A New Linguistic Description Approach for Time Series and its Application to Bed Restlessness Monitoring for Eldercare

Carmen Martinez-Cruz, Antonio J. Rueda, Mihail Popescu, James M. Keller, *Life Fellow, IEEE*

Abstract

Time series analysis has been an active area of research for years, with important applications in forecasting or discovery of hidden information such as patterns or anomalies in observed data. In recent years, the use of time series analysis techniques for the generation of descriptions and summaries in natural language of any variable, such as temperature, heart rate or CO2 emission has received increasing attention. Natural language has been recognized as more effective than traditional graphical representations of numerical data in many cases, in particular in situations where a large amount of data needs to be inspected or when the user lacks the necessary background and skills to interpret it. In this work, we describe a novel mechanism to generate linguistic descriptions of time series using natural language and fuzzy logic techniques. The proposed method generates quality summaries capturing the time series features that are relevant for a user in a particular application, and can be easily customized for different domains. This approach has been successfully applied to the generation of linguistic descriptions of bed restlessness data from residents at TigerPlace (Columbia, Missouri), which is used as a case study to illustrate the modeling process and show the quality of the descriptions obtained.

Index Terms

Time series, linguistic summaries, fuzzy set theory, natural language generation, eldercare, bed restlessness.

I. Introduction

The Generation of Linguistic Descriptions of Time Series (GLiDTS) has been studied by many authors [1], [2], [3]. The aim of GLiDTS is to describe in natural language numerical raw data through the construction of customized summaries where the most relevant information is stressed in a specific context.

The generation of linguistic descriptions makes it possible to provide this summarized information to people that may: i) have little knowledge about a specific area of interest, ii) do not have access to graphic descriptions, such as those with vision problems, or iii) simply have a natural inclination to written language as opposed to other more visual (and possible more technical) representations.

In this paper, we present a novel way of describing time series (TS) using natural language where precision is relaxed to stress different TS characteristics such as trend, volatility, relevant patterns, anomalies, etc. We use fuzzy logic to *compute with words* [4], [5], [6], [7] with these features. A set of experts from a certain domain are needed to define the relevant features and the set of requirements to be accomplished to ensure the quality of the output. The system architecture, which is shown in Figure 1, illustrates the generation process of a linguistic description from a TS.

The proposed method has been successfully applied to the description of real data coming from the TigerPlace residence (Columbia, Missouri). This is an innovative independent living community that is part of the Aging in Place Project, initiated by the Sinclair School of Nursing, specifically to help elder adults age in place in the environment of their choice [8]. The major focus of the research at TigerPlace is to investigate the use of sensors to monitor and assess potential problems in mobility and cognition, detecting adverse health events such as falls or changes in daily patterns that may indicate impending health problems. One of these sensors is a hydraulic bed-sensor [9] that records any fluctuation of mattress

C. Martinez-Cruz and A. J. Rueda are with the Department of Informatica and Centro de Estudios Avanzados en Tecnologías de la Información y de la Comunicación (CEATIC), University of Jaén, Spain e-mail: cmcruz@ujaen.es and ajrueda@ujaen.es

J. M. Keller is with the Department of Electrical Engineering and Computer Science, University of Missouri, USA e-mail: kellerj@missouri.edu

M. Popescu is with the Department of Health Management and Informatics, University of Missouri, USA e-mail:popescum@health.missouri.edu

This document has been downloaded from the Institutional Repository of the University of Jaén. It is an accepted manuscript of an article published in IEEE Transactions on Fuzzy Systems, available on http://dx.doi.org/10.1109/TFUZZ.2021.3052107.

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

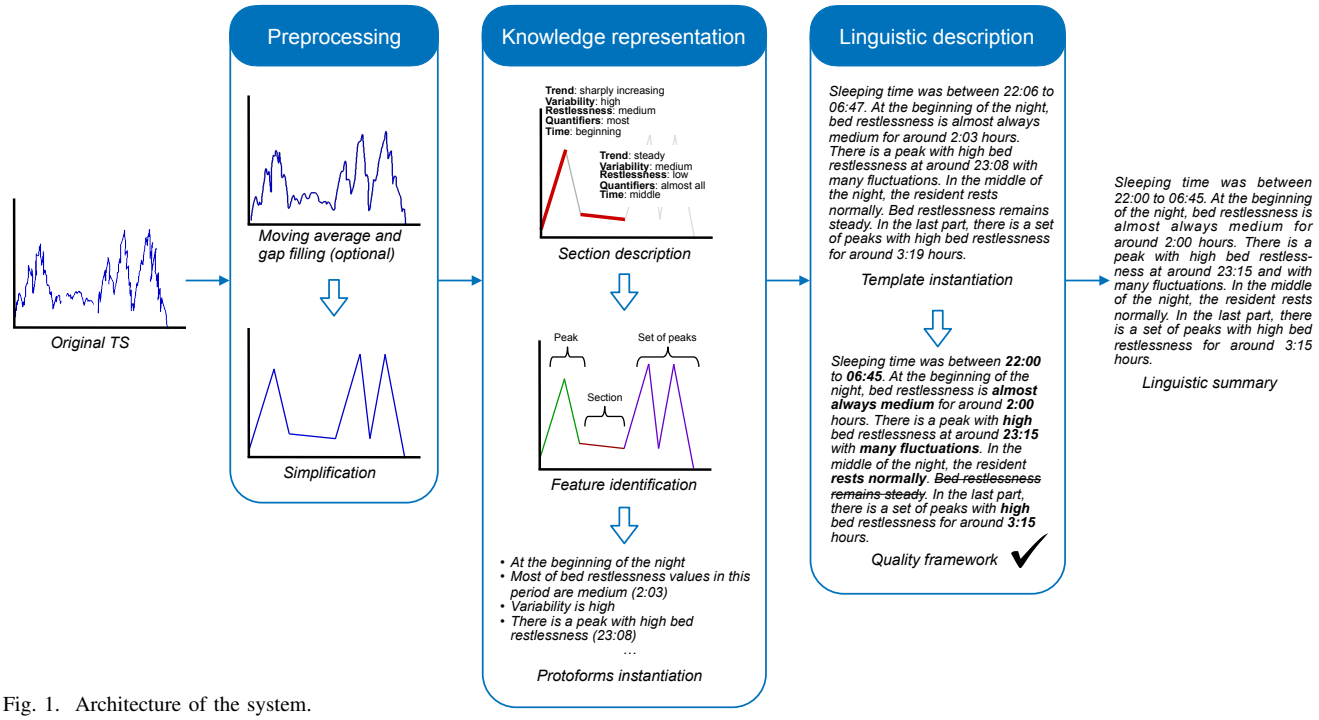


Fig. 1. Architecture of the system.

pressure. Data collected, such as breathing cycles, pulse, or bed restlessness, are processed by the Center for Eldercare and Rehabilitation Technology (ElderTech) at the University of Missouri. The ElderTech brings together researchers from diverse backgrounds, such as Electrical Engineering and Computer Science, Health and Medical Informatics, Nursing, Physical Therapy, Orthopedic Surgery and Social Work with the aim of creating technology to serve the needs of older adults and others with cognitive challenges, helping them to lead healthier and more independent lives. Particularly, *bed restlessness* (BR), the measure that records the movement of a resident during sleep [10], [11], has been used in this paper to provide summaries for two different kinds of users:

- Residents and their families, who would benefit from getting a brief vision of their relative's BR through non-techincal short text messages.
- Nurses caring for residents, that might not be familiar with BR parameters and its graphical description.

The rest of this paper is organized as follows. A brief summary about the state of art of TS linguistic summarization is included in Section II. In Section III and IV the details of our GLiDTS system proposal are described. A real example of generation of summaries from BR data are analyzed in Section V. Finally, the conclusions and future applications of this research are presented in Section VI.

II. Background

The Computing with Words (CW) and Computational Theory of Perceptions (CTP) paradigms proposed by Zadeh [5] have established a framework where words (linguistic expressions in natural language) can be represented and manipulated by a computer using the fuzzy sets theory. Based on these paradigms, Yager [12] proposed the process of linguistic summarization of data by using fuzzy quantified sentences to generate linguistic summaries of one variable. This approach can be used to summarize any kind of data (numeric, text, images, etc.).

GLiDTS applies these and other ideas to describe linguistically the information in a TS, and has been largely studied in recent years, as attested by the surveys of Yager et al. [6], Kacprzyk and Zadrozny [1], Novak et al. [7], [13] and Marín and Sánchez [2]. Some of the most relevant contributors in this field are cited next: Wilbik et al. [14], [15] analyzed the quality of linguistic summaries of TS by the definition of different similarity measures in the health care area among others. Reiter et al. [16] and Ramos-Soto et al. [17] studied the process of summarizing any kind of data using natural language processing, successfully applying their approach to the generation of weather reports. Novak and Perfilieva [7], [18] defined and used the *F-Transform* for the analysis and mining of a TS and the *Fuzzy Natural Logic* for the generation of linguistic descriptions. Kaczmarek-Majer and Hryniewicz [19] applied linguistic summaries for the forecast of TS. Trivino et al. [20] applied the Granular Linguistic Model of a Phenomenon (GLMPL) [21] to generate summaries of streaming data in several application areas such as the gait analysis, energy consumption, etc. Banaee et al. [22] proposed a partial trend

detection algorithm to describe particular changes of health parameters in physiological TS data. Marín et al. [23] developed a mechanism that describes TS where time is expressed in different granularities. In the health field, among other pertinent authors, we can find: Pelaez et al. [24] which evaluated heart rate streams of patients with ischemic heart disease using a linguistic approach and Jain et at. [25] who developed a monitoring system for elderly people that generates linguistic summaries of several health features when an alert of one of these features is activated. This system implements a new method to describe trends in a TS and generates summaries in which all the relevant health features involved in an alarm are compared.

