IEEE Instrumentation and Measurement
Technology Conference
Budapest, Hungary, May 21-23, 2001.

A Methodological Approach to Multisensor Classification for Innovative Laser Material Processing Units

Cesare Alippi[1], Pietro Braione[1], Vincenzo Piuri[2], Fabio Scotti[1]

[1] Department of Electronics and Information, Politecnico di Milano, 20133 Milano, Italy, {fscotti,alippi}@elet.polimi.it
[2] Department of Information Technologies, University of Milan, 26013 Crema, Italy, piuri@elet.polimi.it

Abstract *- **Online quality detection and online laser beam control are important research topics to improve the overall quality of nowadays laser beam material processing units. In both cases innovative units are at study where the state is monitored by a set of heterogeneous in-process sensors conveying a large amount of information. However, low experiment reproducibility, lack of dominion knowledge and high costs greatly limit our ability of finding an optimal solution.***
In this paper we propose a methodology to guide the engineer's design choices towards an optimal implementation of the inductive classifier.*
Keywords *– **Multisensor classification, Laser processing, Quality detection, Sensor fusion, Feature extraction.***

I. INTRODUCTION

Nowadays an increasing attention towards material laser material processing is registered in those industrial sectors where narrow process windows and high quality levels are mandatory. Welding metals with laser is especially attractive for many aspects: Laser devices can concentrate enormous amounts of power on very narrow spots, without needing a complex setup. A correctly executed laser weld will have small area, high penetration depth, optimal mechanical properties, often even better than those of the base metal, and will be stable with time. What is more surprising, industrial laser welding processes have also a degree of flexibility incomparable to that of any other welding process: The only requirement for a piece to be welded is its optical visibility, since no contact between the welding head and the piece is required. All these features will come only at the premium of high capital costs due to the necessity of producing a stable beam, controllable in intensity and duration, with a very high output power and a correct wavelength. In facts, metals are very good light reflectors, even better at the typical emission wavelengths of the most common laser devices. As the metal melts, its light reflectivity suddenly lowers and the laser power is more promptly transferred to the workpiece. The main consequence is that the final quality of a weld is strongly dependent on the actual workpiece surface conditions at the welding spot, where small variations in metal reflectivity may cause high variations in the absorbed energy. This affects negatively the reproducibility of laser-based metal welding process, the situation being worsened by the fact that the beam must be correctly focused on a small area. Small laser spot areas are the key for high weld depth/width ratios, but also require low tolerances on weld geometry and beam focal spot position. But the major cause of low process reproducibility are perhaps the emissions from the weld area, -reflected radiation, vapours and ejected material, which re-enter in the welding system and damage the most fragile parts, namely the optics.

At the present time all these issues greatly limit the industrial fields where a laser based process can be applied. Fine-tuning the process parameters over factors like reflectivity and actual geometry of the workpiece is a demanding issue, which is still unsolved. Nowadays it is common industrial practice to set the process parameters in an open-loop fashion, usually trusting on the experience of a human operator to fine tune the parameters when process drifts excessively increase the percentage of rejects. Improving the overall economicity of laser material processing units by detecting process drifts as soon as possible (i.e., moving from post-process statistic inspection towards online quality monitoring and automatic process tuning) is the key which would enable moving laser material processing to industrial environments with too high yield requisites for current processes. This is true, e.g., for microelectronic spot welding processes, where much work is in progress to bring the advantages of laser welding to processes where high reflectivity and conductivity materials, like copper and aluminium, are employed. In a laser material processing unit the light radiation interacts with matter in a chaotic and rapid manner (a laser spot weld for electronics may last as few as 10^{-3} – 10^{-2} sec) and the parameters affecting the final quality are many, so that a model of the welding process cannot be explicitly derived with sufficient precision. As a consequence the road followed by many researchers is that of inductively finding a relationship between some property of the weld/cut piece and the observable variables of the ongoing process by using some inductive method like artificial neural networks (ANN). For example, in [1] the spectral content of acoustic emission has been related to the penetration depth in a CO_2 laser weld setup, while in [2, 3] weld depth has been related to plasma radiation and plasma charge voltage. However, we have no notice of studies where an integration of those sparse sources of knowledge in a unified methodological framework has been obtained.

