DOI 10.5162/AHMT2014/4.3

Application-Specific Partially Automated Design of Multi-Sensor Intelligent Lab-on-Spoon System

Kittikhun Thongpull, Andreas König

Institute of Integrated Sensor Systems, University of Kaiserslautern, Kaiserslautern,
Tel.: 06312053124, E-Mail: thongpull@eit.uni-kl.de

Abstract

In this work, we pick up the ISE research on automation of intelligent integrated systems design and apply and adapt the concept, methodology, and tool implementation. In particular, a new platform, based on the proven machine-learning tool ORANGE will make the work multi-platform and open-access. We demonstrate the current status by the application to our Lab-on-Spoon (LoS) multi-sensory system and corresponding data. The creation of a an efficient Support-Vector-Machine (SVM)-based hierarchical classification architecture and options of automated feature selection (AFS) are investigated in the experiments. An improved in two cases to 100% classification accuracy with regard to the previous flat approach was achieved.

Keywords : System Design Automation, Intelligent Systems, Machine Learning, Support Vector Machine, Impedance Spectroscopy, Lab-on-Spoon

1 Introduction

The ongoing technological advance from sensor, electronics to methods and algorithms allows the realization of more and more capable systems in many fields, such as, e.g., distributed intelligence, ambient intelligence, ambient assisted living (AAL), automation, and general smart environment applications i.e., smart-kitchen activities. Lab on Spoon [1], with rich potential in assisted living, medical application, as well as food manufacturing process and food analysis related applications, is a system that provides versatile sensory context for smart-kitchen scenarios to determine food ingredient quantity and quality, i.e., identify the type of ingredient as well as its quality, in particular, detect and quantify existence of the specific substances for contamination detection. Thus, to achieve these abilities, the efficient realization of a multi-sensor measurement system is pursued from standard and customized sensors and electronics, sophisticated data acquisition and embedded processor, and, in particular, effective computation intelligent processing for pattern recognition tasks. The multi-sensor information thus available can be exploited by efficient application of specific fusion and classification. Therefore, designing such a system imposes significant effort on a designer and expectation of expertise. With the research goals to alleviate these requirements, speed-up design, and potentially improve solution quality, an automation of intelligent system design, based on prior ISE research [2] [3] [4], as an enhancement of expert-centered design, is pursued in this work with a focus on multi-sensor integrated sensory system, such as, e.g., the LoS [1] as well as DeCaDrive system for driver monitoring [5]. Based on the prior ISE research work, the general methodology is picked-up and implemented in a new framework based on the ORANGE [6] system. In contrast to previous tool implementations, e.g., proprietary QuickCog [3] or GENESIS [4], ORANGE is a multi-platform open-access tool, so that our work can be of general validity and allow reuse in research. ORANGE system, has been extended for sensor system interfacing and gives access to advanced methods for visualization, analysis, and classification of the LoS data. The aspired methodology and framework will allow the rapid and optimized design of intelligent systems tailored to the application. This includes also the leanness of the aspired system by including according constraints in the optimization. LoS applications are various and for each new or enlarged task, e.g., increase of ingredients or substances to discern, a new recognition systems has to be conceived. The partial automation of this task will be studied in our work for a set of typical examples, regarding, e.g., automated feature selection or nested training of hierarchical classifiers for decision level fusion techniques. The organization of this paper is as follows. Section II introduces the architecture of the multi-sensor intelligent system and describes the implementation in the Lab on Spoon case study. Further, it introduces a hierarchical SVM classification topology with automated parameter search techniques. Data acquisition, experiments and results, and discussion will be provided in Section III.

DOI 10.5162/AHMT2014/4.3

2 Automated Multi-Sensor System Design

Our goals are to contribute to the automation of intelligent sensor systems design for potentially complex recognition system composed of various methods and algorithms [7]. For this aim, based on well established as well as newly emerged and evolved signal processing and computational intelligence, we have developed a proposed concept, methodology, and a framework for automated design of intelligent multi-sensor systems [2] [3] [4]. The standard architecture blocks of automated intelligent multi-sensor systems related to recognition applications is illustrated in Fig. 1. The highlighted box indicates the focus of research in this paper, which is described in the following subsections.