According to Yager [26], there are several basic elements involved in a linguistic description. Two of them are:

Summarizer It refers to a linguistic term or fuzzy predicate (e.g., *low*) defined in the domain of a given attribute (e.g., *bed restlessness*). This concept is also defined in the Fuzzy Natural Language (FNL) theory [27] as *evaluative linguistic expressions*.

Quantity in agreement (or linguistic quantifier) It provides an approximate idea of the number (or proportion) of elements of a subset fulfilling a certain condition. It can be defined by a fuzzy quantifier (e.g., *most*) and a calculation method.

In most of GLiDTS proposals, the representation of the knowledge extracted from the TS has been done by means of *protoforms (prototypical forms)*. Two of the basic protoforms defined by Zadeh [4], [5] are outlined below:

Protoform 1 It has the form A *is* P (e.g., *restlessness is low*, or *volatility is medium*), where P is a summarizer (e.g., *low*) defined on the attribute domain A (e.g., *restlessness*).

Protoform 2 It is a type 1 quantified sentence in the form Q A *are* P (e.g., *Most of the bed restlessness values are low*) where Q is a relative linguistic quantifier [28].

Each protoform has associated a *degree of truth*, which is a measure that represents the validity of a summary or protoform in a range of [0,1].

There are many methods for evaluating linguistic summaries and for calculating the degree of truth of quantified sentences [6]. For instance, Yager [26] presented a way of summarizing numeric and non-numeric data, Sanchez et al. [29] extended Yager's work by defining a method based on the representation of fuzziness through levels, Novak [30] formalized intermediate quantifiers in the Fuzzy Type Theory and Jain and Keller [31] described a way of computing the truth values of linguistic summaries attending to the semantic order of language that they are representing. In this paper, the degree of truth of protoform 2 is computed with the method described by the latter authors.

The contributions of the present work are summarized below. Firstly, we describe an integral GLiDTS system where knowledge modeling and linguistic representation are defined explicitly and in a decoupled manner, in contrast to previous approaches [2]. We propose the use of a TS simplification method, and the description of the resulting sections through a set of summarizers defined by very different fuzzy sets. Another contribution is the definition of a feature identification stage, e.g. high-level features that are meaningful for human users. This provides the generated summaries with more richness and descriptive power.

III. Generation of linguistic descriptions of time series

The process of generating linguistic summaries of TS proposed here follows the general architecture described by Marín and Sánchez [2], and consists of three different phases, as shown in Figure 1. Firstly, TS are decomposed into segments in order to simplify them and to identify their most representative parts. Secondly, the sections resulting from the segmentation are analyzed and described using a set of summarizers. Then high-level features are discovered and described using the summarizers. To complete this phase, a set of protoforms is generated from each feature. In the last phase the protoforms are translated into sentences through a set of linguistic templates. The final summary is obtained after applying the rules of the quality framework that improves its readability and ensures that it meets the needs of the target user.

A. Time series simplification

Many authors have addressed the problem of TS segmentation as an important step to simplify its structure and facilitate its analysis. A review of the most relevant segmentation techniques can be found in the surveys of Höppner [32], Keogh et al. [33] or Fu [34].

First of all, we will define a TS of a variable of interest, such as heart rate, blood pressure, bed restlessness, etc. as: $Y = y_t | t = 1, \ldots, n$, where $y_t \in \mathbb{R}$ is the value of the variable at the timestamp t and n is the total number of observations of the variable. For the sake of simplicity we consider only consecutive, evenly-spaced observations. As is well known, TS are commonly represented graphically by using a 2D Cartesian coordinate system where the horizontal axis represents the time variable and the vertical axis the variable of interest. The points of this graph take the form $P_t = (t, y_t)$ and are typically connected by straight lines.

Here, we implement a piecewise segmentation of the TS by means of a geometrical top-down simplification technique known as Iterative End-Point Fit (IEPF) or Ramer-Douglas-Peucker algorithm [35], [36], later denoted as the Perceptually Importance Points (PIP) technique in [37]. The election of this segmentation method was primarily motivated by its simplicity, its intuitive distance criterion for setting the level of simplification and by the preservation of the most relevant features of the TS. Several of the numerous existing segmentation methods based on piecewise linear approximations may have similar or even better characteristics, and although we tested a few alternatives, a more in-depth study is needed in future work.

This method, described in Algorithm 1, follows an incremental approach that successively adds new segments until the approximation satisfies a maximum distance criterion (defined by a threshold ϵ) with respect to the points of the TS. The algorithm proceeds as follows: first, it finds the furthest point $P_{i^{max}}$ of the TS with respect to the segment defined by its first and last point ($\overline{P_1P_n}$); if the distance to the point is greater than ϵ, it performs two recursive calls processing the TS intervals defined by the first point and $P_{i^{max}}$, and by $P_{i^{max}}$ and the last point, respectively. Otherwise the result of the call is the segment defined by the first and last point of the current TS interval. Figure 2 illustrates the simplification of an example TS with the IEPF algorithm. The threshold ϵ must be chosen according to the domain experts, who verify that the simplification obtained retains the general shape of the TS without discarding any valuable information. The distance point-segment $dist(\overline{P_1P_n}, P_i)$ is computed as the distance from point P_i to its projection on the line defined by segment $\overline{P_1P_n}$:

$$dist(\overline{P_1P_n}, P_i) = \left\| P_i - P_1 + \frac{\overrightarrow{P_1P_i} \cdot \overrightarrow{P_1P_n}}{\left\|\overrightarrow{P_1P_n}\right\|} \overrightarrow{P_1P_n} \right\| \tag{1}$$

where $\overrightarrow{P_aP_b}$ denotes the vector from point P_a to point P_b.

The resulting segments divide the original TS (after its preprocessing) into *sections*. In the next phase of our GLiDTS system, these sections are analyzed to describe the TS semantically.

Algorithm 1 IEPF(P, ϵ)

Input: $P = \{P_1 \dots P_n\}$: a graphical representation of a TS as n points in $\mathbb{R}^2$; ϵ: a distance threshold.
Output: $S_\epsilon(P)$: a subset of points of P that satisfies the distance criterion given by the threshold ϵ.

```
d^max ← 0
for i ← 2 to n − 1 do
    d ← dist(P_1P_n, P_i)
    if d > d^max then
        d^max ← d
        i^max ← i
    end if
end for

if d^max > ε then
    S_ε(P) ← IEPF(P_1 ... P_{i^max}, ε) ⋃ IEPF(P_{i^max} ... P_n, ε)
else
    S_ε(P) ← P
end if
```

B. Knowledge representation model

The *knowledge representation model* of the system defines the semantics of the description by means of protoforms or other template instantiations. This model, described by Marín and Sánchez [2] as the first approximation to a TS description, is based on the Computing with Words and Perceptions paradigm [4] that uses the fuzzy set theory to compute the degree of truth (validity) of a summary.

Our knowledge representation model uses protoforms of type 1 and type 2 (defined in Section II) to describe the sections obtained after the simplification of the TS in the previous phase.

Definition of summarizers: The knowledge of our system is adapted to a specific domain by the instantation of a set of protoforms, such as, *Temperature is low* or *Most blood pressure values in this period are very high*. To do that, a set of summarizers (or linguistic terms) has been defined to describe each of the sections of the simplified TS [1]:

Value summarizers They describe the values of the variable represented by the TS in the given section. It depends on the analyzed variable, therefore, in the case of BR we can use the summarizers *low*, *medium* or *high* to describe with words each measure.

