CATEGORIZATION OF RESEARCH SURVEYS AND REVIEWS ON HUMAN ACTIVITIES

A. Angeleas, N. Bourbakis
CART
Wright State University
Dayton, OH, USA
angeleas.2@wright.edu, nikolaos.bourbakis@wright.edu

G. Tsihrintzis
University of Piraeus
Computer Science Department
Greece
geoatsi@unipi.gr

***Abstract* — Due the recently increased number of survey papers, we consider important to offer an evaluation of the different types of current survey papers to assess their benefits and inform the readers regarding their categories. The field selected here is the Human Activity Recognition (HAR), although its idea can be applied to other research areas as well. Thus, a variety of survey papers are studied thoroughly and compared for this categorization. From each paper a short synopsis is presented in order the reader to understand the concepts behind it. In order to objectively assess their importance, we formulate a maturity metric by using a set of significant features, that best describe the significance and contribution of the HAR research. Finally the anticipated result of this paper, is to serve as a reference point for the "dos' and don'ts" regarding the survey papers, and lay a foundation for further research, thoughts, discussion, and improvements of them, towards maturity and usefulness.**

Keywords—human activity recognition; human motion recognition; human pose estimation; human motion analysis; survey papers; design methodology; applications; implementation; comparative study

I. INTRODUCTION

With this study we want to provide a categorization of survey papers and an overview of their significance for the reader. We won't provide a critical overview, but because of the innumerable methodology specific studies, surveys are a worthy tool for a researcher to take a broader look to the area of Human Activity Recognition. Primarily we are going to propose our categorization for the selected Survey Papers. Finally with the use of a Maturity Metric, we'll evaluate the papers significance for each of our user categories, and provide statistical data of overlaps between them.

A. Categories

In this section, we present our categorization for the Survey Papers. We have focused on studies, published during the past 5 years, since a large number of research efforts and products that deal with HAR, were published and implemented. Primary research articles are innumerable, so we decided to take a more holistic view, via the survey reports. We attempt to categorize them based on the approach they take in order to present their findings. We propose to use two main categories named Comparison (C) and Review (R). These two categories are furtherly divided to two subcategories, Deep Evaluation (DE) and First Stage Evaluation (FSE), and Review to Analysis (RA) and Brief Reviews (BR). The reasoning behind the aforementioned categorization was to differentiate and understand the value of each survey paper, in order to be able to distinguish them accordingly [1, 2 and 3].

In both Comparison and Review categories we mainly found articles from professional journals, which meet the criteria of the research output of the writer. The main difference between these two general categories is the depth that researchers study each methodology, approach or system. The main characteristics of the subcategories are:

Deep Evaluation: they describe methodologies and products, but also test them on the same datasets, this category helps to understand the robustness, the ease of implementation, as well as the performance of each researched methodology or product. When we talk about 'scientific research methods', this is what most people immediately think of, because it passes all of the definitions of 'true science'. The researcher accepts or refutes the null hypothesis. The results generated are analyzable and used to test hypotheses, with statistics giving a precise and unambiguous picture.

First Level Evaluation: we shall see differences pointed out between methodologies, products, and a more critical view of each approach. How each approach works over time, its level of maturity and insights into input and output, interesting observations can be provided and different perspectives can be understood easier.

Descriptive Reviews: most researches have selected articles and described them in a way that the reader, will be able to understand the "know how" of each methodology and the needs of it in order to be implemented. This is a group of different research methods where researchers try to observe a phenomenon without interfering too much.

Brief Reviews: here we'll find plain reviews of methodologies and products. The beneficial point of this category is that the readers can learn about different

approaches and focus to the ones that are important to them. Often used as a pre-cursor to more rigorous studies, and avoid the problem of the experiment environment, which affects the behavior of an approach.

