SCIENTIFIC REPORT

inProfilePhoto: a mobile app to assist people with visual disabilities in taking profile photos

Patricia Conde-Clemente, Jose M. Alonso, Gracian Trivino
European Centre for Soft Computing
{patricia.conde,jose.alonso,gracian.trivino}@softcomputing.es

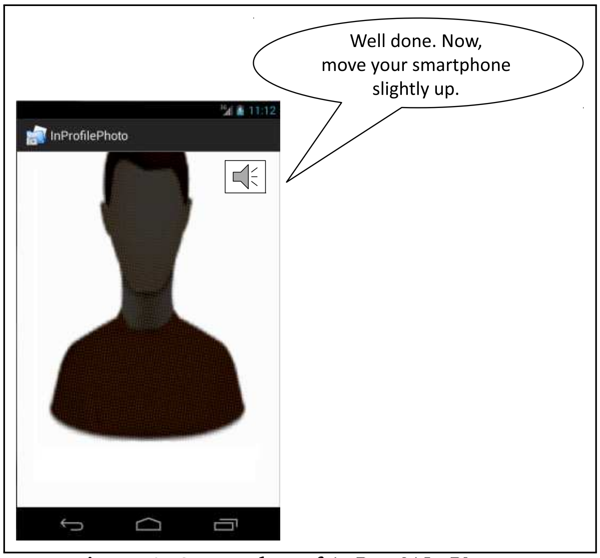

Figure 1. Screenshot of `inProfilePhoto`.

`inProfilePhoto` performs a seemingly simple task but really hard for people with visual disabilities, such as taking a profile photo correctly framed. It interacts in real-time with the user through linguistic commands related to the movements he/she has to perform in order to get framed (Fig. 1). When the user hears the instruction, he/she is expected to react with the movement advised. Then, the system analyzes the last movement made by the user and informs him/her about the effectiveness of the movement. Once the user is properly framed, the picture is taken automatically.

We have developed a new technology for building interpretable control systems including human-in-the-loop. It is supported by the Computational Theory of Perceptions (CTP) [4] which provides a framework to implement computational systems with the capacity of computing with the meaning of natural language expressions, i.e. with the capacity of computing with imprecise descriptions of the world in a similar way that humans do.

Our approach based on CTP, for developing computational systems able to generate linguistic descriptions of data, is called granular linguistic model of phenomena (GLMP)[1] [3]. The effectiveness of CTP relies on human-centric interpretability of the designed models. The human-centric character of interpretable fuzzy systems is highly appreciated in many applications, especially in those involving high interaction with humans. By combining GLMP with the Highly Interpretable Linguistic Knowledge methodology (HILK)[2] [1] we are able to yield a highly interpretable GLMP.

1 Theoretical Background

GLMP consists of a network of perception mappings (PMs). Each PM receives a set of computational perceptions (CPs) and transmits upwards a CP. In the network, each CP covers specific aspects of the phenomenon with certain degree of granularity. Using different aggregation functions and different linguistic expressions, the GLMP paradigm allows the designer to model computationally his/her perceptions.

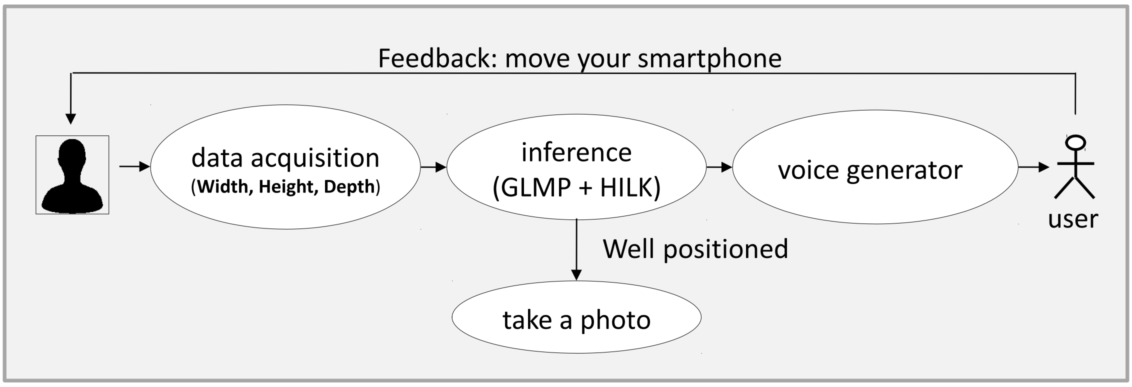

Figure 2. Flowchart.