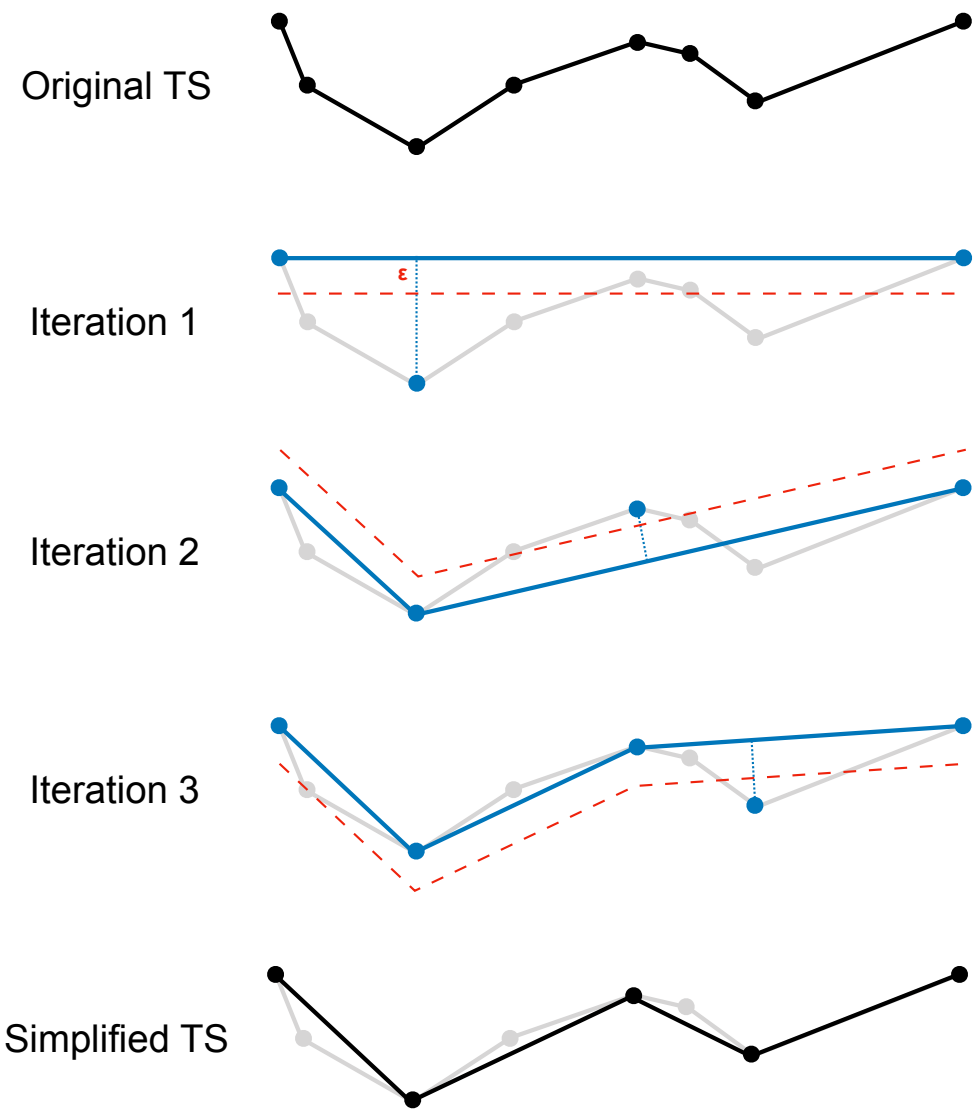


Fig. 2. Example of a TS simplification by Ramer-Douglas-Peucker/IEPF/PIP algorithm.

Local trend summarizers The local trend represents the speed of change of the variable of interest, which is calculated from the slope of the orthogonal regression of the TS points belonging to the period of time comprised by the section. The line obtained by this regression has a similar slope to that of the segment associated with the section, since it uses the same distance criterion as the IEPF algorithm (particularly for a small ϵ threshold). However, similar to how human perception works, it captures the trend of a section of the TS by using all the points within it instead of the first and the last. The most usual summarizers defined for local trend are: *increasing*, *decreasing* or *constant* to describe the slope, or *high* or *low*, to describe its magnitude or steepness.

Volatility summarizers Normally, the trend by itself cannot describe the complete behavior of the TS in the given section. There is an extra component, usually of random nature, that we denote as *volatility* or *variability*. This measures how far are the actual TS values from the trend in a given period, and we estimate it by using a scaled Mean Squared Distance (MSD_ϵ):

$$MSD_\epsilon = \frac{1}{(f-l+1)\cdot\epsilon^2}\sum_{i=f}^{l} dist(\mathbf{R}_{fl}, P_i)^2 \tag{2}$$

where f and l are the indexes of the first and last points of the section, $\mathbf{R}_{fl}$ is the line obtained from the orthogonal

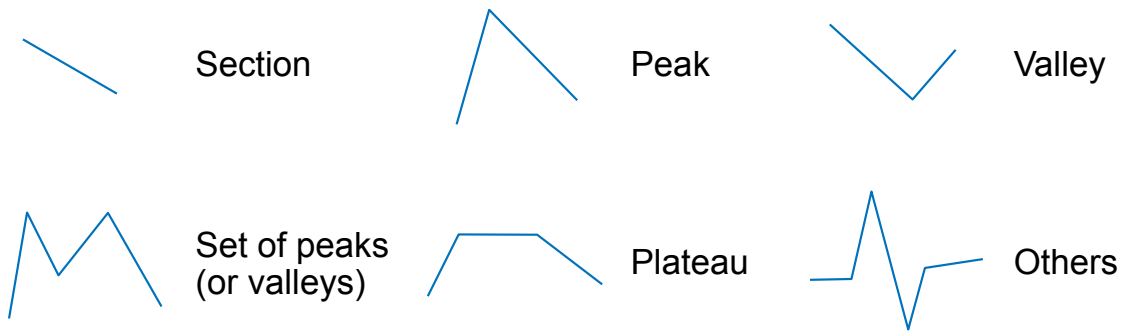


Fig. 3. Common features or patterns in time series.

regression of the points $P_f, P_{f+1} \dots P_l$ and ϵ is the threshold used in the IEPF algorithm. The distance $dist(\mathbf{R}_{fl}, P_i)$ can be computed using equation 1 by taking a segment defined by two arbitrary points of line R_{fl}. Notice that the MSD without the scaling is similar to the Mean Squared Error (MSE) but using an orthogonal distance instead of a simple difference between an estimated value and its actual value. The MSD_ϵ gives values between 0 (the trend perfectly matches the values) and 1 (the values of the TS loosely follow the trend but are far from their expected values).

Time period summarizers The time variable of the TS can be organized into several flexible time periods, such as the *first*, *second*, *third* or *fourth* quarter, or *the beginning*, *the middle* or *the end*. Each section is associated with one or more of these periods.

We also define a **fuzzy quantifier** (denoted henceforth simply as quantifier) to describe the amount of samples of the TS that fits into a given summarizer. Many quantifiers have been proposed (see Section II). For relative quantifiers, which measure the number of elements that satisfy a condition in relation to the total number of elements, we can find several examples, such as: *about a half* of the balls are big [31], or *many* balls are big [14], where *about a half*, *many* or *a few* are quantifiers.

Each summarizer and quantifier is usually modeled by a fuzzy set defined over the particular domain. This is illustrated in Section IV-C for the domain of bed restlessness.

Feature identification: Up to this point, the process of knowledge extraction remained as a simplification and characterization of the TS at the geometric level. In the context of our problem, when experts observe a graphical representation of a TS, they tend to identify common features that provide rich high-level semantic information and a compact description of evolution of the variable (e.g., peaks, valleys, flat areas). Higher-level features can be identified as *patterns*, and their importance for the human mind, to encode and integrate perceived information, has been recognized [38]. These high-level features or patterns are used for decision-making and to transfer information to other individuals. The identification of these features from the geometric description of the TS is essential for the generation of quality linguistic descriptions from a higher level perspective.

Some features that can be used to summarize the shape of a TS are illustrated in Figure 3. It is important to notice that the domain of the problem establishes the set of patterns that are significant. For example, in TS of stock market data, both peaks and valleys are relevant and provide useful information for interpreting data or predicting its evolution. However, in the BR context only peaks, which represent episodes of increased unrest in bed, are relevant, as we will see in Section IV-C.

Strictly speaking, a section cannot be considered a relevant feature if we adhere to the definition given in Section III-A as a piece of a TS segmentation. However, in order to simplify the process of knowledge representation we treat sections as simple or special features whose main purpose is to serve as basis for the discovery of higher-level features and patterns in an iterative process. The identification of these new features is driven by very simple rules. For instance, a *section* feature with a trend described as *increasing* followed by another one *highly increasing* can be replaced by a simple *section* feature described by a summarizer (*increasing* or *highly increasing*) that has to be calculated. This allows consolidating consecutive sections with similar slopes and it is useful in many application domains where it is common for users to focus more on the general trend of a set of consecutive sections than on their particular slopes. Similarly, a *section* feature whose trend is described as *highly increasing* followed by another that is *highly decreasing* identifies a peak in the TS. This new high-level feature replaces the two previous sections. Then in the next iteration two or more consecutive *peak*

features can be consolidated into a *set of peaks* pattern, and so forth. This process finishes when no more features can be discovered in the next iteration.

Once a feature has been identified, it is described using the summarizers of the previous section as follows. i) The time period is reevaluated as the total time span of the sections which form it, ii) volatility is recalculated as the average of the volatility of its sections. For example, in the case of a peak, it is calculated as the mean of the volatility values of its two associated section features. iii) The value summarizer is quantified using all the TS points in the time period of the feature to highlight their predominance, and iv), in the case of high-level patterns, additional summarizers or descriptors can be calculated, such as the maximum value for a *peak* feature, or the number of peaks in a *set of peaks* feature.

The result of this phase is a compact, high-level description of the TS that resembles its abstract representation in the human mind.

Protoforms instantiation: In this final step of the knowledge representation phase, each feature identified in the previous phase is translated into the preestablished protoforms, that are instantiated using the associated summarizers. Optionally, different types of patterns can trigger the generation of different groups of protoforms. If required, a set of protoforms describing global features of the TS can also be added: average or highest value reached, duration, etc. The obtained knowledge representation is already a raw linguistic description of the TS.

C. Linguistic expression model

The generation of linguistic expressions is the last phase of the GLiDTS [2] and it is in charge of turning the simple messages represented by the instantiated protoforms into a text that suitably communicates the knowledge to the target user. Here the preferences of the designer of the system, the linguistic context itself and the needs of the target users must be considered. The linguistic expression model uses a set of syntactic templates that translates the protoforms into natural language expressions as natural and close as possible to the language typically used by the practitioners of the domain.

Unfortunately, simply instantiating a set of linguistic templates does not always provide a satisfactory result. A quality framework is required to improve the readability of the description obtained and to ensure that it meets the needs of the receiver [2]. For this purpose we propose a set of selection and transformation rules that determine aspects such as which specific language should be used in the templates, what information is relevant (and therefore, which protoforms should be translated into text), and how the instantiated templates can be combined.

IV. GLiDTS system for bed restlessness

The process of summarizing a TS is highly dependent on the application context, both in the knowledge modeling process, where the relevant features of the TS are discovered, and in the linguistic expression model, that captures the specific communication style of the target user. In this section we describe the case study of the linguistic summarization of the evolution of BR during night time for residents in TigerPlace. The protocol for this research was approved on 07/21/2016 by the University of Missouri Institutional review Board (approval number: 2005938). All subjects signed an inform consent form before being enrolled in the study and the processed data in this paper was completely deidentified.