During the last year the authors worked over weld quality classification based on the information from a set of heterogeneous in-process sensors (from now multisensor

0-7803-6646-8/01/$10.00 ©2001 IEEE

classification). During our work clearly emerged the necessity of a methodological framework to cope with the complex issues that arise from the peculiar nature of the problem, where lack of a priori information and scarce experiment reproducibility greatly limit the possibility to find a desirable solution. In this situation is of primary importance asking us whether or not the objectives can be achieved with the available data, and how an improvement can be obtained.

Our approach is methodological in the sense that it organizes a priori information and data with reasonable work assumptions in a sequence of loosely coupled methodological steps. The steps are ranked so that the reuse is maximized when work assumptions are removed as new dominion knowledge becomes available. Our approach also harmonizes several theoretical results from different fields, offering a consistent framework that guides the engineer's design choices.

This paper will be structured as follows; In Section 2 we will expose the issues which motivate a methodological approach; In Section 3 we will describe our approach, explaining each step with meaningful examples; In section 4 we will show how applying a methodological approach yields meaningful advantages over an unstructured approach, for what concerns both understanding the problem dominion and defining a solution close to the optimal one.

II. MOTIVATIONS

The main issues concerning multisensors classification for laser processing applications can be quickly summarized.

- Difficulties with identifying the physical process, due to its rapid and chaotic nature (a typical temperature cycle from room temperature to the metal boiling point and back to room temperature can be as less as 10^{-2} sec [4]). As a result, the correlation between the physical process values and the final piece features are largely unknown.
- Difficulties with sensing, mainly due to finding observable process variables related to specific quality factors and to sensor calibration for range and bandwidth (see for example [5]).
- Difficulties with quality assessment, due to lack of a reproducible quality parameter. Let us consider, for instance, a spot weld process for microelectronics. Many different factors can make a weld fail, although the most important is insufficient penetration depth. Among the many we have the formation of shrinkage cracks, contaminant inclusions and porosity, HAZ embrittlement, local alloy sensitisation, warping and misalignment [4]. What is more troublesome, the presence of one of more of these factors does not necessarily imply a failure (see for instance [6]).
- Difficulties due to poor experiment reproducibility caused by operation stress which alters the laser processing unit itself during time. One of the main factors which limit experiment reproducibility is the reflected radiation and the emissions from the weld pool, which return towards the welding system, damaging and polluting the sensible parts.
- Finally, the results of an experimental setup cannot be easily generalized to a different one, for instance with a different substrate.

We can resume all these issues by stating that laser beam processes are characterized by scarce a priori information, and that limited experimental data can be obtained from them in a reproducible way. Differently from what happens, for example, with numerical transmission, where large amount of real or simulated data can be obtained with a relatively low cost, setting up for laser welding and cutting has high capital costs, both for the laser welding device and for the sensors, and perhaps high operative costs. Moreover, thoroughly exploring the space of configurations often means the need to operate beyond the ordinary welding conditions. Physical changes in the sensible parts due to operation stress alter the experimental condition, thus severely compromising experimental reproducibility. To limit this issue we can avoid exploring the zones of the parameter space that may cause more stress to the sensitive parts of the setup, trading off polarization against reproducibility, but the issue generally remains serious. The situation is worsened by the fact that we do not know whether the observed variables are related to any or all the meaningful process parameters, neither how these parameters affect the final weld quality, not even if "weld quality" can unambiguously be defined. Finally, the lack of a mathematical model for the process makes impossible to obtain more data by simulation. All these aspects imply that we must use the available data efficiently, looking for the best trade-off between complexity and performance. Moreover, understanding what we can (and what we cannot) obtain is of primary importance to rethink the requisites in the (likely) eventuality they would reveal unattainable.