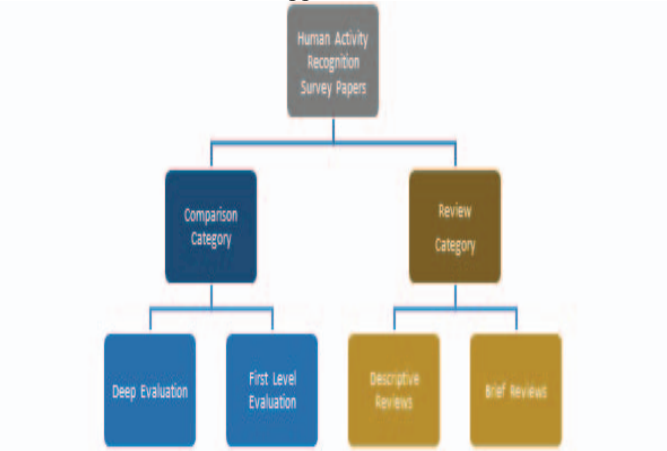

Fig. 1. Visualization of the proposed categorization

II. SURVEY CATEGORIES

In the previous section, the general information about our categorization schema was presented. Brief descriptions of each selected studies are provided in the following sections and show the classification and the main point of research that are leaning on. Our proposed classification method is generic enough, in order to divide the categories into an easy to view and understand way, while each category is meaningful and retains specific characteristics.

A. Deep Evaluation

Here we shall find researches, that examine each methodology separately and in order to evaluate each one of them, they are using datasets. The papers included in this category are [4-7].

In [4] the authors provide a comprehensive introduction to the field of the activity recognition with the use of body-worn sensors, and they compare different methodologies for specific hand gesture activities with the use of datasets.

Another study was conducted in methodologies that use Fuzzy Logic in [5]. Approaches that depend on the use of Microsoft Kinect and machine learning techniques were studied, the taxonomy of the study is organized according to the pipeline of the human motion analysis, from high-level to low-level.

In [6], the authors dwell with methodologies that use spatio-temporal interest point (STIP) detectors. They study selected methods in public datasets, in order to compare the performance on the basis of STIP detection ratio, and on the basis of action recognition ratio.

In the last study of this category [7] the authors studied recent developments of multi-view approaches for 3D human pose estimation and activity recognition. They categorized the approaches on 2D and 3D, finally they've done a qualitative and quantitative comparison of selected reviewed methods for activity recognition. Main concern is recognition accuracy, the number of actions that are used, and whether the results are based on all views or cross-view recognition.

B. First Level Evaluation

As we noted earlier this category includes studies that describe methodologies to some depth to state important details and offer a light comparative evaluation. However some papers adopt a more critical view towards them. The papers included in this category are [8 – 14].

In [8] the authors dwell with the problem of human detection in order to automatically locate people. The approach that it is adopted in organized in the thread of human object descriptors, the process is used in order to achieve that starts with candidate extraction, continues with human description, classification and post-processing until the final detection result is achieved.

A summarized review of methodologies that were previously explored is done in [9], in order to identify advantages and disadvantages of these approaches. The taxonomy they use is an approach-based, the first level of categorization was made between hierarchical and single-layered approaches. Finally they provide the best applicable usage for a handful of methodologies.

Using depth data in the field of human motion analysis, as they do in [10], which has reached a new high with the development of Time-of-Flight cameras and other depth sensors. Face and hand motion are the center of the research in this survey. Finally an informatory summary and thorough evaluation for each area of study and different methodologies is provided.

Furthermore in [11] they deal with detection techniques, datasets and application. The first area interests us the most, which is further categorized into initialization, tracking, pose estimation and recognition. A comparison of the approaches is provided, where advantages and limitations of each one of them is discussed. A critical point that stands out in this paper and hasn't been mentioned up until now in our research is the assumptions and limitations that most researchers use to make, so that the methodologies produce acceptable results.

In [12] the authors are concentrated on approaches that aim on classifying full-body motions, like punching, kicking, and waving. The main drawback that is pointed out, is that significant progress needs to be made in uncontrolled environment in which the absence of background knowledge hinders the robustness of these methods.

A vision-based comprehensive review conducted in [13], considering a hierarchical activity scheme for the human body. A first level self-evaluation of methodologies is proposed towards improvement with the use of maturity metrics. This evaluation incorporates sets of features that found to best describe the importance of aspects of each methodology, regarding the operation, performance, and others, and acclaimed weights to them by their importance.

