This document is published in:

2012 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Lugano, 19-23 March 2012. IEEE pp. 865 - 870.
DOI: 10.1109/PerComW.2012.6197633

Ontological representation of time-of-flight camera data to support vision-based AmI

M.A. Serrano, J. Gómez-Romero, M.A. Patricio, J. García, J.M. Molina
Applied Artificial Intelligence Group (GIAA)
Universidad Carlos III de Madrid, UC3M
Colmenarejo, Madrid, Spain
Email:{miguel.serrano, juan.gomez.romero, miguelangel.patricio, jesus.garcia, josemanuel.molina}@uc3m.es

***Abstract*—Recent advances in technologies for capturing video data have opened a vast amount of new application areas. Among them, the incorporation of Time-of-Flight (ToF) cameras on Ambient Intelligence (AmI) environments. Although the performance of tracking algorithms have quickly improved, symbolic models used to represent the resulting knowledge have not yet been adapted for smart environments. This paper presents an extension of a previous system in the area of video-based AmI to incorporate ToF information to enhance scene interpretation. The framework is founded on an ontology-based model of the scene, which is extended to incorporate ToF data. The advantages and new features of the model are demonstrated in a Social Signal Processing (SSP) application.**

***Keywords*-Visual sensor networks; Time-of-Flight camera; Ontologies; Ambient Intelligence; Social Signal Processing**

I. INTRODUCTION

AmI aims at the development of computational systems that apply Artificial Intelligence techniques to process information acquired from sensors embedded in the ambience in order to provide helpful services to users in daily activities. AmI objectives are: (i) to *recognize* the presence of individuals in the sensed scene; (ii) to *understand* their actions and estimate their intentions; (iii) to *act* in consequence.

The use of visual sensors in AmI applications has been poorly studied, even though they can obtain a large amount of interesting data. Some reasons have been usually argued to explain this absence: the economic cost of visual sensor networks, the computational requirements of visual data processing, the difficulties to adapt to changing scenarios, the disadvantages respect to other sensor technologies, and so forth.

In the last decade, a new visual sensor technology has emerged: ToF cameras. ToF cameras provide both intensity and distance information for each pixel of the image, thus offering 3-dimensional imaging. Recently, the cost of ToF sensors has dramatically reduced, which has lead to a widespread adoption of this technology, now even present in consumer electronics like the Kinect™ peripheral for Microsoft XBox™ system.

New computer vision algorithms have been proposed to detect and track human movements from ToF data. To name some application areas, ToF-based systems have been used in SSP to classify human postures [21], and in Ambient Assisted Living to detect people falls [15].

Unfortunately, current approaches do not offer a well-defined model to capture the semantics of ToF data. In this paper, we argue that the use of a formal conceptual model to represent ToF data offers several advantages at a low cost. Among other features, formal models allow us to establish a common symbolic vocabulary to describe and communicate camera data while providing support for logic-based reasoning. Symbolic language is closer to human language, and therefore it is easy to interact and interpret system inputs and outputs. Reasoning, in turn, can be applied to check the consistency of the models and to infer additional knowledge from explicit information.

This paper describes an ontology-based representation model for data acquired from ToF technologies. This model is incorporated into a framework for contextual fusion of 2-D visual information previously proposed by our research group [5]. The ontologies of the initial framework have been extended to include ToF data, specifically:

- An additional Euclidean dimension for the position of ToF objects. This is easily achieved by relying on the qualia approach used in the original ontology model to represent properties and property values.
- A new definition of the concepts that represent human entities in the scene. Essentially, Person concept is now associated to a description of anatomical joints and limbs. This description has been formalized according to existing patterns to represent part-whole relations with ontologies and current ToF-based computer vision models for articulated bodies.

A case study based on a SSP environment is presented to illustrate the functioning of the extended framework. The goal is the formal representation of complex activity recognition data through ontologies. The example explains a novel application of ToF cameras for live market researches. Finally a straightforward rule is presented to describe the ability of the model to express the semantics of real situations.