The only alternative to improve reproducibility as well as economicity, is designing experiments with a reduced number of samples. In this situation we must pay attention to objective definition and feasibility rather than simply trying to maximize performance. In a contest where the production of a working solution is required, this means both integrating reasonable a priori assumption with available knowledge and methods to produce an (heuristically) optimal solution, and pointing out the limits of this solution, also indicating the possible ways to overcome these limits.

A methodological approach relieves for some of those issues and helps us to focus on the relevant problems. Our approach is aimed at:

- Understanding, with the use of statistical techniques, the intrinsic limits of the automatic classification problem.
- Making reasonable work hypotheses in order to achieve a good data generalization, and point out how those hypotheses affect the validity of the solution.
- Selecting a reduced subset of meaningful features from a wide set of candidates.
- Tuning the model complexity over the available data to ensure good generalization properties.
- Evaluating and comparing different classification algorithms for meaningful merit figures (accuracy, latency, silicon area).

– Integrating new information (data and a priori knowledge) when it becomes available.

We will briefly discuss each point.

The Bayesian theory of classification tells us that there exists a lower bound on classification accuracy, which does not depend on the particular classification algorithm adopted, or on the number of available samples, but only on the distribution of samples in the feature space. If we estimate those distributions, we can have an idea on what will be the maximum performance attainable with the available dataset.

Introducing work hypotheses (a priori assumptions) where a priori information is incomplete, or not available at all, is what one commonly does every time he faces a classification problem in an inductive way [7], and this is especially true in our case, where scarce a priori information is available. A priori assumptions are the inductive bias of the methodology and, altogether with the intrinsic amount of information contained in the experimental data, are the factors that determine the limits of the methodological approach. Our methodological approach helps the engineer to understand what hypotheses more affect the validity of the final solution, and thus to formulate "reasonable" hypotheses.

It is necessary to extract from the signals produced by the sensors those features that are relevant to the classification problem, thus making simpler the classifier design and improving classification accuracy. Here the problem is understanding which features describe the process in a way almost accurate as the full signals, and which constitute a minimal set, avoiding both the "curse of dimensionality" and incomplete process description.

Since the famous "Occam's Razor" reducing model complexity emerged as an important issue. For an inductive classifier it means lowering its implementation costs, reducing its size, improving noise rejection and loosening its inductive bias, thus improving its generalization ability.

When new a priori information becomes available, we are allowed to remove some of our hypotheses and replace them with the new available knowledge. A priori assumptions may also be altered or removed if we discover later in the methodology that they excessively limit our approach. This implies rethinking a part of the work done, returning back to the methodological step based over the removed assumption and propagating the change forward to the subsequent steps. Our methodology organizes the work so that "what can be saved will be saved". This objective is met by structuring the methodology so that the assumptions which are more likely to be removed, or whose removal will more affect the work done, are made later in the methodology, when most of the work has already been done.

III. A METHODOLOGICAL APPROACH TO MULTISENSOR CLASSIFICATION

In Figure 1 the structure of the methodology is reported as a flow chart, where each step is represented. The methodology has a waterfall structure, but some of the methodological steps can be followed in parallel. In this section we will briefly discuss each of the methodological steps.

