*Quality of Life through Quality of Information*
*J. Mantas et al. (Eds.)*
*IOS Press, 2012*

*doi:10.3233/978-1-61499-101-4-138*

# Scale-Space Methods for Live Processing of Sensor Data

Stein Olav SKRØVSETH [a,1], André DIAS [a,b], Lukas GORZELNIAK [b,c], Fred GODTLIEBSEN [d] and Alexander HORSCH [b,e,f]
[a] *Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway*
[b] *Institut für Medizinische Statistik und Epidemiologie, Technische Universität München, Germany*
[c] *Institute of Epidemiology, HelmholtzZentrum München, Germany*
[d] *Department of Mathematics and Statistics, University of Tromsø, Norway*
[e] *Department of Computer Science, University of Tromsø, Norway*
[f] *Department of Clinical Medicine, University of Tromsø, Norway*

**Abstract.** A temporal scale-space is a vector space spanned by time and a scale parameter, and by constructing the scale-space correctly a causal structure can be imposed on the scale-space. This enables early warning of significant changes in sensor data at an early time, and on any scale. We describe a feasibility study on how to use these ideas for live surveillance of monitoring processes such that important features can be visualized and users warned about changes an early stage. Sensor data from motion sensors on patients with chronic obstructive pulmonary disease are used as the example of such system, where important pattern are found and visualized using significance plots.

**Keywords.** Sensors, scale-space, COPD, chronic diseases, early warning.

## Introduction

Motion sensors and accelerometers are technologies among many others commonly applied in telemedicine and e-Health settings that involve generation of vast amounts of data. In many situations these data are not visualized during the data generation due to the complexity involved, and also due to the limited extent the data are used. Typical uses are for pattern detection purposes, e.g., detecting critical changes of a patient's condition. Another reason is that evaluating the data or visualizing them efficiently and close to real-time is difficult and requires substantial amount of work on behalf of the investigator, whereas the chance of providing valuable insight is limited due to the complexity of visualizing the data. Data-driven feedback mechanisms using data as they are amassed and presenting informative feedback on changes can be valuable for patients, health professionals, and researchers.

Physical activity (PA) plays an important role in the prognosis of Chronic Obstructive Pulmonary Disease (COPD), and a reduction in PA may be an early, though unspecific, symptom [1]. PA has been related to lung function decline and

[1] Corresponding Author. Stein Olav Skrovseth E-mail: stein.olav.skrovseth@telemed.no

linked with an increased risk of hospitalization and mortality in COPD [2]. Using accelerometers provides an objective method for assessment of PA [3]. It has been shown that the level of PA diminishes with the progression of the disease, reaching its lowest level in the most severely ill patients [4].

When analyzing data, including time series data such as that provided by accelerometers, choosing a scale on which to investigate is crucial. Scale is a feature of the world most people are intuitively familiar with, for example when investigating maps on different scales or zooming in on digital images. Researchers almost invariably face decisions that involve choosing a scale, consciously or not, either in study design, data gathering or analysis. Choosing the wrong scale of operation in planning or analysis entails discarding all other scales, and thereby loosing potentially valuable information. Using scale-space as a visualization tool to find important information on any scale has become a well-developed statistical technique, though still relatively rare in practical analysis. SiZer (Significant Zero-crossings of derivatives) was developed as such a display device, and has subsequently been extended to several related ideas [5,6], including efficient computational techniques for two-dimensional data such as high-resolution images [7]. Regarding temporal data SiZer has the drawback that it is a tool for retrospective analysis, i.e., it can only be applied when the data set is complete, and has limited value during data gathering or in monitoring of live processes.

Using temporally adapted scale-space methods on accelerometer output can provide the patient or health professional with important insight into changes in their PA level on multiple scales, and be an indicator of changes that are predictors for how their disease progresses.

## 1. Methods and Data

We restrict analysis to temporal scale-spaces, i.e., vector spaces spanned by time *t* and scale *h*, where we have observations $\{(t_i, y_i), i = 1, \ldots, N\}$ of some parameter $y_i$ at times $t_i$. For a given scale $h$, the smoothed kernel regression curve $\hat{f}_h(t)$ is computed based on standard kernel methods [8]. Using appropriate estimators for the mean and standard deviation, one can find a confidence interval for the derivative with respect to $t$, and conclude that for a given location in scale-space $(t, h)$, the data show a significantly increasing, decreasing, or no trend [5]. If the data in a region of scale-space is too sparse for a normality assumption to hold, the hypothesis test is not performed, and one does not conclude on the sign of the derivative. Doing this procedure for all points in scale-space, one arrives at a map showing regions of interest, and structure on all scales. Since this procedure involves massive multiple testing, the significance level must be corrected for using, e.g., simultaneous quantiles based on an estimate of the number of independent blocks. The method has proven to be a valuable approach in a variety of settings, including analysis of blood glucose in patients with type 2 diabetes [9] or circulatory research [10].