The paper is organized as follows. Section 2 presents the state of the art of ToF camera applications and the Kinect™ sensor technology. Section 3 includes an overall description of the new features added to the existing ontology-based computer vision approach. An ontology-

based human skeleton representation is explained in Section 4. Section 5 depicts a case study to detect interesting situations in a SSP scenario. Section 6 summarizes the conclusions obtained and proposes some directions for future work.

II. Time-of-flight camera applications and Kinect™ sensor

ToF cameras provide support for new application thanks to their genuine features, compact structure, low weight, reduced power consumption, low price, high resolution and real-time intensity and distance acquisition. These sensors implements a non-invasive technology to obtain 3-D data, which avoid the use of embedded hardware like inertial devices. ToF cameras have been mainly applied to human activity recognition and Human Machine Interaction (HMI). These contributions can be adapted to AmI environments.

Several works have been aimed at improving person and people tracking by relying on ToF sensors. Kahlmann et al. [12] presented a tracking algorithm for the detection of moving people. The approach is based on a Recursive Bayesian Filter, more specifically the authors applied on a flexible and general solution named Condensation algorithm [11]. A fast multi-person tracking approach in 3-D environments is Shape from Silhouettes (SfS) [6]people using few cameras. Another proposal on real-time multi-person tracking algorithm is presented in [1]. This work delves into partial occlusions and close interactions between several people under severe low-lighting conditions.

Eye-safeness of ToF cameras facilitates the development of applications for human face detection. A nose detection algorithm is presented in [8]. The combined use of range and amplitude data achieves a robust identification in a wide range of head orientations. Another example is [9], which presents a boosting method based on the use of both gray scale and depth images.

ToF sensors can be also applied for hand tracking. Breuer et al. [3] focused on the problem of reconstruction –i.e. inferring the various degrees-of-freedom of the hand from sensor data. A real-time dynamic gesture recognition can be found in [20]. The research is focused on 3-D medical data exploration.

ToF cameras are useful to detect and track human articulations for full-body reconstruction. Knoop et al. [14] presented a framework to fuse information acquired from different sensors –stereo, ToF and monocular–. This research is based on a 3-D body model composed of cylinders and different kinds of joint to track complex movements. The approach by Holte et al. [10] combines both intensity and depth data for body gesture recognition. The proposal is trained from a specific point of view, and is able to recognize gestures from different points of view by using a spherical harmonic context representation.

All these researches were developed with sensor technologies prior to Kinect™. Kinect™ have meant a breakthrough in the hardware features of these devices. The sensor is based on a structured near-infrared light, and a standard CMOS image sensor used to receive the reflected light. A multi-sense system-on-chip provides synchronized real-time depth image, color image and audio stream. Kinect™ operates at a maximum frame rate of 30 fps. The capture range is between 0.8 and 3.5 meters with a maximum resolution of 640x480.

Kinect™ is supported by three freely available libraries. OpenKinect[1], CL NUI[2] and OpenNI[3]. OpenKinect is an open source project dual-licensed under Apache 2.0 and GPL2. The library provides drivers for the sensor and a cross-platform API that works on Windows, Linux, and OS X; wrappers for different languages such as Python, C++ and C#; and an analysis library which is expected to include among other things, hand tracking, skeleton tracking and 3-D reconstruction. The Windows Kinect Driver/SDK - CL NUI plataform provides a SDK with freely available C, C++ and C# libraries, an API and a stable driver for Xbox NUI Audio, NUI Camera and NUI Motor and Accelerometer devices on Windows machines. OpenNI is a not-for-profit organization composed of several companies. The purpose of the organization is to promote the compatibility and interoperability of Natural Interaction (NI) middlewares, applications and devices, like Kinect™ . OpenNI developed a framework that can be used across different platforms and devices. The framework is based on a set of middleware libraries that convert raw data from a compliant device to application data.