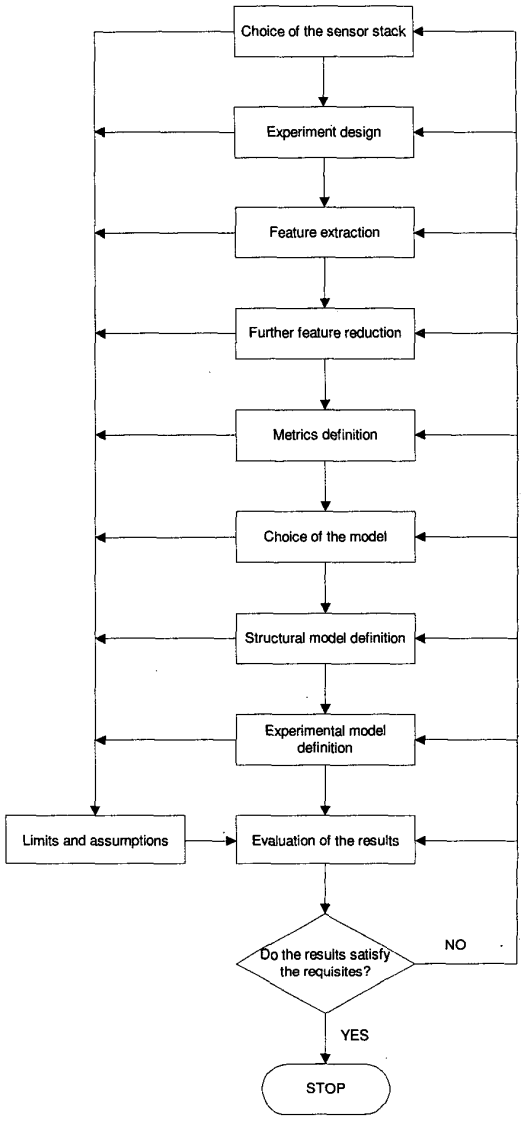

Figure 1 – Structure of the methodology

The choice of the sensors which surround the welding head is driven by two contrasting issues: That of keeping the setup as simple – and robust – as possible, and that of obtaining a thorough view of the ongoing process. Since scarce a priori information is available on which signals convey meaningful or redundant information, choosing a wide set of sensors during the experimental phase and deciding with the subsequent feature selection step whether some sensor can be discarded or not seems to be the most reasonable solution. All the meaningful acoustic, thermal, visual and electromagnetic phenomena from the ongoing welding process can be monitored to collect as more information as possible over it.

Experimental sensor stacks have been realized containing temperature sensors, on-axis and off-axis reflected radiation monitors, sonic and ultrasonic sensors, CCD cameras, eddy current sensors and plasma radiation sensors.

Experiment design is mainly driven by requirements on confidence: The less the number of available samples, the less the confidence over the performance estimates of the obtained classifier [8]. Experiment design must also consider all the issues that can limit our ability of exploring the configuration space, thus biasing our "vision" of the process. This is particularly true for laser processing units, where operation stress limits experiment reproducibility, as pointed out previously. A tradeoff must be met between the need to do as few experiments as possible, in order to reduce costs and improve experiment reproducibility, and the need to raise the confidence level of our estimates. Normally the need to improve experiment reproducibility dominates, so we cannot expect consistent experimental datasets with more than few thousands of samples.

From all of these signals features can be extracted which are related ([1], [2], [9], [10]) to meaningful final quality factors both in laser cutting and in laser welding applications. The feature extraction phase, driven by the a priori knowledge of the welding physics, generates from the sensors' time signals a vector of features more suited for feeding a classifier.

The first three methodological steps are driven by high-level knowledge and methods from statistics and welding physics Only economic requirements and very generic work assumptions (that a solution exists, that the final weld quality is related mainly to observable physical quantities) may limit this phase. From this point of view we say that they are problem-centric steps.

The following steps are at a lower level of abstraction, and thus more subject to rethinking due to the availability of new a priori knowledge and to the need to remove some work assumption when the objectives are not reached. Indeed, repeating those steps implies less costs than, for instance, having to repeat the whole experimental set because we discovered that a previously unexplored sensor may convey useful information about some meaningful physical phenomena not detected by the others.

Being sensor selection and feature extraction problem-centric steps, a high number of features may be produced in order to avoid excluding information that may be meaningful. Given the limited size of experimental datasets, a further data reduction step is mandatory in order to prevent the "curse of dimensionality" problem. The data reduction step differs from feature extraction as it follows a solution-centric rather than problem-centric approach. Many approaches are available, like PCA ([8], [11]), ICA ([12]), or simple feature selection. In Section 4 we will present a growing feature selection heuristic based on Bayesian error minimization, and we will compare it with a "traditional" data reduction technique, PCA. In this phase some considerations about the theoretical limits on classification precision may be drawn, as we will show in Section 4.