A. Bed restlessness description

A BR TS represents the pressure fluctuation of four hydraulic bed-sensors installed under a mattress, and serves as indicator of the discomfort or lack of quality sleep of the user during the night. These sensors, which have been developed by the Center for Eldercare and Rehabilitation Technology at the University of Missouri, collect and process signals from the mattresses in the residents' bedrooms at TigerPlace.

A BR measure represents the maximum pressure read from the four under-mattress sensors, and varies from 0 (meaning no movement) to 15 (maximum level of movement). Data are recorded every 15 seconds, interrupting the recording process between 5-30 seconds to send the data to the database server once each 20 minutes on average. The process stops when the sensor does not detect the presence of a human in the bed.

Figure 4 shows three histograms of BR data, comprising more than ten million records collected over 5 years. BR values have been normalized to [0, 1]. The overall histogram and the histograms of two particular residents have a similar shape, clearly right-skewed, showing a prevalence of low values (between 0 and 0.2). Obviously, for most of the time in bed, residents are in deep sleep, with low restlessness values, while episodes of high restlessness usually concentrate on short periods during the night (e.g., at the beginning) or during certain sleepless nights. Medium/high restlessness episodes, particularly if they continue over time, are relevant and may be a sign of a physical or mental condition.

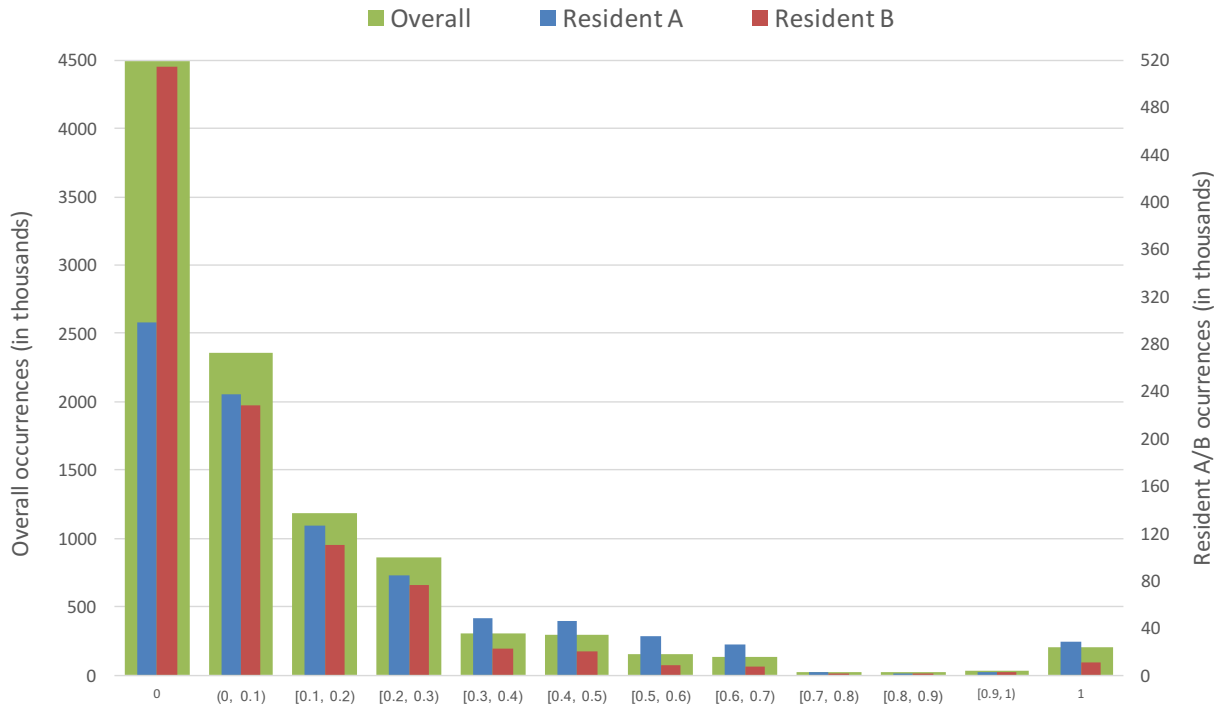


Fig. 4. Histogram of BR data collected at TigerPlace, and individual histograms for two sample residents.

B. Preprocesssing

BR database records require a preprocessing step to fix the gaps in data stream, reduce the number of samples and remove the noise and high-frequency fluctuations of the signal of the BR sensor. This preprocessing follows these phases:

Downsampling As we explained above, during normal operation a BR measure is gathered every 15 seconds. Missing values during the data upload to the server are estimated by using the average of the 4 nearest neighbors. Once we have four values per minute, these are aggregated into a single one to reduce the number of samples of the TS.

Out-of-bed labeling A period of 8 minutes without restlessness activity is considered as *out of bed*. A *NaN* value is assigned to each minute until new values are read.

Smoothing A moving average is calculated to smooth out the noise and emphasize the medium-term trend of the BR TS. The size of the sliding window used is 5.

C. Knowledge representation model

Following a similar organization to that of Section III-B, we first describe the modeling of the summarizers for the BR domain, then the relevant features and finally the protoforms for the representation of knowledge.

Definition of summarizers and quantifiers: Each of the summarizers described in Section III-B is defined for the context of BR through a fuzzy set. The corresponding membership functions have been revised and approved by a group of experts.

Value summarizers After the analysis of the BR database of TigerPlace residents, we have modeled five summarizers with a set of fuzzy sets defined by the trapezoidal membership functions included in Table I[1].

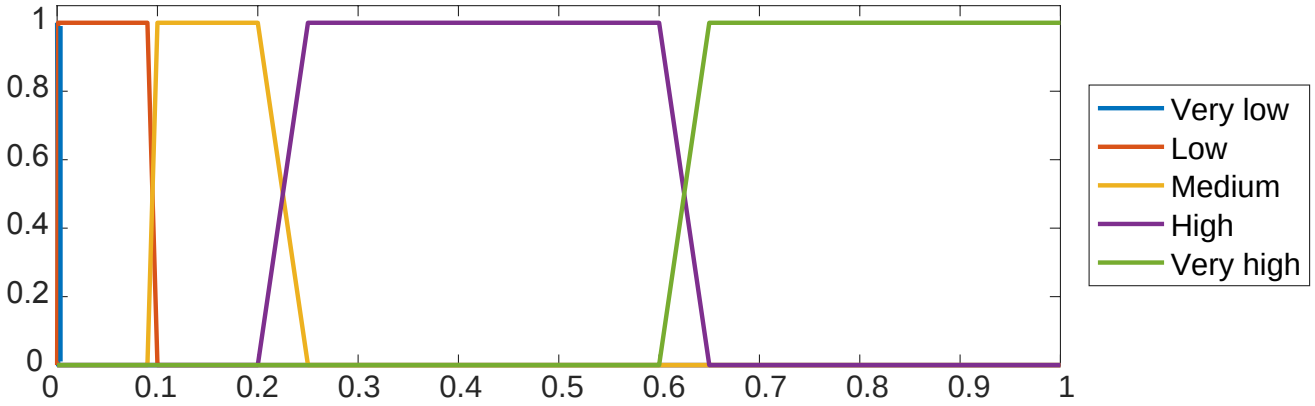
Local trend summarizers In the same context, we have defined 5 summarizers to represent the local trend of the TS, which is described by its slope, or equivalently, by its change per hour. These summarizers and their corresponding membership functions are shown in Table II [1] [2].

Volatility summarizers As we described in Section III-B, this describes how far are the actual TS values from the simplification resulting from the segmentation stage. We have defined four different summarizers for describing

[1] We would like to notice that the terms *sharply* and *very* in Tables I and II are not used in this paper as modifiers (hedges or extensions) defined by Zadeh [39] or Novak [13]. Here, each summarizer has an exclusive representation meaning *approximately* some value.

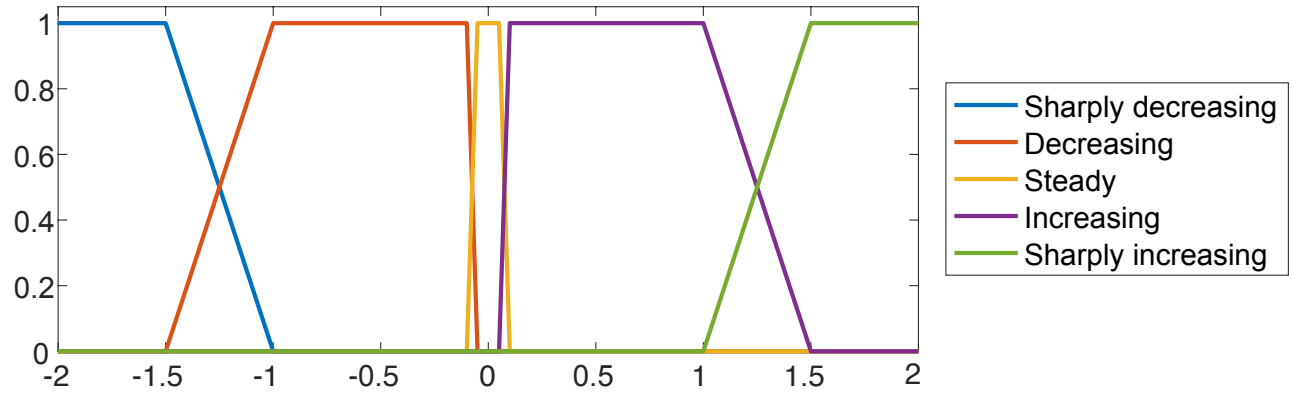
[2]The *Very Low* summarizer can not be appreciated in the graphical depiction due to its values.