III. Ontology-based computer vision model and ToF technology integration

The framework for computer vision representation presented in [5] is based on an ontological model for the representation of context and scene entities. This model is organized into several levels compliant with the Joint Directors of Laboratories (JDL) model [16]. Each layer includes general concepts and properties to describe general computer vision entities and relations at different abstraction level. Concepts that belong to a less abstract ontology are the building blocks of concepts corresponding to a more abstract ontology. Current implemented levels are:

- Tracking Entities (TREN) level, to model input data coming from the tracking algorithms.
- Scene Objects (SCOB) level, to model real-world entities, properties, and relations.
- Activities (ACTV) level, to model behavior descriptions.

[1]OPENKINECT OpenKinect Main Page. http://openkinect.org/

[2]LABORATORIES, C. About: CL NUI Platform. Code Laboratories, http://codelaboratories.com/kb/nui

[3]OPENNI OpenNI. http://openni.org/

The model has been designed to promote extensibility and modularity. Ontologies may contain both perceptual and context data. Perceptual data is automatically extracted by the tracking algorithm, while the context data is external knowledge used to complete the comprehension of the scene. For example, the description of a sensorised static object –size, position, kinematic features, type of object, and so on– is regarded as context data.

Some changes are needed to model tracking data coming from ToF devices. The priority to adapt these changes is to maintain the compatibility with the previous approach.

A. Three dimensional representation

The introduction of new devices requires upgrading the capacity of spatial representation in the model from two to three dimensions. These changes concern both perceptual data captured by ToF cameras and context data representing physical objects. The previous model followed the qualia approach used in the upper ontology DOLCE [4]. This modeling pattern distinguishes between properties themselves and the space in which they take values. The values of a quality –e.g. Position– are defined within a certain conceptual space –e.g. 2DPoint. To adapt the ontology-based model to this new quality space, the 3DPoint concept, which represents a position using three coordinates, is included as a subclass of PositionValueSpace, which represents the space of values of the physical positions.

B. Real-world entities

Current Kinect™ algorithms are able to detect real-world entities; e.g. a person including data related to the human limbs and joints. Our ontology-based model represents these kind of real-world data at the SCOB level. However, SCOB assertions must be supported by TREN data. TREN is adapted to represent low level data of human members and joints –position, size, kinematic state, and so on– this information is associated to the Track concept.

The inclusion of limbs and joints is compliant to the previous version of the TREN ontology. The applied part-whole pattern (see below) allows keeping backwards compatibility. In fact, this model can combine 2-D monocular cameras and ToF devices using the same set of ontologies.

IV. General model for ontology-based human skeleton representation

There are a lot of existing ontologies designed to share and reason with structured data representing human anatomy [19]. Unfortunately, these ontologies have been developed in biomedical environments and define a complex conceptualization which is not useful to our needs. There are also other ontologies that represent the human body in a more simplified way [7]; however these ontologies are not designed to deal with different value spaces in a cognitive environment. A general pattern based on part-whole relationships is proposed to cover the semantic representation of data captured using ToF sensors. The designed ontology adapts the patterns presented in [18] and follows the conceptualization of articulated bodies shown in [13] while keeping compatibility with DOLCE. Our proposal can be broadly adaptied to other fields. Some examples could be formal representation of the cognitive vision discipline [17] and automatic code generation for virtual worlds [2].

Real-world knowledge is organized by using mereological –part-whole– relationships. A clear example is how the human mind divides the structure of a body in subjetive parts. Kinect™ skeletal view (see Fig.1[4]) is able to describe a detected person in terms of two kinds of attributes: (i) body members –hands, feet, thigh, and so on; (ii) joints –shoulders, elbows, wrists, knees, and so forth. TREN represents the attributes detected and the limbs composed by these attributes as a conceptualization. Resulting concepts represent the parts of the human body which is embodied in the Track concept.