Many inductive classification paradigms are available to the designer. Among these the statistical classifiers, like KNN or Parzen classifiers [8], ensure a design that aims to minimize the Bayesian error over the available samples. Soft-computing methods like neural networks, on the other hand, often achieve performances very close to that of statistical classifiers with a structure several orders of magnitude simpler, but the training methods are not so strongly related to Bayesian error minimization [13]. Also, some models like feedforward non-linear network requires a training phase which conditions the actual performance of the classifier. In this case we must select a training strategy and a criterion to select from a set of trained models the optimal one. The criterion commonly followed is performance over a cross-validation data subset, not used during the training phase.

The last steps of the methodology require the definition of a set of meaningful metrics, in order to compare, select and evaluate the trained models. We may define a performance metric, like the estimated misclassification error, and an implementation dependent cost metric, like silicon area for minimal latency implementation. Problems may arise when reduced datasets are available (as is our case), or when different model paradigms must be compared one against the other. In Section 4 we will propose a performance metric that permits a comparison between KNN and neural classifiers.

The final methodological steps further lower the level of abstraction, as they tune the model over data by maximizing the performance metric previously defined. A structural model definition phase is necessary for those model families which depend on a structural parameter, like the number of neurons in the hidden layer for two-layer feedforward neural networks, the K for KNN classifiers, the spread for kernel functions in the Parzen classifier, etc. The following step, experimental model definition, determines the model parameters by training over available data. A model selection phase may be necessary for those model families which strongly depend on the training phase, as discussed previously, but different solutions, like building a majority-voting classifier with all the obtained trained models, are available. Some model families, also, do not require a true training over data. This is the case of the KNN classifier where, once a suitable value for K has been established during the structural model definition phase, the classifier is immediately built from data. This permits us to avoid using some of the (few) available samples to build a cross-validation set for the model selection phase, so that all of them can be used to build the classifier.

Finally, if the cost or performance objectives are not met, we can iterate over the methodological steps, moving back in the hierarchy. We can repeat the model selection phase, retrain the networks, try with more hidden neurons, choose a different classifier (e.g. a Parzen classifier rather than a KNN one), look for other features, and at last the options with higher costs: doing more experiments, adding new sensors and looking for a better (a priori) knowledge of the physical process.

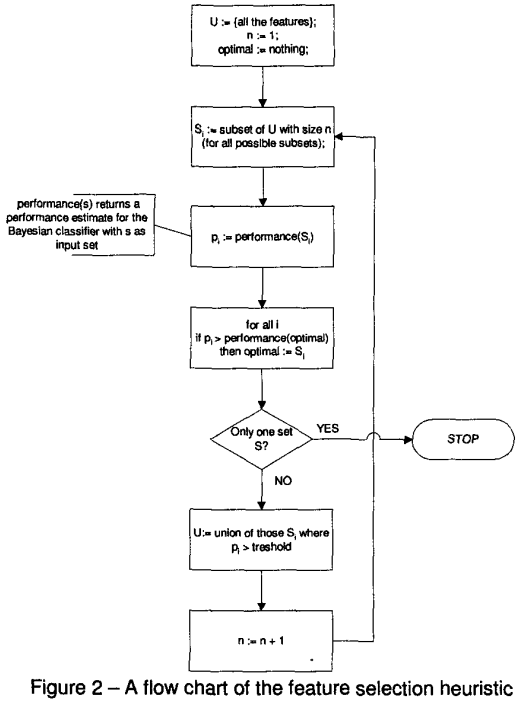

Figure 2 – A flow chart of the feature selection heuristic

IV. A PRACTICAL APPLICATION

Our methodology has been applied to an experimental dataset of 599 spot welds over a stainless steel substrate, produced during a feasibility study for automatic weld classification by Philips AG. The setup was completed by a set of sensors detecting the incident and the on-axis reflected radiation, the sonic emissions, the charge of the plasma, and the temperature of the weld pool. The setup was dismantled after the production of the experimental data and was not available at the time we started our work.