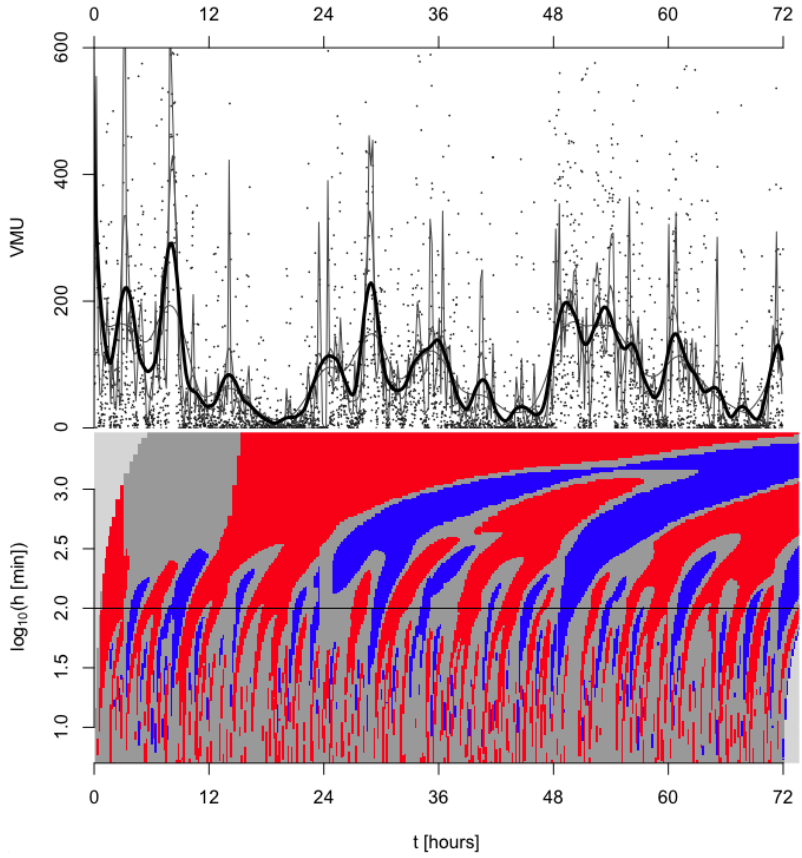


**Figure 1:** *Upper panel:* The sensor samplings for each minute in three days for a COPD patient as dots overlaid by a family plot of kernel smooths with four different bandwidths $h = \{10,50,100,200\}\text{min}$ with $h$ = 100 min highlighted. For clarity, the upper limit of the y-axis is cropped such that some data points are not shown. *Lower panel:* The c-SiZer plot with statistically significant changes color coded as red decreasing and blue increasing activity level. The highlighted bandwidth in the upper panel is indicated by the horizontal line. Light gray areas in the upper left and lower right corners are where the effective sample size is too small for a normality assumption to hold, and dark gray areas have no significant gradient.

The original scale-space methodology as described above does not respect causality in the physical (as opposed to statistical) sense. That is, a data point at time $t_0$ affects the estimate of $\hat{f}_h(t)$ for both past and future values of $t$. Indeed, if the kernel is Gaussian, which is by far the most popular choice, the range of a single data point is infinite into the past and future. However, for live processes, where observations are made sequentially, a future estimate of $\hat{f}_h(t)$ is not meaningful. In order to respect causality in this sense, we shift the kernel back, such that for a given scale $h$, and a kernel with finite support on $[-h,h]$, we map the regression curve $\hat{f}_h(t) \mapsto \hat{f}_h(t-h)$. Thus causality is preserved, and we can infer that any change observed at a time results from the data up to that particular time point. Furthermore, we can impose a causal structure on the vector space such that one can reliably infer at what point in time a change likely originated. We denote these annotated scale-spaces and the associated

causal structure c-SiZer. Though one cannot infer a causal relationship in a statistical sense between events in this manner, the results can be interpreted in light of colluding information from other sources, such that this information is particularly useful when analyzing data originating from several heterogeneous sources. c-SiZer is implemented with an R interface [11], core modules are implemented in C for computational efficiency.

We based our work on data from a previous study, in which patients with COPD who had been admitted for a 3-week in-patient physical recovery program with the diagnosis of smoking-related COPD of stage IV according to the GOLD classification [12]. We processed the data from a male, 65 years old, body mass index 39, classified as having severe COPD. PA level at the hip was assessed using the triaxial RT3 accelerometer (Stayhealthy, Monrovia, CA), worn in a holster at the non-dominant side of the waist. The RT3 is small and lightweight and records activity of the three orthogonal directions as vector magnitude units (VMU) [13] in time intervals of 1 minute.