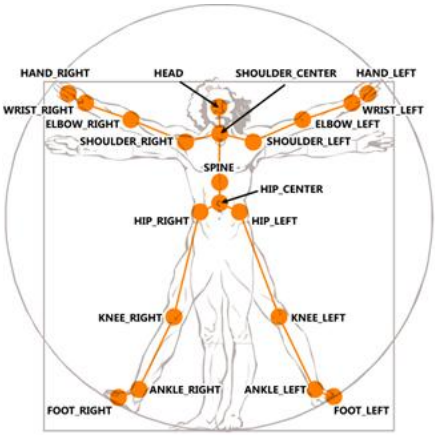

Figure 1: Joints captured by Kinect™ skeletal view

Two properties are used to represent part-whole relationships: (i) partOf; (ii) partOf_directly –a partOf subproperty. partOf is a transitive property whose goal is establishing the correspondences between the parts and all the entities containing them. partOf_directly defines the subjetive relation among a part and the next direct level of composed entities. These properties are necessary since cardinality restrictions over transitive properties, such as partOf, are not allowed by OWL-DL. Therefore, partOf_directly is used to define restrictions to mantain the cardinality consistency, partOf is used to infer both direct and indirect parts by means of transitive characteristic and the instances of partOf_directly property.

The ontology is extended with classes to represent direct parts –e.g. TrackPartDirectly– and the overall set of part-

[4]Fig. 1 source: http://embodied.waag.org

whole relationships –e.g. TrackPart. TrackPartDirectly subsumes direct parts of a Track such as Head, UpperLimb, LowerLimb, and so forth. TrackPart subsumes the set of parts of the Track concept. For example, the direct parts of an UpperLimb concept, namely Arm, Forearm, Hand, Shoulder, Elbow and Wrist, are classified as subclasses of TrackPart; however they are not considered subclasses of TrackPartDirectly.

The classes hosting parts state existential range restrictions –`owl:someValuesFrom`– over part properties. To improve the consistency, cardinality restrictions –exactly 1– are stated over partOf_directly as necessary conditions into the concepts corresponding to body members and joints. This means "a part only belongs directly to the next level entity and just to that entity".

The combined use of the part properties and the restricted classes leads reasoners to automatically infer new taxonomies derived based on part-whole relationships. Fig. 2 illustrate an example of a taxonomy inferred from a explicitly stated taxonomy. Unfortunately, adding qualitative cardinality restrictions on each concept could significantly affect the performance of the reasoner. Some other configurations for this pattern are possible and also valid. This implementation tries to reduce the classification time complying the semantics of the human body domain.

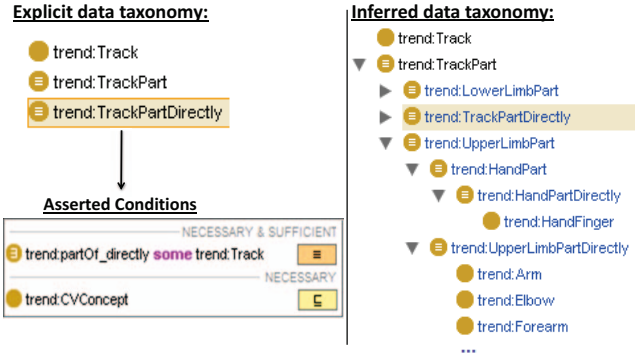

Figure 2: An example of explicit and inferred taxonomies

The classification of joints is inspired by the virtual model shown in [13]. Similar to this article, the model comes from the application of computer vision techniques in ToF devices. There are three types of joints (see Fig. 3) depending on the degrees of freedom (DoF): (i) UniversalJoint, three DoF; (ii) HingeJoint, one DoF and two restricted DoF; (iii) EllipticJoint, three restricted DoF. Joint concepts store important data such as the articulated body members and the angle between them. This data is basic to mantain the consistency and to improve the semantic capacity of the model.

The model is designed by taking into account future changes in the granularity of the obtained data. New devices able to offer an accurate definition of the body members –e.g the fingers of a hand– are easily adaptable. The larger the number of levels in the model, the greater amount of data is inferred. More details and additional information about the data described in this section can be found in the authors' web page[5].