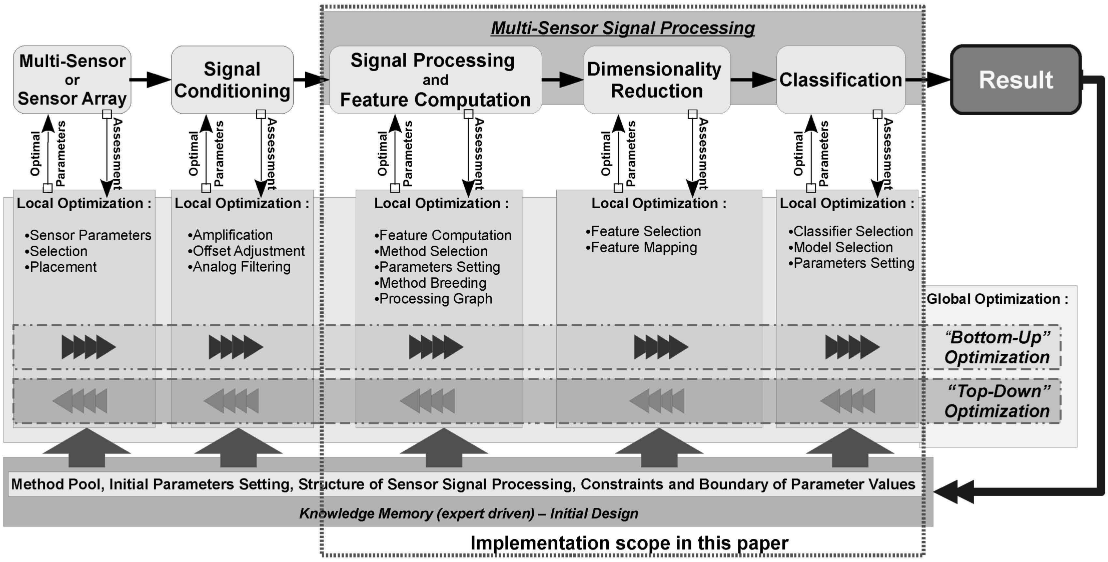
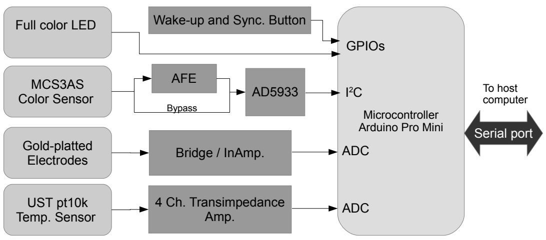


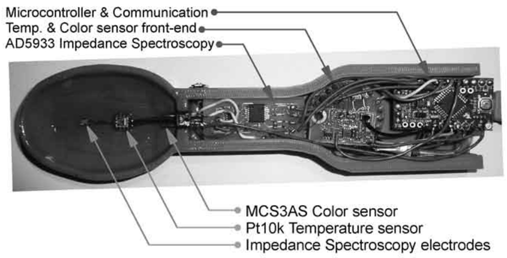
Figure 1: The concept of an automated intelligent multi-sensor system design.

2.1 Multi-Sensor Intelligent Lab-on-Spoon System

The system is based on the novel concept of combining sensors for visual and impedance spectroscopy with an embedded processor and communications to achieve an autonomous, well discerning system. The outcome of multi-sensor solution will be exploited in later advanced computational intelligence techniques, e.g, the automated selection of features and, in particular, advanced classification methods based on automated approach. The current LoS prototype includes a ceramic substrate pt10k temperature sensor with a custom calibrated analog circuit, the MCS3AS true color sensor and corresponding transimpedance 4-channel amplifier chip, and an embedded impedance spectroscopy measurement, the AD5933 network analyzer chip with gold-plated electrodes applied. For low-impedance measurements, typically below 1kΩ, extended by an analog front-end (AFE). The programmable amplification of the AD5933 input stage has to be set with regard to the aspired impedance measuring range. An Arduino Pro mini microcontroller module is in charge of sensors reading and measurement routines as well as data transferring to host computer. The analog sensors are measured using integrated 10-bit resolution ADC module of Arduino's microcontroller. The LoS's hardware block diagram as well as the current LoS prototype printed in a spoon shape package from our MakerBot 3D printer are shown in Fig. 2a and 2b.