TABLE I
DESCRIPTION OF BR SUMMARIZERS.



Summarizer	Membership function	Number of occurrences
Very High	$trap(0.6, 0.65, 1, 1)$	278369
High	$trap(0.2, 0.25, 0.6, 0.65)$	887180
Medium	$trap(0.09, 0.1, 0.2, 0.25)$	2051906
Low	$trap(0.0005, 0.001, 0.09, 0.1)$	2360102
Very Low	$trap(0, 0, 0.0005, 0.001)$	4488737

TABLE II
DESCRIPTION OF LOCAL TREND SUMMARIZERS.



Summarizer	Membership function	Change/hour
Sharply decreasing	$z\text{-}shape(-1.5, -1)$	$(-\infty, -1.5]$
Decreasing	$trap(-1.5, -1, -0.1, -0.05)$	$(-1, -0.1)$
Steady	$trap(-0.1, -0.05, 0.05, 0.1)$	$(-0.05, 0.05)$
Increasing	$trap(0.05, 0.1, 1, 1.5)$	$(0.1, 1)$
Sharply increasing	$s\text{-}shape(1, 1.5)$	$[1.5, \infty)$

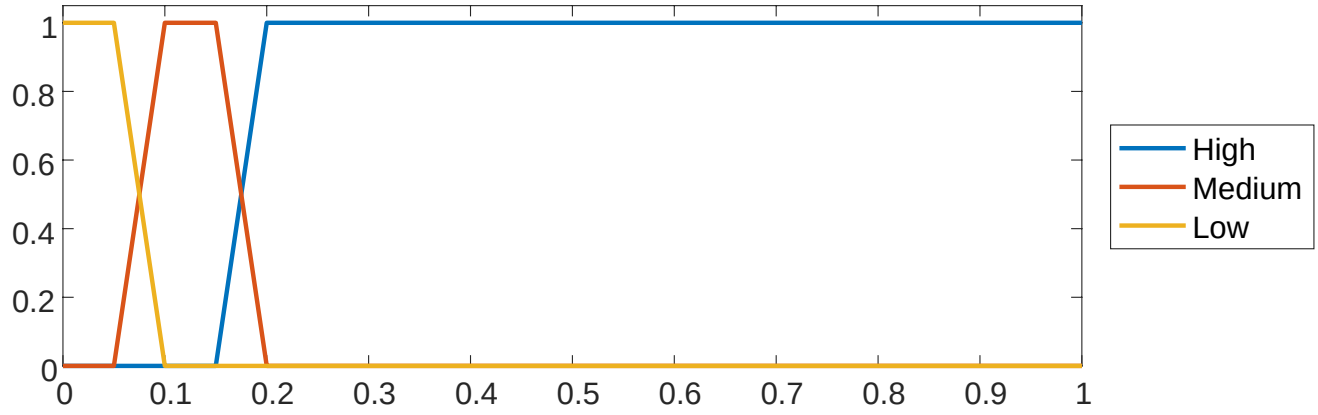
volatility in the context of BR, along with their corresponding membership functions, which are illustrated in Table III. These membership functions are defined from the values of the scaled MSD (Eq. 2).

Time summarizers Using a specific number of hours is of little interest since sleeping habits vary between different residents or even for a resident in different nights. Because of that, we use three summarizers described as fuzzy sets to define approximate percentages of the sleeping time of the resident, as shown in Table IV.

Similarly, quantifiers are described by a set of linguistic labels and membership functions shown in Table V.

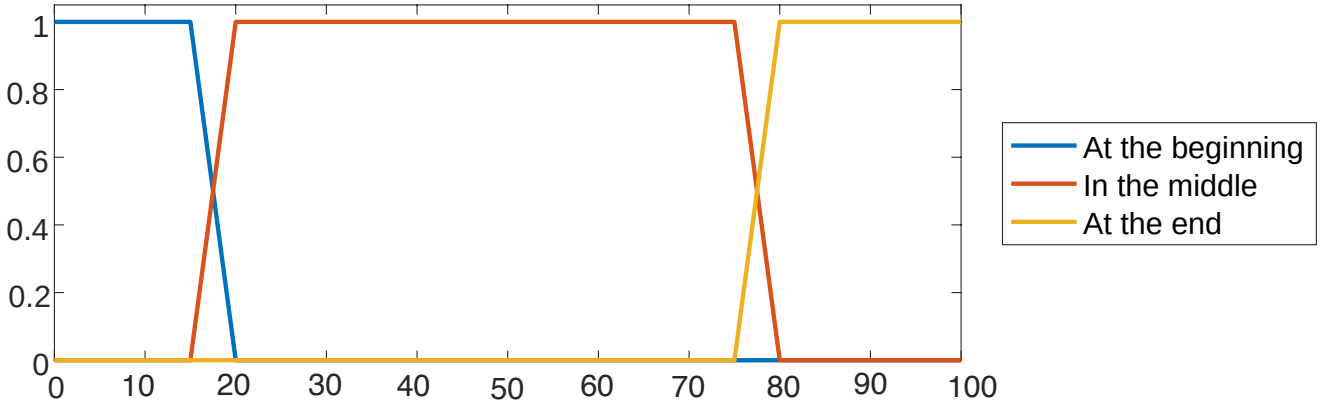
Notice that only descriptions that represent a large number of values are used in our summaries (i.e., *Almost all*, *Most* and *Many*) because they stress the relevant TS information and also, a minimum truth degree of 0.7 is required to keep the resulting quantifier. The quantifiers that describe a few occurrences of a particular value do not provide useful information

TABLE III
DESCRIPTION OF LOCAL VOLATILITY SUMMARIZERS.



Summarizer	Membership function	Scaled MSD
High	$trap(0.15, 0.2, 1, 1)$	$(0.2, 1)$
Medium	$trap(0.075, 0.1, 0.15, 0.2)$	$(0.1, 0.15)$
Low	$trap(0, 0, 0.075, 0.1)$	$(0, 0.075)$

TABLE IV
DESCRIPTION OF TIME SUMMARIZERS.



Summarizer	Membership function	Percentage of sleeping time
At the beginning	$trap(0, 0, 15, 20)$	Around the first 15% of the period.
In the middle	$trap(15, 20, 75, 80)$	Until around 80% of the period.
At the end	$trap(75, 80, 100, 100)$	Around the last 20% of the period.

for the kind of BR summaries we are interested in.

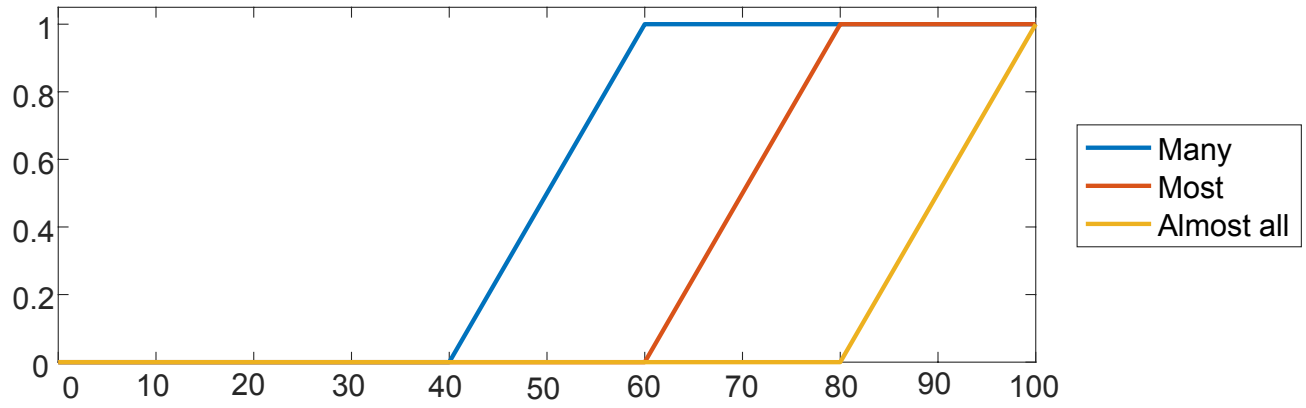
Features: In the context of BR four features have been considered as useful to highlight the most relevant restlessness-related events during the night and make the description of the TS useful for nurses or resident's relatives. These correspond to the *section*, *peak* and *set of peaks* features depicted in Figure 3 plus the *gap* feature.

Gap This is used to represent time periods for which no BR data exists, usually because the resident is out of bed.

Section As we explained in Section III-B, this is the basic feature to describe the shape of a TS, serving as basis to discover higher-level patterns and simplify the TS by consolidating consecutive sections with similar local trends.

Peak Two consecutive *section* features with *(sharply) increasing*, or *(sharply) decreasing* trends respectively, define a moment of maximum restlessness. This is potentially a relevant event that is captured by the *peak* feature and normally reported in the summary. The two sections are required to have semantic distance of at least two consecutive linguistic terms in the trend variable. For example, a *(sharply) increasing* section followed by a *decreasing* section has a semantic distance of two terms: *increasing* and *steady*. (see Table II).

TABLE V
QUANTIFIERS AND MEMBERSHIP FUNCTIONS FOR BR VALUES.



Quantifier	Membership function	Number of occurrences
Almost all	$trap(80, 100, 100, 100)$	Almost 100%.
Most	$trap(60, 80, 100, 100)$	More than 80%.
Many	$trap(40, 60, 100, 100)$	More than 60%.

Set of peaks Two or more successive peaks in a relative short period of time is in itself more significant than the peaks separately, since it is related to repeated rest/wake up episodes during the night, and therefore, low quality sleep. This is captured by this higher-level pattern, that replaces two or more consecutive *peak* features in the representation.