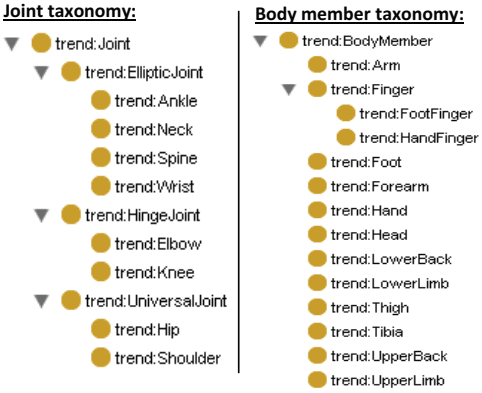

Figure 3: Explicit taxonomies for joints and body members

V. Case Study: Live Market Research

Learning about relationships between the customer and the product at the point of sale is a very interesting knowledge in many economic fields, such as sales or marketing. Body gestures and spatial relationships contain useful knowledge about the sensations and intentions of shopping experiences. The model hereby presented can be used to automatically build live market research works based on the reactions and interactions of customers with the products. Next subsections describes our system gesture instantiation procedure and a description of the expressiveness of the ontology model by presenting an activity recognition example.

A. Gesture instantiation procedure

A data set containing the skeleton representation of several –11– people was designed to test the new representation. These body structures were captured by using a Kinect™ sensor. For each person five types of upper limbs gestures were stored: down, open, up, diagonal and akimbo. A control system based on the OWL API[6] functionalities automates the assertion of data in the form of axioms from the capture device to the ontology formalism. The control system manages the classification of the individuals received from the Kinect™ sensor, the explicit property instantiations such as partOf_directly and the instantiation of properties that represent the articulation of body member through a joint. The control system also manages the automatic calculation of datatypes from the received data, such as the size of the body members, angles formed between them and so forth.

[5]Additional resources: http://www.giaa.inf.uc3m.es/miembros/jgomez/et/
[6]OWL API: http://owlapi.sourceforge.net/

An example instatiation of data to describe a left upper limb with down gesture for the person in Fig 4., would include: (i) classification of joint instances (see Fig. 3); (ii) partOf_directly property instantiations (see Fig. 2); (iii) joint positioning data.

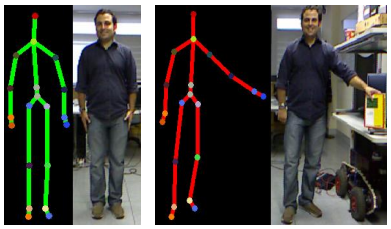

Figure 4: Gesture instantiation and action example

B. Activity recognition example: Picking up an object

Activity recognition usually requires composition of simple activities along the time. Therefore temporal analysis is required in order to recognize complex activities [10]. Our ontology model is expressive enough to represent the temporal dimension of the activities. The representation capabilites resulting from the combined use of Kinect™ and the ontology-based model offer simple but very expressive tools to detect interesting activities for a market research confection.

Interesting activities for current market researches may be: stand in front of, look at, point at and touch a product, and compare two products. Comparing products normally implies the recognition of simple interactions between different body members and several static objects that are part of the context. Recognition of simple interactions generally starts with a body member picking up an object; these facts can be detected, for example, if there is a spatial relationship between a hand and an object. This process is more robust if the object includes sensors able to detect state features –kinematic state, position, and so on.

In order to demonstrate the expressiveness of our representation, a syntactically relaxed nRQL –the query language of the RACER reasoner– rule is presented. This rule finds picking up activities between persons and smart objects.