## 2. Results

In Figure 1 we show the data from the patient's sensors for three full days expressed as VMU along with the corresponding c-SiZer plot in the lower panel. Since the sampling rate is fixed, there is a fixed lower limit on the scale on which the data can be analyzed, which in this case is at $\log h = \log_{10} 5 \approx 0.7$ as we have chosen a required estimated sample size of 5. The number of underlying data points is large and there is much significant variation, and therefore there is much structure in the data that is picked up in the c-SiZer plot. On the scale of $h = 10^2 \text{min} \approx 17 \text{ h}$ there is a clear day/night pattern, while at larger scales a decreasing overall trend is visible. At the smallest scales, interpreting the pattern is difficult without knowledge of the patient's situation, but it is likely to be meaningful as a live information feedback to those close to the situation. On scales between these ranges non-trivial patterns emerge such as regular forks in the significance regions during the day. These types of trends would be very hard to spot investigating the raw data, and requires a specific bandwidth choice if doing kernel regression, and could be easily missed by choosing another bandwidth. Only by using a multiscale approach do such features become apparent.

## 3. Discussion

Sensors have become ubiquitous in our day-to-day living. They provide a useful way to monitor processes, including the activity levels of chronic patients. Clever use of these types of technologies is a way to help patients manage their own diseases, and alleviate a challenged health care system. However, this approach relies on the individual's ability to correctly interpret the presented information and to be able to act upon it. This is far from trivial, but visualization techniques as described offer good chances to overcome the problem. c-SiZer allows for early detection of trends on any scale, such that the patient or anyone monitoring the patient's status can be alerted at the earliest possible time that a significant change has occurred. As is clear from the example, there are many features at the smallest scales that are not very important from this point of view, and might even be obvious to the user without the c-SiZer plot. Choosing an

appropriate range of scales to focus on is important, depending on the application and interest, but this is a far more flexible choice than that of choosing a single bandwidth. While patients obviously cannot be expected to investigate full scale-space plots, these give a statistically sound underpinning to present essential changes in their activity. While we have used raw data to investigate the applicability of c-SiZer plots, one can easily use the same technique on adjusted data, e.g., by adjusting for daily trends we can eliminate the structure that related to these trends to more easily access long-term changes in activity level.

The c-SiZer technique is not limited to sensor data, but can be utilized for a number of settings where non-parametric regression is applicable, even situations with unevenly sampled data, such as self-measured blood glucose values in diabetes patients. By managing the problem of investigating multiple scales simultaneously, one can detect trend changes at an early point, and get a retrospective estimate of the originating time.

## Acknowledgments

This research was performed as part of Tromsø Telemedicine Laboratory (TTL) funded by the Norwegian Research Council under grant no 174934. AD is supported by scholarship SFRH/BD/39867/2007 of the Portuguese Foundation for Science and Technology. LG and AD are supported by the Graduate School of Information Science in Health (GSISH) and the Technische Universität München Graduate School.

## References

[1] Polkey MI, Rabe KF. Chicken or egg: physical activity in COPD revisited. Eur Respir J. 2009; 33(2):227–229.
[2] Baarends EM, Schols AM, Westerterp KR, Wouters EF. Total daily energy expenditure relative to resting energy expenditure in clinically stable patients with COPD. Thorax 1997; 52(9):780–785.
[3] Pitta F, Troosters T, Probst VS, Spruit MA, Decramer M, Gosselink R. Quantifying physical activity in daily life with questionnaires and motion sensors in COPD. Eur Respir J. 2006; 27(5):1040–1055.
[4] Watz H, Waschki B, Meyer T, Magnussen H. Physical activity in patients with COPD. Eur Respir J. 2009; 33(2):262–272.
[5] Chaudhuri P, Marron J S. SiZer for exploration of structures in curves. J Am Stat Assoc. 1999; 94(447):807–823.
[6] Erästö P, Holmström L. Bayesian Multiscale Smoothing for Making Inferences About Features in Scatterplots. J Comput Graph Stat. 2005; 14(3):569–589.
[7] Thon K, Rue H, Skrøvseth SO, Godtliebsen F. Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. Comput Stat Data An. 2012; 56(1):49–61.
[8] Fan J. Design-adaptive nonparametric regression. J Am Stat Assoc. 1992; 87(420):998–1004.
[9] Skrøvseth SO, Godtliebsen F. Scale space methods for analysis of type 2 diabetes patients' blood glucose values. Comput Math Methods Med. 2011; 2011: 672039.
[10] Mortensen KE, Godtliebsen F, Revhaug A. Scale-space analysis of time series in circulatory research. Am J Physiol Heart Circ Physiol. 2006; 291(6):H3012-22.
[11] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
[12] Gorzelniak L, Dias AS, Wittmann M, Schultz K, Karrasch S, Jörres R, Horsch A. Characterization of physical activity (PA) patterns by use of multiple accelerometers in COPD patients undergoing long-term oxygen therapy (LTOT) within a pulmonary rehabilitation (PR) program. European Respiratory Society Annual Congress 2010
[13] Powell SM, Jones DI, Rowlands AV. Technical variability of the RT3 accelerometer. Med Sci Sports Exerc. 2003; 35(10):1773–1778.