Notice that in contrast with peaks, valley features are in general of minor importance in the context of BR, particularly in the case of wide valleys, since they reflect a normal situation of quality rest during the night.

Protoform generation: The description of the features identified in the previous phase is translated into the protoforms shown in Table VI. Protoform 3.1 is used with normal sections, while protoform 3.2 highlights the presence of a singular feature (only *peak* and *set of peaks*). Gaps are described using protoform 5. Finally a global protoform (protoform 6) has been added to describe the duration of the resident's night rest.

TABLE VI
PROTOFORMS IN THE BR KNOWLEDGE MODEL

N.	Protoform	Summarizers
	Per-segment/section protoforms	
1	*At/In the **[part]** of the night.*	**part**: *beginning, end, middle.*
2	***[quantif]** of bed restlessness values in this period are **[value]**(**duration**).*	**quantif:** *almost all, the majority, many.* **value**:*very high, high, medium, low, very low.* **duration**: duration of the event.
3.1	*Bed restlessness is **[local trend]**.*	**local trend**: *sharply increasing, increasing, decreasing, sharply decreasing, steady.*
3.2	*There is a **[feature]** with **[value]** bed restlessness (**[hour]** / **[duration]**).*	**feature**: *set of peaks/peak* **value**: *very high, high, medium, low, very low.* **hour**: hour of the peak. **duration**: duration of the set of peaks.
4	*Volatility is **[var]**.*	**var**: *high, medium, low.*
5	*Getting out of bed at **[hour]** for the **[num]** time (**[duration]**).*	**num**: 1st, 2nd, 3rd, etc. **duration**: duration of the period.
	Global protoforms	
6	*Sleeping time was between **[start hour]** and **[end hour]**.*	**start hour**: starting hour of the TS. **end hour**: end hour of the TS.

D. Linguistic expression model

All the meaningful protoforms obtained in the previous knowledge extraction phase are processed to generate a final linguistic summary that effectively translates this knowledge to the target user in the BR context. For each protoform we define a corresponding syntactic template, shown in Table VII, that approaches the target user's language style and needs. These linguistic templates are defined according to our quality assurance model for the context of BR. Our quality framework also defines the selection and transformation rules outlined in Table VIII, following the guidelines provided by the nurses at TigerPlace and from the Eldertech team. Basically, these rules stress the following aspects:

TABLE VII
BR LINGUISTIC TEMPLATES.

N.	Template	Description
1	At/In the **[part]** of the night, /After that, /Then, /Later, /Next, /Afterwards, /Subsequently, /In the last part,	**part**: *beginning, middle, end*. Description of a period (beginning or continuation using different connectors).
2	(a) bed restlessness is **[quantif] [value]** for around **[duration]**[.] (b) the resident rests normally[.]	(a) **quantif**: *almost always, mostly, very often*. **value**: *very high, high, medium*. **duration**: duration of the value. BR is expressed as an uncountable quantity. Protoforms are transformed as follows: *almost all* $\rightarrow$ *almost always*, *most* $\rightarrow$ *mostly*, *many* $\rightarrow$ *very often*. (b) Use when **quantif** is *low* or *very low*.
3.1	(a) {with a(n)} / {bed restlessness has a(n)} **[local trend]** trend[.] (b) and steady/bed restlessness remains steady[.]	(a) **local trend**: *sharply increasing, increasing, decreasing, sharply decreasing*. (b) Only when **local trend** is *steady*. One of the two forms is used to describe *section* features depending on whether or not template 2 has been instantiated.
3.2	(a) There is a **[feature]** with **[value]** bed restlessness (b) for around **[duration]**[.] (c) at around **[hour]**[.]	(a) **feature**: *set of peaks/ peak*. **value**: *very high, high*. (b) **duration**: duration of the *set of peak* value. (c) **hour**: time of the *peak*.
4	(a) with/and many fluctuations. (b) with/and some fluctuations	(a) Only with *high* volatility. (b) Only with *medium* volatility. The connector *with* is used if it has not been used before.
5	The resident got out of bed **[num]** times at around **[hour1]** (**[duration1]**), **[hour2]** (**[duration2]**), ...	**num**: number of times the resident got out of bed. **hourn**: approximate time of the nth occurrence of a patient getting out of bed. **durationn**: duration of the nth occurrence of a patient getting out of bed.
6	The sleeping time was between **[h]** and **[h]**.	**h**: time of beginning/end of the user rest time.

- Use non-technical natural terms (r. 1 and 4).
- Descriptions should focus on medium to very high BR since they correspond to abnormal sleep patterns (r. 2).
- Features such as peaks and sets of peaks are relevant sleep patterns that should be highlighted (r. 3).
- Avoid irrelevant information (r. 5 and 6).
- Sentences should be properly combined and connected to produce a fluent text (r. 8-10).

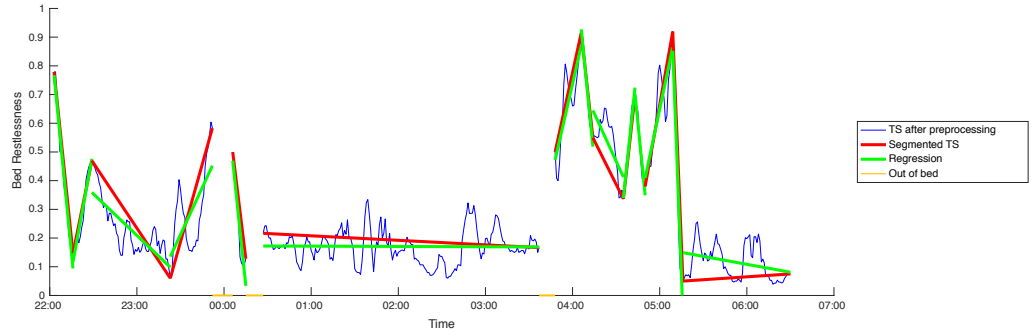
TABLE VIII
RULES OF THE QUALITY FRAMEWORK FOR BR TS

N.	Description
1	Use *bed restlessness* (uncountable) instead of *BR values* (countable) in the final description. Transform the quantifiers in the protoform to an uncountable counterpart (template 2a).
2	The description should emphasize abnormal sleep episodes (i.e., *medium, high* or *very high* BR). For *low* or *very low* BR use the default message given by template 2b.
3	Isolated *peaks* and *sets of peaks* (template 5) are relevant events during the night but only if their associated BR values are *high* or *very high*.
4	Only *high* volatility is relevant. Use *with many fluctuations* instead of the more technical expression *with high volatility* (template 4).
5	Discard descriptions of short features (less than 2 minutes of BR data).
6	Time and duration should always be shown in format hh:mm and hh***h***:mm***m*** respectively and be rounded to the closest quarter for the templates 2, 3.2, 5 and 6. Template 5 shows the exact duration of the time out of bed in minutes.
7	A brief summary of the number of times that a resident has got out of bed, including time and duration, should appear at the beginning of the description, up to a maximum of five. A larger number must be replaced by simply *many times* (template 5).
8	Connectors between sentences must change from one sentence to the next. Several levels of connectors are established to avoid repeating them. Description of the last feature starts with *In the last part* (template 1).
9	Consecutive sentences with a similar description are combined. For instance, two consecutive sentences in the form: *At the beginning of the night bed restlessness is medium and steady* and *In the middle of the night bed restlessness is medium and steady* are merged into the sentence *From the beginning to the middle of the night bed restlessness is medium and steady*. If volatility is different, the sentence is merged anyway but volatility is established to the maximum.
10	When a statement qualifies a previous one in an opposite sense, connect both using the conjunction *but*. For instance: *Bed restlessness is mostly high with a decreasing trend.* is rearranged as *Bed restlessness is mostly high but with a decreasing trend*. There are three cases where this rule is applied: • High BR with [sharply] decreasing trend. • Medium BR with [sharply] increasing trend. • Steady trend with many fluctuations.
11	Do not describe trend or volatility with low restlessness (use only template 2b).
12	Bed Restlessness is written only in the first sentence together with its acronym as *bed restlessness (BR)*. The remaining cases it will be replaced by the acronym BR.
13	Volatility is not described with *peaks* and *sets of peaks*.

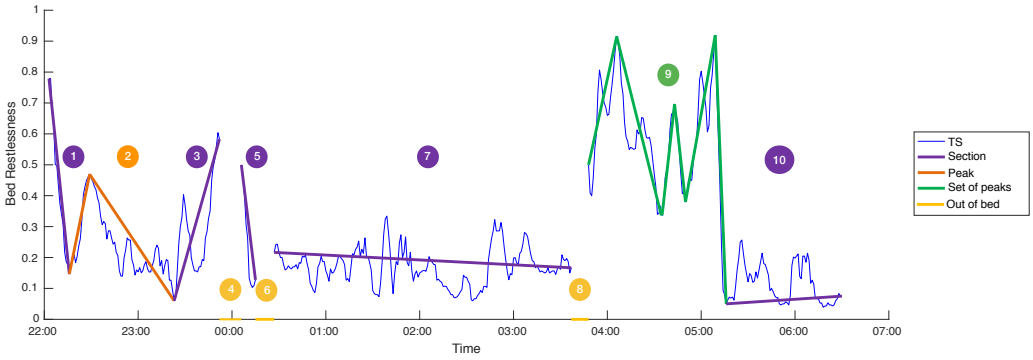
V. Experimental results

We have tested our bed restlessness GLiDTS system with real data from TigerPlace. In this section we illustrate how the system works by describing in detail an specific case study with the BR data of a resident during the night of June 15 to 16, 2014. After that, an expert assessment has been included on these results.