First, different variables that act along the rule are declared (3-9). Then, a correspondence between Tracks and Persons is performed (10). Hands and Elbows pertaining to the Track are retrieved (11-12). The rule checks if these individuals are parts of the same UpperLimb (13-15). The act of picking up an object usually means that the Elbow is maintained at over 90 degrees (16). Afterwards the spatial relationships between Hands and Products are retrieved (17). Finally, to increase the accuracy the rule considers if the object involved in the situation is currently moving (18). If it does not exist any active pick up relationship that acts along the same Hand and Product and the antecedent conditions are satisfied, then the consequent is applied. The consequent creates a PickUp activity (20) with a known begining (21) and an unknown ending (22) as well as a relationships among the new activity with the passive (23) and the active subject (24).

```
1  (firerule
2     (and //Antecedent
3         (?currentFrame #!tren:CurrentFrame)
4         (?person #!scob:Person)
5         (?track #!tren:Track)
6         (?hand #!tren:Hand)
7         (?elbow #!tren:Elbow)
8         (?product #!scob:StaticObject)
9         (?product "type" #!scob:Type)
10        (?person ?track #!scob:hasAssociatedTrack)
11        (?hand ?track #!tren:partOf)
12        (?elbow ?track #!tren:partOf)
13        (?hand ?upperLimb1 #!tren:partOf_directly)
14        (?elbow ?upperLimb2 #!tren:partOf_directly)
15        (equal ?upperLimb1 ?upperLimb2)
16        (?elbow (>= #!tren:angle 90)
17        (not (?*hand ?*product :dc)
18        (?product (> #!tren:velocity 0)
19    (//Consequent
20        (instance (new-ind ?pickUpAct) #!actv:PickUp)
21        (related (?pickUpAct ?currentFrame #!tren:isValidInBegin)
22        (related (?pickUpAct "unknown_frame" #!tren:isValidInEnd))
23        (related (?pickUpAct ?object #!actv:pickedUp)
24        (related (?pickUpAct ?person #!actv:pickingUp)
25 )
```

Figure 5: Rule to exemplify expressiveness

Improved functionalities for activity recognition algorithms can be offered by relying on the semantic expressiveness of the model. It is possible to use techniques to refine the search space; for example, by considering the type of object analyzed. If we are only interested in knowing current interactions of the customers with a specific kind of objects, it is only necessary: (i) to find the tracks with a proper part spatial relationship with the area type corresponding to the object of interest; (ii) to browse for a spatial relationship between the product and the hands of the previously retrieved tracks.

Other rules can be defined to extract interesting market research data. For instance counting the kind of people pointing at a product. The recognition of this activity only involves the analysis of the state of the joints of an UpperLimb to infer the pointed object. Hence, using the size of the limbs allow us to infer addtional data, such as the range of ages –e.g. child or adult– of the people who is attracted by the product.

VI. Conclusion and Future Work

This article has presented a general ontology-based model for formal representation of the human body. This model can be exported to other fields such as cognitive vision or code generation from ontologies. The model has been embedded into a previous computer vision framework by relying on part-whole patterns and DOLCE recommendations. The proposal accomplishes an extension, which includes Kinect™ skeletal view data representation with backward compatibility. To illustrate the functioning of the extended framework, a case study for live market research with a simple activity recognition example rule has been described.

Future works will address the application of the entire model to a real life scenario combining monocular and ToF sensors. This application should include a probabilistic mechanism to reason with real world data asserted in the model, which may be imprecise or uncertain.

ACKNOWLEDGMENT

This work was supported in part by Projects CICYT TIN2011-28620-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485) and DPS2008-07029-C02-02.

REFERENCES

[1] A. Bevilacqua, L. D. Stefano, and P. Azzari. People tracking using a Time-of-Flight depth sensor. In *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06)*, pages 89–89, Sydney, Australia, Nov. 2006.

[2] W. Bille, O. D. Troyer, F. Kleinermann, B. Pellens, and R. Romero. Using ontologies to build virtual worlds for the web. In *Proceedings of the 3rd IADIS International Conference WWW/Internet (ICWI'04)*, pages 683–690, Madrid, Spain, Oct. 2004.

[3] P. Breuer, C. Eckes, and S. Mller. Hand gesture recognition with a novel IR Time-of-Flight range camera–a pilot study. In *Proceedings of the 3rd International Conference on Computer Vision/Computer Graphics Collaboration Techniques (Mirage'07)*, volume 4418, pages 247–260, Rocquencourt, France, Mar. 2007.

