von ASICs für musterbasierte Fuzzy-Klassifikationsverfahren (Design and Application of ASICs for pattern-based Fuzzy-Classification), Ph.D. Thesis, Technical University Chemnitz, Germany

Hall, D. L. & Llinas, J. (2001). Multisensor Data Fusion, Second Edition - 2 Volume Set, CRC Press, 0849323797, Boca Raton, USA

Hall, D. L. & Steinberg, A. (2001a). Dirty Secrets in Multisensor Data Fusion, http://www.dtic.mil, last download 01/04/2010

Hempel, A.-J. & Bocklisch, S. F. (2008). , Hierarchical Modelling of Data Inherent Structures Using Networks of Fuzzy Classifiers, Tenth International Conference on Computer Modeling and Simulation, 2008. UKSIM 2008, pp. 230-235, April 2008, IEEE, Piscataway, USA

Hempel, A.-J. & Bocklisch, S. F. (2010). Fuzzy Pattern Modelling of Data Inherent Structures Based on Aggregation of Data with heterogeneous Fuzziness Modelling, Simulation and Optimization, 978-953-307-048-3, February 2010, SciYo.com

Herbst, G. & Bocklisch, S.F. (2008). Classification of keystroke dynamics - a case study of fuzzified discrete event handling, 9th International Workshop on Discrete Event Systems 2008, WODES 2008 , pp.394-399, 28-30 May 2008, IEEE Piscataway, USA

Jolliffe, I.T. (2002). Principal Component Analysis, Springer, 0-387-95442-2, New York

Larsen, H.L. (2003). Efficient Andness-Directed Importance Weighted Averaging Operators. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 11(Supplement-1) pp. 67–82

Liggins, M.E.; Hall, D. L.; Llinas, J. (2008). Handbook of Multisensor Data Fusion: Theory and Practice (Electrical Engineering & Applied Signal Processing), CRC Press, 1420053086, Boca Raton, USA

Lohweg, V.; Diederichs, C.; Müller, D. (2004). Algorithms for Hardware-Based Pattern Recognition, EURASIP Journal on Applied Signal Processing, Volume 2004 (January 2004) pp. 1912-1920, 1110-8657

Lohweg, V.; Dyck, W.; Schaede, J.; Türke, T. (2006a). Information Fusion Application On Security Printing With Parametrical Fuzzy Classification, Fusion 2006-9th International Conference on Information Fusion, Florence, Italy

Lohweg, V.; Li, R.; Türke, T.; Willeke, H.; Schaede, J. (2009). FPGA-based Multi-sensor Real Time Machine Vision for Banknote Printing, Proceedings, IS&T/SPIE 21th Annual Symposium on Electronic Imaging, Vol. 7251, No. 7251-28, 9780819475015, San Jose Convention Centre, CA, January 2009, SPIE, Bellingham, USA

Lohweg, V.; Schaede, J.; Türke, T. (2006). Robust and Reliable Banknote Authentication and Print Flaw Detection with Opto-Acoustical Sensor Fusion Methods, Proceedings, IS&T/SPIE 18th Annual Symposium on Electronic Imaging, Vol. 6075, No. 6075-02, 0277-786X, San Jose Convention Centre, CA, January 2006, SPIE, Bellingham, USA

Luo, R.C. & Kay, M.G. (1989). Multisensor integration and fusion in intelligent systems, Systems, IEEE Transactions on Man and Cybernetics, vol. 19, no. 5, pp. 901-931, Sep/Oct 1989, IEEE Piscataway, USA

Mönks, U.; Lohweg, V.; Larsen, H. L. (2009). Aggregation Operator Based Fuzzy Pattern Classifier Design, Workshop Machine Learning in Real-Time Applications (MLRTA 09), Artificial Intelligence 2009, Paderborn, Germany

Mönks, U.; Petker, D.; Lohweg, V. (2010). Fuzzy-Pattern-Classifier Training with Small Data Sets, In: Information Processing and Management of Uncertainty in Knowledge-Based Systems, E. Hüllermeier, R. Kruse and F. Hoffmann (Ed.), Vol. 80, pp. 426 – 435, Springer, 978-3-642-14054-9, Heidelberg

Niederhöfer, M. & Lohweg, V. (2008). Application-based approach for automatic texture defect recognition on synthetic surfaces, IEEE Int. Conference on Emerging Technologies and Factory Automation 19, pp. 229-232, Hamburg, IEEE Piscataway, USA

Polyanin, A.D. & Manzhirov, A.V. (2007). Handbook of mathematics for engineers and scienctists, Chapman & Hall/CRC, Boca Raton

Ross, A. & Jain, A. K. (2006). Multimodal Human Recognition Systems, In: Multi-Sensor Image Fusion and its Application, R. S. Blum and Z. Liu (Ed.), pp. 289-301, CRC Press, 0849-334-179, Boca Raton

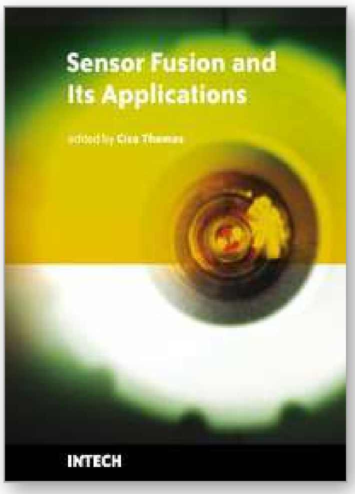
Schlegel, M.; Herrmann, G.; Müller, D. (2004). Eine neue Hardware-Komponente zur Fuzzy-Pattern-Klassifikation (A New Hardware Component for Fuzzy-Pattern-Classification), Dresdener Arbeitstagung Schaltungs- und Systementwurf DASS'04, Dresden, April 2004, pp. 21-26

Schölkopf, B. & Smola, A.J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press

Wald. L. (2006). Some terms of reference in data fusion, IEEE Transactions on Geoscience and Remote Sensing, No. 37(3), pp. 1190-1193, IEEE, Piscataway, USA

Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking, Systems, Man and Cybernetics, IEEE Transactions on 18(1) pp. 183–190

Zadeh, L. (1965). Fuzzy sets, Information Control, 8(3), pp. 338-353



Sensor Fusion and its Applications

Edited by Ciza Thomas

ISBN 978-953-307-101-5
Hard cover, 488 pages
Publisher Sciyo
Published online 16, August, 2010
Published in print edition August, 2010

This book aims to explore the latest practices and research works in the area of sensor fusion. The book intends to provide a collection of novel ideas, theories, and solutions related to the research areas in the field of sensor fusion. This book is a unique, comprehensive, and up-to-date resource for sensor fusion systems designers. This book is appropriate for use as an upper division undergraduate or graduate level text book. It should also be of interest to researchers, who need to process and interpret the sensor data in most scientific and engineering fields. The initial chapters in this book provide a general overview of sensor fusion. The later chapters focus mostly on the applications of sensor fusion. Much of this work has been published in refereed journals and conference proceedings and these papers have been modified and edited for content and style. With contributions from the world's leading fusion researchers and academicians, this book has 22 chapters covering the fundamental theory and cutting-edge developments that are driving this field.