The original BR data spanned 8 hours and 23 minutes of the resident's rest, and consisted of 1834 samples, that were reduced to 508 after the preprocessing (see Section IV-B). The TS was segmented with the IEPF algorithm using as threshold ϵ the value 0.25. Figure 5 (a) illustrates the TS after its preprocessing and the result of the IEPF algorithm, which reduced it to 18 samples and 14 segments. This figure also compares these segments with the orthogonal regression of the points of each of the sections. The resident was out of bed three times, that are highlighted in yellow. Figure 5 (b) shows the identified features, namely 3 gaps, 5 ordinary sections, a peak, and a set of peaks.



(a) TS after its preprocessing, segmented TS, orthogonal regression of the points in each section and time out of bed.



(b) Features identified in the TS.

Fig. 5. BR data of a resident for the night of June 15-16, 2014.

The protoforms generated for the features shown in Figure 5 (b) are given in Table IX. Therewith, the information in the protoforms is translated into the linguistic templates, producing the following linguistic summary for the TS:

Sleeping time was between 22:00 and 06:30. The resident got out of bed 3 times at around 23:45 (00h14m), 00:15 (00h12m) and 03:30 (00h11m). At the beginning of the night, bed restlessness (BR) has a sharply decreasing trend. After that, there is a peak with high BR at around 22:30. Then, BR has an increasing trend with many fluctuations. In the middle of the night, BR has a sharply decreasing trend. After that, BR is very often medium for around 3h00m and steady. At the end of the night, there is a set of peaks with high BR for around 01h30m. In the last part, BR remains steady.

A quantitative analysis of the outputs of the different stages of our system for the same TS using an increasing threshold (ϵ) in the segmentation algorithm is shown in Table XI. It is clear from the table that the number of protoforms and the length of the generated summary are determined by this threshold. In the BR context we concluded that a value of 0.25 for the threshold ϵ is appropriate, taking into account the quality criteria defined in the model, together with the descriptiveness of the segmentation and the level of abstraction and length of the resulting summaries.

TABLE IX
PROTOFORMS GENERATED FROM THE BR TS OF THE RESIDENT DURING THE NIGHT OF JUNE 15-16, 2014.

Feature	Protoforms
1	*At the beginning of the night. Bed restlessness is sharply decreasing. Volatility is low.*
2	*At the beginning of the night. There is a peak with high bed restlessness (22:29). Volatility is low.*
3	*At the beginning of the night. Bed restlessness is increasing. Volatility is high.*
4	*In the middle of the night. Getting out of bed at 23:52 for the 1st time (00:14).*
5	*In the middle of the night. Bed restlessness is sharply decreasing. Volatility is low.*
6	*In the middle of the night. Getting out of bed at 00:15 for the 2nd time (00:12).*
7	*In the middle of the night. Many of bed restlessness values in this period are medium (3:01). Bed restlessness is steady. Volatility is medium.*
8	*In the middle of the night. Getting out of bed at 03:36 for the 3rd time (00:11).*
9	*At the end of the night. There is a set of peaks with high bed restlessness (01:28). Volatility is medium.*
10	*At the end of the night. Bed restlessness is steady. Volatility is medium.*
Global	*Sleeping time was between 22:03 and 06:29.*

TABLE X
RESULT OF APPLYING THE QUALITY FRAMEWORK TO THE LINGUISTIC DESCRIPTION GENERATED FROM THE PROTOFORMS IN TABLE IX.

Rule	Res.	Description
1	✓	*Very often* is used instead of *many* for feature 7.
2	✓	Feature 7 emphasizes *medium* values of BR.
3	✓	All the peaks have *high* values.
4	✓	It has been applied to feature 5.
5		n/a.
6	✓	It has been applied to all the sentences that include hours or duration but the instantiation of template 5 that include the precise duration of the time out of bed.
7	✓	It has been done. There are less than 5 gap/out-of-bed features.
8	✓	Connectors do not repeat after two successive sentences and the last sentence has been described as *In the last part*.
9, 10, 11		n/a.
12	✓	All bed restlessness appearances have been replaced by BR with the exception of the first one.
13	✓	The volatility of the set of peaks has not been shown.

Ensuring the interpretability of the linguistic summaries generated by a GLiDTS is a major concern [40]. Our approach helps this interpretability through a modular architecture that is comprehensible to the end user [41] . The system generates a set of simple, meaningful and traceable sentences through the execution of three transparent and well-structured phases: The first phase involves the fragmentation of the TS in a set of relevant segments; in the second one, the segments are analyzed and their main features are characterized using protoforms; and finally, these are rendered into a set of sentences that follow some linguistic rules tailored to the needs of the end users. The set of simple assumptions and rules that govern these three phases have been approved by our final users.

Validation of the results

A battery of linguistic summaries generated by our system was presented to the nurses working at TigerPlace and from the ElderTech team getting a positive assessment. The general consensus was that they capture the essential information of the BR of a resident during the night, and that it is a useful tool for caregivers working at TigerPlace. However, it is pending to do a more rigorous and thorough evaluation in a close future.

VI. CONCLUSIONS

In this work we propose a novel approach that follows the general GLiDTS architecture described in [2] to generate quality summaries of TS. Two different models have been defined: the knowledge representation model and the linguistic generation model. Both of them are used to extract the knowledge of a TS and generate the final linguistic summary with a quality that fulfills the user requirements.

When experts examine a TS, they usually focus on the most relevant characteristics to extract the valuable information (local trends, singular events, singular features, etc.). Aiming to imitate this human behavior, we propose the use of the IEPF algorithm to simplify and reveal the basic structure of the TS, and we define a knowledge representation model that includes a novel mechanism to describe a TS linguistically by translating relevant features into words. In order to do that, several fuzzy linguistic variables have been defined, modeled and used: local trend, volatility, time periods, and their aggregated values. A set of summarizers and membership functions has been defined in an specific context and used in the generation of the linguistic description. Moreover, several high level features have been defined to better capture

TABLE XI
QUANTITATIVE ANALYSIS OF THE RESULTS OBTAINED FOR DIFFERENT TS SIMPLIFICATION LEVELS.

Theshold (ϵ)	**N. segments**	**N. sections**	**N. peaks**	**N. set of peaks**	**N. protoforms**	**N. Words**
0.15	22	6	2	1	12	140
0.25	*14*	*5*	*1*	*1*	*10*	*114*
0.35	10	5	0	1	9	98
0.45	7	7	0	0	10	96
0.55	7	7	0	0	10	96
0.65	4	15	0	0	7	75

remarkable or repetitive aspects of the TS that can be described using natural language in a more accurate and simpler way.

Our approach has been successfully applied to the description of bed restlessness data collected from TigerPlace residents (Columbia, Missouri). The nurses of the TigerPlace staff and the researchers of ElderTech (University of Missouri) have contributed to generate, analyze and provide the expert knowledge modeled in our system.

As future work we aim at improving the linguistic representation model in order to generate better linguistic descriptions. This implies replacing the simple template-based approach with one based on natural language processing techniques. We also plan to use our knowledge representation for TS not only for generation of linguistic summaries but for TS qualitative comparison. We believe that the flexibility and descriptive power of words can help to compare several TS and describe their similitude and differences linguistically.

ACKNOWLEDGMENT

The protocol for this research was approved on 07/21/2016 by the University of Missouri Institutional review Board (approval number: 2005938). All subjects signed an inform consent form before being enrolled in the study and the processed data in this paper was completely deidentified.

We would like to thank Daniel Sánchez and Nicolas Marín, professors in DECSAI department at Granada University (Spain) for their advice and support for describing our system using their general approach for GLiDTS [2].

Funding for this research is provided by EU Horizon 2020 Pharaon Project *Pilots for Healthy and Active Ageing*, Grant agreement no. 85718 and the Spanish Ministry of Science, Innovation and Universities and the European Regional Development Fund - ERDF (Fondo Europeo de Desarrollo Regional - FEDER) under project PGC2018-096156-B-I00 *Recuperación y Descripción de Imágenes mediante Lenguaje Natural usando Técnicas de Aprendizaje Profundo y Computación Flexible*. Moreover, this contribution has been supported by the Andalusian Health Service by means of the research project PI-0387-2018 and the *Action 1 (2019-2020)* no. EI_TIC01 of the University of Jaén. Finally, this research has been also backed by the National Library of Medicine (NLM) funded project (Popescu, NIH-NLM #R01LM012221).