(a) Hardware block diagram of Lab-on-Spoon.



(b) Lab-on-Spoon first prototype.

Figure 2: The concept and the implementation of Lab on Spoon

2.2 Automated Feature Selection

Further, we focus on automatically reducing the dimension of feature space in the early stage of computation. An Automated Feature Selection [8] is already available in the QuickCog and other systems and will be added to ORANGE in the aspired research together with advanced optimization techniques. AFS leads to more lean, in some cases better performing decision systems. However, the SVM does not depend on this to achieve excellent recognition ability, but it can still benefit due to reduced system complexity, resource, and energy consumption. In particular, the spectroscopic methods, as e.g., impedance spectroscopy can benefit from application-specific selection, because the measurement time as well as the effort will be substantially reduced, knowing the key frequencies to inspect for a certain task. So, AFS can also be understood for LoS as a data analysis and knowledge acquisition step applied to data for full-scale measurements to achieve lean and fast measurements in following applications.

2.3 Hierarchical SVM Classification

We have developed a multi-level architecture in order to enhance classifying performance of full spectrum data processing advancing from conventional flat SVM classification [9]. The purpose of the desired architecture is to obtain and achieve a more balanced weighting of sensory channels by fusing on the decision level. Thus, each individual sensory channel is assigned to an independent first-level SVM classifier in the hierarchy. Thus, local optimizations are solved in each sensory channel giving dedicated optimal parameters that fit to a particular sensory channel. In training, in the first-level stage, for each sensory channel an SVM with optimum parameters is generated. These generated SVM classifiers estimate probabilistic values for the different classes using an embedded Platt's posterior probabilities [10] estimator from LIBSVM [11] library returning a class probabilistic vector, denoted as class-P vector in the following. The dimension of this class-P vector depends on the number of classes as well as the implemented multi-class classification approach. In addition, SVM is an intrinsic binary classifier and the combination of multiple binary SVM classifiers is generally employed to achieve multi-class classification problems. Considering the one-against-one (pairwise) approach [11] , the total number of pairwise binary SVM classifier is derived from $k(k-1)/2$. The estimation of pairwise of k class probabilities of input data x from ith and jth classes is formulated as eq.1.

$$p_i = P(y = i \mid \boldsymbol{x}), i = 1, ..., k \qquad r_{ij} \approx P(y = i \mid y = i or j, \boldsymbol{x})$$

$$r_{ij} \approx \frac{1}{1 + e^{A\hat{f}+B'}} \tag{1}$$

where $\hat{f}$ is the decision value of trained SVM classifier model, A and B are estimated from training data by minimizing the negative log-likelihood function. The $\hat{f}$, A and B arguments are computed in a particular sensory channel of training procedure. These are also used to produce class-P vectors of test data for prediction procedures. All class-P vectors are concatenated in a global class-P vector that corresponds a probabilistic patterns of classes of all sensory channels. In the top level stage, the final SVM classifier performs the final class output determination based on class probabilistic information obtained from the previous level stage. The proposed approach is illustrated in Fig.3.