[1] http://www.softcomputing.es/ldcp
[2] http://www.softcomputing.es/ifs

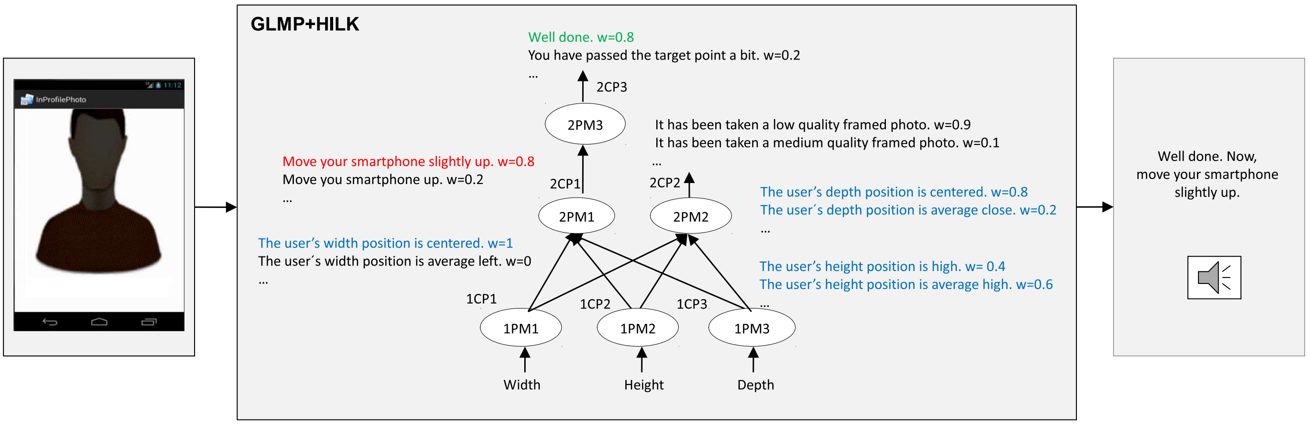

Figure 3. Illustrative example.

1.1 Computational perception (CP)

A CP is the computational model of a unit of information acquired by the designer about the phenomenon to be modeled. In general, CPs correspond with specific parts of the phenomenon at certain degrees of granularity. A CP is a couple (A,W) where:

A = $(a_1, a_2, .., a_n)$ is a vector of linguistic expressions (words or sentences in natural language) that represents the whole linguistic domain in CP.

W = $(w_1, w_2, .., w_n)$ is a vector of validity degrees $w_i \in [0, 1]$ assigned to each a_i in the specific context.

1.2 Perception mapping (PM)

We use PMs to create and aggregate CPs. A PM is a tuple (**U**,**y**,**g**,**T**) where:

U = $(u_1, u_2, .., u_n)$ is a vector of n input CPs $u_i = (A_{u_i}, W_{u_i})$. In the special case of first order perception mappings (1PMs), these are the inputs to the GLMP and they are values $z \in \mathbb{R}$ being provided either by a sensor or obtained from a database.

y = (A_y, W_y) is the output CP.

g is an aggregation function employed to calculate $W_y = g(W_{u_1}, W_{u_2}, ..., W_{u_n})$ from the input CPs. In Fuzzy Logic, many different types of aggregation functions have been developed. For example, **g** might be implemented using a set of fuzzy rules. In the case of 1PMs, **g** is built using a set of membership functions.