We first selected 20 meaningful features among those selected by Philips, and then we applied a feature selection heuristic approach to the data reduction phase. In Table 1 the results of each iteration of the heuristic have been reported; The features have been reported with their progressive number in order to preserve the non-disclosure agreement with Philips AG. The heuristic uses the KNN classifiers as estimators for the Bayesian error probability of each possible 1-input classifier to rank all the features with the leaving-one-out method. We chose to use the 1NN classifier because of its statistical properties, strongly related to the Bayesian error probability [8].

A set of candidates is selected among the highest-rank ones, then the procedure is iterated with all (or some of) the two, three, ... input classifiers that can be built from all the possible unions of the feature sets produced at the previous iteration, obtaining at the n-th step a certain number of feature sets each containing n features. This procedure rapidly converges to a single set of features, or may be stopped before convergence. We can thus characterize our heuristic as a greedy growing iterative procedure. Our heuristic also considered the estimates yielded by the 3NN and the 5NN classifiers to rank the features. Classifiers with higher K were discarded for their excessive cost in terms of number of units, and even values of K were not considered for their poor performance due to their wide rejection regions. The estimates of the Bayesian error for each set of features obtained with the leaving-one-out method have the desirable property to be the ones with the highest level of confidence among those achievable with 599 samples and a KNN estimator [8].

Iteration	Error threshold	Features selected
1	< 35%	3;8;20;21;23;24;32
2	< 25%	(3,8);(3,20);(3,32);(8,20); (8;24);(8,32);(20,32)
3	< 20%	(3,8,32);(3,20,32);(8,20,32)
4	-	(3,8,20,32)
5	Adding one by one all the others	No additional gain

Table 1 – Iterations of the heuristic feature selection algorithm

Model	Error	Confidence	Complexity	Notes
KNN	11%	~ 8 – 18%	>300,000 Ne	K = 5
ff NN	9%	~ 5 – 16%	12 Ne	best of 100

Table 2 – Classification error and confidence for the best classifier (minimum latency implementation)

We compared the results obtained with our feature selection approach against those achievable with a traditional feature reduction approach as PCA. Our heuristic scored for better performances with lower K values: 11% of misclassified samples with K=5 and 4 features against 12% with 2 features and K=11. On the other hand our heuristic is sensibly slower (hours against seconds) than PCA, but is this is not an issue as feature reduction is a design step performed offline just once. Moreover, feature selection helps us to reduce the overall complexity of the final system as this is largely dominated by the feature extraction step (see Table 3). Please note that, while the classification steps may be significantly parallelized, this is not the case of the feature extraction step, as features are present whose value cannot be calculated until another feature's value is known.

Model	Feature extraction	Preprocessing	Classifica-tion
KNN	>200,000 flops	~10 flops	~200 flops
ff NN			~10,000 flops

Table 3 – FLOating Point operationS necessary to classify a sample

We may assess a lower bound for confidence over performance by applying the results exposed in [8], which establish how the optimal confidence value the one obtainable with the optimal Bayesian classifier. In our work this step proved very useful to understand how "trustable" are the performance figures obtained for our classifiers. In an

industrial environment this step would be done "a priori" by choosing the number of samples during the experiment design methodological step. Here a suitable trade-off must be met between a high degree of confidence, which would require to perform as many experiments as possible, and the need to keep the welding setup consistent by avoiding stress-due process drifts. The feature selection phase showed its validity when more than the four selected features were used to estimate the Bayesian error. No improvement was registered over the estimated error: On the contrary, the estimated precision worsened about 2% to 4%, showing clearly the disadvantage of using a dimensionality greater than 4.