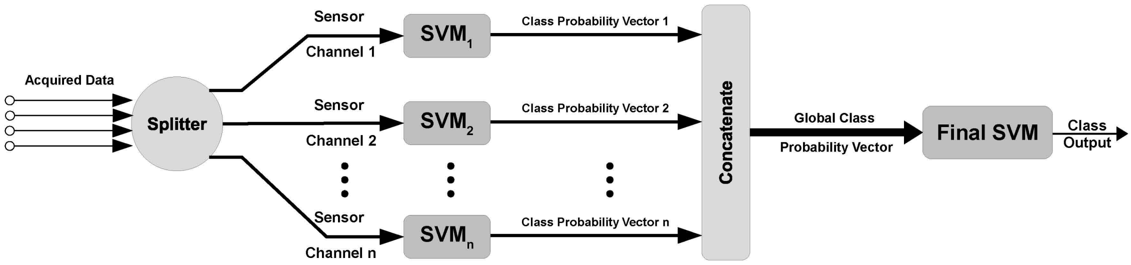


Figure 3: Hierarchical SVM Classification

2.4 System Implementation in ORANGE

ORANGE is a machine learning and data mining suite for data analysis through Python scripting and visual programming. We have been developing various additional functions, features, and effective methods of computational intelligence extended from standard available modules in ORANGE. The LoS Serial Port Interface module is a Python-based and consisted of a LoS device data communication function via serial port interface for sensor reading and measurement procedures as well as importing and transforming the acquired data for post processing in ORANGE system. The LoS Feature Selector module is the data selector that recognize the acquired multi-sensor data in LoS format. The feature selector selects data corresponded LoS's sensor type domain e.g. temperature, color spectrum, impedance. An ORANGE's inclusive C type SVM classifier with a RBF kernel function, implemented from the library LIBSVM in ORANGE, is chosen for all classification operators in this work. In the hierarchical SVM Classification implementation Python Script widgets are used, because the desired algorithm is not implemented by standard widgets and modules. The optimal SVM parameters are obtained with an automated parameter search by using embedded LIBSVM's procedures internally implemented with the Grid-search method and Cross-validation techniques.

3 Experiments and Results

The data sets used in this work are obtained from six different ISE's experiments in food ingredients scenarios of recently experiments demonstrated an early prototype of LoS. The data sets are: Soy data set consists of four liquid cooking ingredients, i.e., soy sauce, vinegar, and tab water, Beer data set consists of four different brands of beer, Oil data set of three kinds of cooking oil, Wine data set consists of seven kinds of wine, Used oil data set was extracted from fresh and used cooking oil, and Glycol in Wine consists of pure wine and wine contaminated with Glycol. These data sets are named as Soy, Beer, Wine, Oil, Used Oil and Glycol in Wine repectively. 30 measurement repetitions have been set as the standard for each substance or ingredient. LoS data output of a single measurement represents an instance of data vector consisting of one temperature value, three RGB color values, and 512 complex impedance values including magnitude and phase which amounts to 1028 entries stored in double precision of ORANGE's data format. The transimpedance of the LoS color sensor is set to 500 kΩ. The AD 5933 acquires impedance spectrum in the frequency range of 10kHz-100kHz. The measurement proceeds in three phases, measuring the temperature, followed by accumulative color registration of 50 samples concluding with the impedance spectroscopy measurement. Currently, we consider the temperature only as a context information and constrain multi-sensor processing and classification to color and impedance only. Every data set is separated by a hold-out random data splitting, 50% for training and 50% for testing. The experiments were divided into three steps. In the first step, we demonstrate the conventional flat SVM classification method where the complete sensor data was employed in a single SVM classifier computation. In the second step, instead of computing with the complete data and all features, we applied AFS function employing QuickCog to reduce the input vector space of the flat SVM. In the third step, to prove the hierarchical classification concept, three SVM are individually trained from color, magnitude, and phase channel outputting class distribution probability vectors. The final class output is computed from the concatenated class distribution probability vector from three sensor channels that are given to the final SVM classifier. The automated parameters search procedure is enabled in all employed SVM finding an optimal value for parameter C, the penalty parameter for the error term, and γ, the kernel parameter, with searching range of 1-512 for C and 0-8.00 for γ. The implemented hierarchical SVM classification in ORANGE is shown in Fig. 4. Table 2 compares the results from conventional (Flat SVM) of both full feature and AFS feature data, and hierarchical (H-SVM) method, classification accuracy computed from testing data set are given here. The generated SVM parameters from automated parameters search function for all experiments and data sets are given in table 1. The performance of Flat SVM gives perfect classification except in two cases.The proposed H-SVM classification returns 100% accuracy in all cases. Number of support vectors (no.SVs) extracted from classification models indicate the computation load and results demand for embedded system. The example of feature map plots of the Wine data set scattered by using a multidimensional scaling technique in Fig. 5 applied with data from three different cases: full feature, AFS features and global class-P vector. The plot of global class-P vector, the color represents different classes, are well separated with substantial boundary and realized capabilities of the proposed approach.