T is a text generation algorithm that allows generating the sentences in A_y. In simple cases, **T** is a linguistic template, e.g., "It has been taken a {low, medium, high} quality framed photo", but it can be customized according to user preferences, mood, etc.

1.3 Interpretability-guided design of GLMP

HILK is a fuzzy modeling methodology that was conceived for carefully integrating expert and induced knowledge under the fuzzy logic formalism. It enables the user to follow a step-by-step procedure in the generation of all elements involved in a fuzzy knowledge base, starting from the design of fuzzy partitions, going through the rule-based learning and ending up with a knowledge base improvement stage which iteratively refines both partitions and rules. For ensuring interpretability of the GLMP we have carried out the following steps:

- Define CPs as linguistic variables with small odd number of linguistic terms. Each linguistic variable is characterized by a Strong Fuzzy Partition (SFP) in its universe of discourse, as recommended by HILK. As a result, global semantics is defined. Moreover, SFPs satisfy most constraints (coverage, distinguishability, overlapping, etc.) demanded to have interpretable partitions. Increasing the granularity of the underlying fuzzy partitions produces an increase in the number of linguistic expressions given to the user. This point is very important, because the GLMP should contain only the strictly necessary and sufficient information to describe the phenomenon.
- Define CP linguistic rules of form "IF *premise* THEN *conclusion*". Both *premise* and *conclusion* are made up of linguistic propositions like "V is a_i" where one of the previously defined linguistic terms is assigned to one of the selected variables. The absence of one variable in a rule means such variable is not considered in the evaluation of the selected rule.

2 Practical Applications

Fig. 2 sketches the architecture of the developed app. It runs on a smartphone equipped with front camera and Android software for detecting person's faces (data acquisition) and for converting the generated text messages into voice messages that can be delivered to the user through the speakers of the smartphone (voice command generator). `inProfilePhoto` is freely available at:

`http://www.softcomputing.es/inprofilephoto`

The steps for taking a profile photo correctly framed are the following:

1. The app receives a profile photo in which a face is detected. The values of Width, Height and Depth

which determine the user's position are automatically extracted.

2. Then, it infers the most suitable motion the user should perform in order to get framed according to the developed GLMP+HILK model.

3. Then, a linguistic command is communicated to the user through voice commands. The user is expected to move the smartphone accordingly.

4. Finally, the photo is taken (once the user is properly positioned) and another linguistic expression is conveyed to the user with the aim of informing him/her that the photo was taken successfully.

Fig. 3 shows an illustrative example with a set of linguistic expressions (along with their validity degree) that are generated by GLMP+HILK model for a given input profile photo. The meaning of generated sentences is emphasized in four colors:

- **Blue**: a set of sentences summarizing the current state of the user.
- **Red**: the movement required to the user in order to be properly framed.
- **Green**: feedback given to the user according to his/her reaction (information about the level of fulfillment of the previous command).
- **Black**: sentences not to be communicated to the user.

The interested reader is kindly referred to [2] for further details about the implemented GLMP+HILK .

Acknowledgments

This work was supported in part by the Spanish Ministry for Education (FPI-MICINN BES-2012-057427), the Asturian Government (Grant COF13-052), the Spanish Ministry of Science and Innovation (Grant TIN2011-29827-C02-01), and the Spanish Ministry of Economy and Competitiveness under Grants TIN2011-29824-C02-01 and TIN2011-29824-C02-02 (ABSYNTHE project).

References

[1] J. M. Alonso and L. Magdalena. HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. Soft Computing - A Fusion of Foundations, Methodologies and Applications, 15(10):1959-1980, 2011.

[2] P. Conde-Clemente, J. M. Alonso, and G. Trivino. Interpretable fuzzy system allowing to be framed in a profile photo trough linguistic expressions. In International Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), pages 463-468. Atlantis Press, 2013.

[3] L. Eciolaza, M. Pereira-Fariña, and G. Trivino. Automatic linguistic reporting in driving simulation environments. Applied Soft Computing, 2012.

[4] L. A. Zadeh. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. Journal of statistical planning and inference, 105(1):233-264, 2002.