Two classifier structures, two-layer feedforward neural network with sigmoid hidden neurons and the KNN classifier, were compared against a performance metric and a cost metric. We trained a high number of networks and assumed that the network which best performed over a validation set of samples not used to train the network is also the best network. At this purpose the sample set has been randomly partitioned in a design and a validation subset, and all the networks have been trained/validated over them. We used the available a priori information about the data set structure in order to reduce the polarization introduced when partitioning. The estimates of the neural classifier error, obtained with the sample partitioning method, are less confident than those obtained with the leaving-one-out method for the KNN classifier, this situation being somehow healed by the fact that the best performing neural classifier yields a slightly better error estimate than the KNN one.

We used the results exposed in [14], where a two-layer neural implementation of the 1NN classifier is presented, and then we employed the number of neurons of this classifier as a figure of merit for the KNN minimum latency spatial complexity. Such a figure of merit is a function of the total number of available samples n, which is $O(n^2)$. We can see that in our case the lowest latency KNN classifier has a spatial complexity order of magnitudes lower than the neural one. This statement can be further generalized by recalling the results exposed in [13]: We can thus assert that, when the number of samples is considerably higher than the number of inputs – a condition which is normally met, the spatial complexity of a neural network cannot be more than $O(i\, n^{1/i})$, where i is the number of its inputs, as $O(i\, n^{1/i})$ is the number of neurons of a network with i inputs which can completely overfit aver all the available data. We can conclude that the KNN classifier, whose complexity grows with the number of available samples, scales worse than the neural classifier, which needs a considerably lower number of units to represent decision surfaces yielding comparable performance.

V. CONCLUSIONS

In this paper we presented a methodological approach to multisensor classification, suited for different laser processing units, and its application over real-world data. Nowadays industrial research aims at bringing the experience on multisensor classification outside the laboratories to the production lines. In this context our methodological effort represents an innovative work of practical relevance, which proved its advantages in guiding the designer's choices, producing comparable figures of merit, interpreting the performance indicators, tuning a priori work assumptions and organizing several theoretical as well as practical results in a unified framework.

REFERENCES

[1] D.F.Farson, K.S.Fang, and K.T.Kern. Classification of laser welds by acoustic signature. In *Proceedings of the 29th Conference on Decision and Control*, pages 1620–1621. IEEE, December 1990.

[2] Y.Takanohashi, S.Kimura, Y.Sameda, and K.Miyahara. Quality monitoring technology for laser welding. In *Proceedings of the IECON 1995*, volume 2, pages 1164–1168. IEEE, 1995.

[3] A.I.Gavrilov. Development of automatic control system for laser welding using neural network technology. In *UKACC International Conference on Control 1998*, volume 1, pages 278–283. IEE, September 1998.

[4] L.Migliore. The principles of laser welding. *Optoelectronics world*, July 1998.

[5] R.H.J.N.Bitter. Design and calibration of a temperature sensor for laser spot weld monitoring. Technical report, Stan Ackermans Instituut, 1999.

[6] V.E.Merchant. Laser welding in the pipeline industry. In D.Belforte and M.Levitt, editors, *The industrial laser handbook*. Springer-Verlag, 1992-1993 edition, 1993.

[7] T.M.Mitchell. *Machine learning*. McGraw Hill, 1997.

[8] K.Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, 1972.

[9] W.M.Steen. In *Proceeding of LAMP 1992*, page 439, 1992.

[10] H.-A.Crostack, and M.Maaß. Non-destructive evaluation of seams of laser welded tools using the cs-eddy current technique. In *Proceeding of the IECON 1998*, volume 4, pages 2279–2283. IEEE, 1992.

[11] E.Dunn. *Applied multivariate data analysis*. Arnold Press N.Y., 1991.

[12] A.Hyvarinen. Survey on Independent Component Analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[13] J.Makhoul. Pattern recognition properties of neural networks. In *Workshop on Neural Networks for Signal Processing*, pages 173–187. IEEE, 1991.

[14] Y.Q.Chen, M.S.Nixon, and R.I.Damper. Implementing the k-nearest neighbour rule via a neural network. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 136–140. IEEE, 1995.