[4] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management. 13th International Conference on Ontologies and the Semantic Web (EKAW'02)*, pages 223–233, Sigenza, Spain, Oct. 2002.

[5] J. Gómez-Romero, M. A. Serrano, M. A. Patricio, J. García, and J. M. Molina. Context-based scene recognition from visual data in smart homes: An Information Fusion approach. *Personal and Ubiquitous Computing*, pages 1–23, Sept. 2011.

[6] S. A. Guomundsson, R. Larsen, H. Aanaes, M. Pardas, and J. R. Casas. ToF imaging in smart room environments towards improved people tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, pages 1–6, Anchorage, USA, June 2008.

[7] M. Gutiérrez, A. Garcia-Rojas, D. Thalmann, F. Vexo, L. Moccozet, N. Magnenat-Thalmann, M. Mortara, and M. Spagnuolo. An ontology of virtual humans. *The Visual Computer*, 23(3):207–218, July 2007.

[8] M. Haker, M. Bohme, T. Martinetz, and E. Barth. Geometric invariants for facial feature tracking with 3D ToF cameras. In *International Symposium on Signals, Circuits and Systems (ISSCS'07)*, volume 1, pages 1–4, Iasi, Romania, July 2007.

[9] D. W. Hansen, R. Larsen, and F. Lauze. Improving face detection with ToF cameras. In *Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS'07)*, volume 1, pages 1–4, Iasi, Romania, July 2007.

[10] M. B. Holte, T. B. Moeslund, and P. Fihl. Fusion of range and intensity information for view invariant gesture recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, pages 1–7, Anchorage, USA, June 2008.

[11] M. Isard and A. Blake. Condensation-conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[12] T. Kahlmann, F. Remondino, and S. Guillaume. Range imaging technology: New developments and applications for people identification and tracking. In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE'07)*, volume 6491, 2007.

[13] S. Knoop, S. Vacek, and R. Dillmann. Modeling joint constraints for an articulated 3D human body model with artificial correspondences in ICP. In *5th IEEE-RAS International Conference on Humanoid Robots*, pages 74–79, Tsukuba, Japan, Dec. 2005.

[14] S. Knoop, S. Vacek, and R. Dillmann. Sensor fusion for 3D human body tracking with an articulated 3D body model. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'06)*, pages 1686–1691, Orlando, USA, May 2006.

[15] A. Leone, G. Diraco, and P. Siciliano. Detecting falls with 3D range camera in ambient assisted living applications: A preliminary study. *Medical Engineering & Physics*, 33(6):770–781, July 2011.

[16] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, and F. White. Revisiting the JDL data fusion model II. In *Seventh International Conference on Information Fusion*, pages 1218–1230, Stockholm, Sweden, July 2004.

[17] A. Pinz, H. Bischof, W. Kropatsch, G. Schweighofer, Y. Haxhimusa, A. Opelt, and A. Ion. Representations for cognitive vision: A review of appearance-based, spatio-temporal, and graph-based approaches. *Electronic Letters on Computer Vision and Image Analysis (ELCVIA)*, 7(2):35–61, 2009.

[18] A. Rector, C. Welty, N. Noy, and E. Wallace. Simple part-whole relations in OWL ontologies. Internet draft, 2005.

[19] C. Rosse and J. L. V. M. Jr. A reference ontology for biomedical informatics: The foundational model of anatomy. *Journal of Biomedical Informatics*, 36(6):478–500, Dec. 2003.

[20] S. Soutschek, J. Penne, J. Hornegger, and J. Kornhuber. 3-D gesture-based scene navigation in medical imaging applications using Time-of-Flight cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, volume 1, pages 1–6, Anchorage, USA, June 2008.

[21] F. Wientapper, K. Ahrens, H. Wuest, and U. Bockholt. Linear-projection-based classification of human postures in Time-of-Flight data. In *IEEE International Conference on Systems, Man and Cybernetics (SMC'09)*, pages 559–564, San Antonio, USA, Oct. 2009.