REFERENCES

[1] J. Kacprzyk and S. Zadrozny, "Fuzzy logic based linguistic summaries of time series: a powerful tool for discovering knowledge on time varying processes and systems under imprecision," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 6, no. 1, pp. 37–46, 2015.
[2] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets and Systems*, vol. 285, pp. 6 – 30, 2016, special Issue on Linguistic Description of Time Series.
[3] V. Novak, I. Perfiljeva, and A. Dvorak, *Insight into Fuzzy Modeling*, 05 2016.
[4] L. A. Zadeh, "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, pp. 103–110, 1996.
[5] ——, "From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," in *Intelligent Systems and Soft Computing: Prospects, Tools and Applications*, 2000, pp. 3–40.
[6] F. E. Boran, D. Akay, and R. R. Yager, "An overview of methods for linguistic summarization with fuzzy sets," *Expert Systems with Applications*, vol. 61, pp. 356–377, 2016.
[7] V. Novák, "Linguistic characterization of time series," *Fuzzy Sets Syst.*, vol. 285, pp. 52–72, 2016.
[8] M. J. Rantz, R. T. Porter, D. Cheshier, D. Otto, C. H. Servey, R. A. Johnson, M. Aud, M. Skubic, H. Tyrer, Z. He, G. Demiris, G. L. Alexander, and G. Taylor, "TigerPlace, A State-Academic-Private Project to Revolutionize Traditional Long-Term Care." *Journal of housing for the elderly*, vol. 22, no. 1-2, pp. 66–85, 2008.
[9] D. Heise, L. Rosales, M. Skubic, and M. Devaney, "Refinement and evaluation of a hydraulic bed sensor," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, Aug 2011, pp. 4356–4360.
[10] G. Demiris, M. Skubic, J. Keller, M. J. Rantz, D. Parker Oliver, M. A. Aud, J. Lee, K. Burks, and N. Green, "Nurse participation in the design of user interfaces for a smart home system," in *Proceedings of the International Conference on Smart Homes and Health Telematics, NI Belfast, Editor*, 2006, pp. 66–73.
[11] M. J. Rantz, M. Skubic, G. Alexander, M. Popescu, M. A. Aud, B. J. Wakefield, R. J. Koopman, and S. J. Miller, "Developing a comprehensive electronic health record to enhance nursing care coordination, use of technology, and research," *Journal of GerontoloGical nursing*, vol. 36, no. 1, pp. 13–17, 2010.
[12] R. R. Yager, "On linguistic summaries of data," in *Knowledge Discovery in Databases*, 1991, pp. 347–366.
[13] V. Novák, "On modelling with words," *Int. J. Gen. Syst.*, vol. 42, no. 1, pp. 21–40, 2013.
[14] J. Kacprzyk and A. Wilbik, "Temporal linguistic summaries of time series using fuzzy logic," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods*. Springer Berlin Heidelberg, 2010, vol. 80, pp. 436–445.

[15] A. Wilbik, J. M. Keller, and J. C. Bezdek, "Linguistic prototypes for data from eldercare residents," *IEEE Trans. Fuzzy Systems*, vol. 22, no. 1, pp. 110–123, 2014.

[16] E. Reiter, S. Sripada, J. Hunter, and I. Davy, "Choosing words in computer-generated weather forecasts," *Artificial Intelligence*, vol. 167, pp. 137–169, 2005.

[17] A. Ramos-Soto, A. Bugarín, and S. Barro, "On the role of linguistic descriptions of data in the building of natural language generation systems," *Fuzzy Sets and Systems*, vol. 285, no. C, pp. 31–51, 2016.

[18] V. Novak and I. Perfiljeva, "Time series mining by fuzzy natural logic and f-transform," vol. 2015, pp. 1493–1502, 03 2015.

[19] K. Kaczmarek-Majer and O. Hryniewicz, "Application of linguistic summarization methods in time series forecasting," *Information Sciences*, vol. 478, pp. 580 – 594, 2019.

[20] A. Alvarez-Alvarez and G. Triviño, "Linguistic description of the human gait quality," *Eng. Appl. of AI*, vol. 26, no. 1, pp. 13–23, 2013.

[21] G. Trivino and M. Sugeno, "Towards linguistic descriptions of phenomena," *International Journal of Approximate Reasoning*, vol. 54, no. 1, pp. 22 – 34, 2013.

[22] H. Banaee, M. U. Ahmed, and A. Loutfi, "A framework for automatic text generation of trends in physiological time series data," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 3876–3881.

[23] R. Castillo-Ortega, N. Marín, and D. Sánchez, "A fuzzy approach to the linguistic summarization of time series," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17, pp. 157–182, 2011.

[24] M. Peláez-Aguilera, M. Espinilla, M. Olmo, and J. Medina, "Fuzzy linguistic protoforms to summarize heart rate streams of patients with ischemic heart disease," *Complexity*, vol. 2019, pp. 1–11, 01 2019.

[25] A. Jain, M. Popescu, J. Keller, M. Rantz, and B. Markway, "Linguistic summarization of in-home sensor data," *Journal of Biomedical Informatics*, vol. 96, p. 103240, 2019.

[26] R. R. Yager, "A new approach to the summarization of data," *Information Sciences*, vol. 28, no. 1, pp. 69 – 86, 1982.

[27] V. Novak, "Mining information from time series in the form of sentences of natural language," *International Journal of Approximate Reasoning*, vol. 78, pp. 192 – 209, 2016.

[28] M. Zemankova and A. Kandel, *Fuzzy Relational Databases - A Key to Expert Systems*. Verlag TUV Rheinland, 1984.

[29] M. D. Ruiz, D. Sanchez, and M. Delgado, "On the relation between fuzzy and generalized quantifiers," *Fuzzy Sets and Systems*, vol. 294, pp. 125 – 135, 2016, theme: Rough Sets, Similarity and Quantifiers.

[30] V. Novák, "A formal theory of intermediate quantifiers," *Fuzzy Sets and Systems*, vol. 159, no. 10, pp. 1229 – 1246, 2008, mathematical and Logical Foundations of Soft Computing.

[31] A. Jain and J. Keller, "On the computation of semantically ordered truth values of linguistic protoform summaries," in *FUZZIEEE*, 2015.

[32] F. Höppner, "Time series abstraction methods – a survey," in *Informatik bewegt: Informatik 2002 - 32. Jahrestagung der Gesellschaft für Informatik e.v.* GI, 2002, pp. 777–786.

[33] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data mining in Time Series Databases*, 1993, pp. 1–22.

[34] T. chung Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164 – 181, 2011.

[35] U. Ramer, "An iterative procedure for the polygonal approximation of plane curves," *Computer Graphics and Image Processing*, vol. 1, no. 3, pp. 244 – 256, 1972.

[36] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112–122, Oct. 1973.

[37] F. Chung, T.-C. Fu, R. Luk, and V. Ng, "Flexible time series pattern matching based on perceptually important points," *International Joint Conference on Artificial Intelligence Workshop on Learning from Temporal and Spatial Data*, pp. 1–7, 01 2001.

[38] M. P. Mattson, "Superior pattern processing is the essence of the evolved human brain," *Frontiers in Neuroscience*, vol. 8, p. 265, 2014.

[39] L. A. Zadeh, "A fuzzy-set-theoretical interpretation of linguistic hedges," *Journal of Cybernetics*, vol. 2, no. 2, pp. 4–34, 1972.

[40] M.-J. Lesot, G. Moyse, and B. Bouchon-Meunier, "Interpretability of fuzzy linguistic summaries," *Fuzzy Sets and Systems*, vol. 292, pp. 307 – 317, 2016.

[41] D. Doran, S. Schulz, and T. R. Besold, "What does explainable AI really mean? A new conceptualization of perspectives," in *Proc. of the First International Workshop on Comprehensibility and Explanation in AI and ML*, vol. 2071, 2017.

Carmen Martinez-Cruz received the Ph.D. degree in computer science and artificial intelligence in 2008 from the University of Granada, Spain. She joined the University of Jaén in 2005, where she is currently an non-tenured Associate Professor with the Computer Science Department. Her research has focused on the area of knowledge representation, fuzzy databases, ontologies, fuzzy logic, linguistic summarization and eldercare technologies.

Antonio Rueda received the Ph.D. degree in computer and information systems in 2004 from the University of Malaga, Spain. He is an Associate Professor of Computing at the Escuela Politécnica Superior, University of Jaén (Spain). His main interests are in computer graphics, focusing on design of geometric algorithms, processing of 3D laser scanned data or GPU computing. He has presented his work in over 40 papers and communications in journals and conferences.

Mihail Popescu received his B.S. degree in NuclearA picture containing wall, person, indoor Description automatically generated Engineering from the Bucharest Polytechnic Institute in 1987. Subsequently, he received his M.S. degree in Medical Physics in 1995, his M.S. degree in Electrical and Computer Engineering in 1997, and his Ph.D. degree in Computer Engineering and Computer Science in 2003 from the University of Missouri. He is currently a Professor with the Department of Health Management and Informatics, an Adjunct Associate Professor of Nursing and an Adjunct Associate Professor of Electrical and Computer Engineering at the University of Missouri in Columbia, Missouri, USA. Dr. Popescu is interested in intelligent medical system design and medical data processing. His current research focus is developing decision support systems for early illness recognition in elderly and investigating sensor data summarization and visualization methodologies. He has authored or coauthored more than 200 technical publications. In 2011, Dr. Popescu received the IEEE Franklin V. Taylor Award of the IEEE Systems, Man, and Cybernetics Society. He is a senior IEEE member.

James M. Keller received the Ph.D. in Mathematics in 1978. He is now the Curators' Distinguished Professor Emeritus in the Electrical Engineering and Computer Science Department at the University of Missouri. Jim is an Honorary Professor at the University of Nottingham. His research interests center on computational intelligence: fuzzy set theory and fuzzy logic, neural networks, and evolutionary computation with a focus on problems in computer vision, pattern recognition, and information fusion including bioinformatics, spatial reasoning in robotics, geospatial intelligence, sensor and information analysis in technology for eldercare, and landmine detection. He has coauthored over 500 technical publications. Jim is a Life Fellow of IEEE, a Fellow of the International Fuzzy Systems Association (IFSA), and a past President of the North American Fuzzy Information Processing Society (NAFIPS). He received the 2007 Fuzzy Systems Pioneer Award, the 2010 Meritorious Service Award from the IEEE Computational Intelligence Society (CIS) and the IEEE Frank Rosenblatt Technical Field Award in 2021. He has been a distinguished lecturer for the IEEE CIS and the ACM. He also has been Editor-in-Chief of the IEEE Transactions on Fuzzy Systems, Vice President for Publications of the IEEE Computational Intelligence Society (CIS) and elected CIS Adcom member. He is President-Elect of CIS in 2021.