Table 1: Generated SVM parameters from automated parameters search function

Experiment & Feature	Parameter	Data set					
		Soy	Beer	Wine	Oil	Used Oil	Glycol in wine
Flat SVM Full	C	2	128	32	512	512	512
	γ	0.03125	0.125	0.125	0.125	0.5	2
Flat SVM AFS	C	512	512	512	8	128	512
	γ	0.03125	0	0.03125	0	2	0.5
H-SVM	$C_{\mathbf{Color}}$	512	32	128	512	128	512
	$\gamma_{\mathbf{Color}}$	0	0	8	0.5	2	0.5
	$C_{\mathbf{Mag.}}$	32	128,	512,	512	512	128
	$\gamma_{\mathbf{Mag.}}$	0.5	0	0.125	0.125	2	0.5
	$C_{\mathbf{Phase}}$	128	512	512	8	8	512
	$\gamma_{\mathbf{Phase}}$	0	0.03125	0	0	2	0.03125
	$C_{\mathbf{Final}}$	0.5	2	512	8	512	8
	$\gamma_{\mathbf{Final}}$	0.125	0.125	0	2	2	0.125

Table 2: Experimental results

Data set	Flat SVM (Full features)		Flat SVM with AFS features		H-SVM	
	CA.	no.SVs	CA.	no.SVs	CA.	no.SVs
Soy	100	31	100	15	100	70
Beer	98.33	60	98.33	36	100	112
Wine	99.05	94	98.10	15	100	203
Oil	100	45	100	17	100	60
Used Oil	100	41	100	7	100	62
Glycol in Wine	100	20	100	2	100	45

DOI 10.5162/AHMT2014/4.3

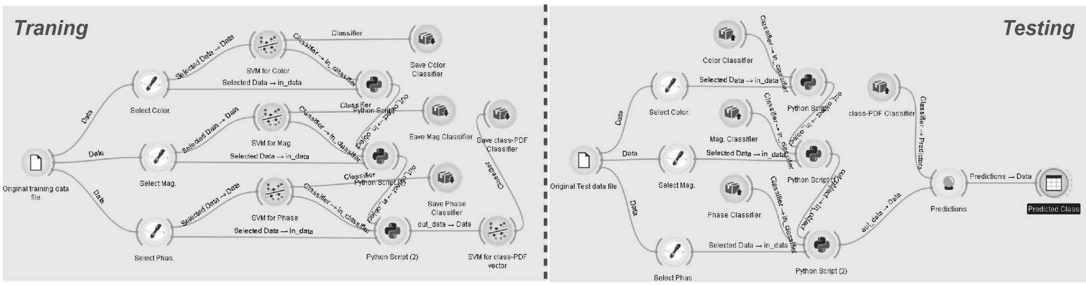


Figure 4: The proposed hierarchical SVM classification in ORANGE

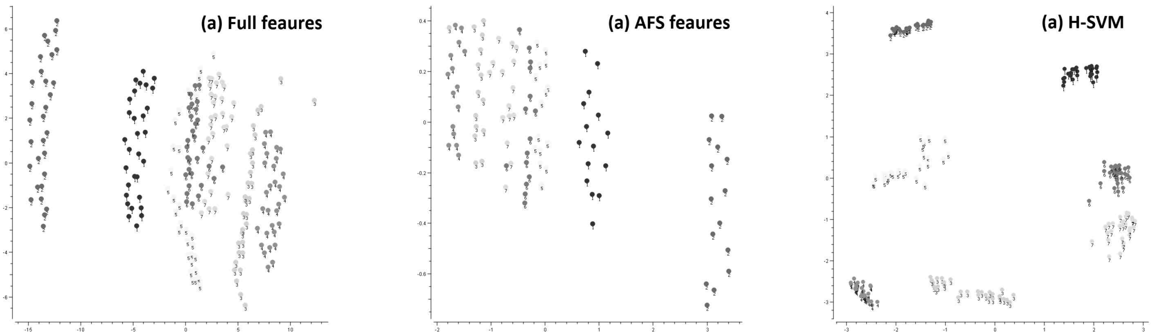


Figure 5: Feature map plot of the Wine data set employing distance preserving dimensionality reduction and interactive visualization

4 Conclusion

In our work, we presented ongoing research on an merging methodology and framework for the automated design of intelligent integrated systems. A strong focus, in contrast to other existing work, is on multi-sensor issues and implementation constraints, e.g., power, speed, size, robustness etc., which are soft constraints in our design optimization to achieve lean yet adaptive or self-x integrated system solutions. In this paper, we apply the current state of our design methodology and system on ORANGE to the multi-sensor Lab-on-Spoon demonstrator, partially automating the design of an improved, hierarchical sensor signal processing and SVM-based recognition system for several task related to food processing, analysis, and safety. We could achieve with the hierarchical classification approach competitive and in some cases even superior results to the previously applied flat approach. In all cases, 100$\%$ recognition rate. The price tag is an substantial increase in parameter finding and training time as well as increased of overall support vectors. In future work, we will combine AFS and related dimensionality reduction techniques with the presented hierarchical approach to reduce the number of support vectors, and, thus, the overall system size or complexity and integrate the required, possibly nested or hierarchical optimization, in our emerging tool architecture and framework.

References

[1] Lab on Spoon Multi-Sensorial 3D-integrated Measurement System for Smart-Kitchen and AAL Applications, `http://www.eit.uni-kl.de/koenig/gemeinsame_seiten/projects/LabonSpoon.html`, last visited: 4.7.2014.

[2] K. Iswandy und A. König, "Methodology, algorithms, and emerging tool for automated design of intelligent integrated multi-sensor systems", *Algorithms* **2**(4), S. 1368–1409, 2009.

[3] A. König, E. Michael und W. Robert, "Quickcog self-learning recognition system - exploiting machine learning techniques for transparent and fast industrial recognition system design", *Image Processing Europe* **Sept./Oct**, S. 10–19, 1999.

[4] S. Peters, *Automatischer Entwurf und Autokonfiguration von Bildverarbeitungssystemen für die industrielle Oberflächeninspektion*. PhD thesis, Lehrstuhl Integrierte Sensorsysteme, Fachbereich Elektrotechnik und Informationstechnik der Universität Kaiserslautern, Germany, 2010.

[5] L. Li, B. Thomas und A. König, "Can impedance spectroscopy serve in an embedded multi-sensor system to improve driver drowsiness detection ?", in *Int. Workshop on Impedance Spectroscopy*, S. 48–49, September 2013.

[6] J. Demšar, T. Curk, A. Erjavec, Črt Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik und B. Zupan, "Orange: Data mining toolbox in python", *J. of Machine Learning Research* **14**, S. 2349–2353, 2013.

[7] *Automating the Design of Data Mining Algorithms*, Springer Berlin Heidelberg, 2010.

[8] H. Liu und H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[9] K. Muller, S. Mika, G. Ratsch, K. Tsuda und B. Scholkopf, "An introduction to kernel-based learning algorithms", *Neural Networks, IEEE Transactions on* **12**, S. 181–201, Mar 2001.

[10] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods", in *ADVANCES IN LARGE MARGIN CLASSIFIERS*, S. 61–74, MIT Press, 1999.

[11] C.-C. Chang und C.-J. Lin, "Libsvm: A library for support vector machines", *ACM Trans. Intell. Syst. Technol.* **2**, S. 27:1–27:27, May 2011.