Another part of human activity recognition that is an active research topic in the computer vision community the last years is Gesture Recognition via visual information [14]. The classification based on recognizing a gesture, extracting the features, the classifiers that are being used and relevant real-time application are the ones that are analyzed in depth throughout the aforementioned research.

C. Descriptive Reviews

Here readers shall come across survey papers that delineate methodologies in some depth. In our opinion in this category we can understand the "what and how" of approach or methodologies. The papers included in this category are [15-21].

In [15] depth imagery and in particular Microsoft Kinect methodologies are studied. 3D human pose estimation is performed, either with the use body part detection or body pose modelling techniques, while human activity recognition is performed by breaking a high-level activity into sub-activities and create a link to hierarchical models.

Another paper [16] that deals with Microsoft Kinect studies the motion recognition. The approaches are divided into two categories template-based and algorithmic recognition. Algorithmic recognition uses manually defined rules in order to recognize a gesture or an activity. In contrast template based approaches, compare a gesture or an activity with pre-recorded template motion automatically, thus they try to find a pattern in order to produce an outcome.

The paper in [17] deals with activity recognition in surveillance mostly. For activity recognition they use the proposed categorization [18] into hierarchical and nonhierarchical approaches. Based on the interpretation, the nonhierarchical approach is divided into two classes, space-time approach, and sequential approach. Similarly, the hierarchical approach can be classified into three categories, statistical, syntactic, and description-based approaches.

In [19] the paper offers studies on human activity recognition approaches for surveillance. A categorization of the activities like in the previous survey is proposed, according to their complexity. The approaches studied for HAR are defined in unimodal and multimodal, and it is done according to the amount and type of data that they depend on, as well as on the ability of the algorithm that they use to deal with large scale and heterogeneous data.

A review of vision-based gait recognition methods for human identification was done in [20]. The authors categorize gait features into two groups, model-based and model-free features. Model-based are extracted via modeling or tracking components of human bodies, while model-free approaches give more emphasis on shapes of silhouettes or the whole motion of human bodies.

The paper in [21] presents recent developments in view-invariant human motion analysis with a given emphasis on view-invariant pose representation and estimation, and view-invariant action representation and recognition. View-invariant pose representation and estimation is separated into three categories based on the use of a prior human model, view-invariant action representation and recognition is categorized into template-based approaches and state-space approaches.

D. Brief Reviews

Throughout this category we are going to dwell with survey papers that offer brief information about researched methodologies and systems, we mostly expect to deal with information like the general field that a methodology belong, like spatiotemporal domain, and some introductory information about them. The research work that belongs to this category is [22-28].

In [22] the authors studied methods that use markerless human motion capture, and classified them according to the method they use to implement it. Throughout this paper we came across three main approaches that distinguish the methodologies that were implemented. First the segmentation of the human from the background, second the reduction of data complexity, and last the reduction of frames, since a general assumption is that the motion between neighbor frames is small.

Another area merely covered is the sensing technologies in [23]. Human Activity Recognition with the use of depth sensors, wearable devices and RGB cameras is studied here. In this paper, HAR is defined and approaches in surveillance, healthcare and human computer interaction are described. There is also a brief description of how the activity recognition is done with the use of the aforementioned sensors.

Human actions with the use of vision-based techniques was studied in [24]. A nicely presented general overview of the recognition task is performed, henceforward the segmentation process is described in more detail with the use of local features because there are more effective against noise and occlusion, in contrast with space time volumes and discrete Fourier transformation. Finally pose estimation technique and motion history are referenced in some detail.

In [25] there is another study that dwells with vision based techniques. The main contributions of this paper are, the general process of feature-based modeling and detection of anomalies in video sequences is described, as well as the research shift from rule-based to statistical-based methods for modeling scene behavior and methods for doing that are pointed out. In the end unsupervised, supervised, and semi supervised frameworks to train the previous approaches, are presented, pros and cons of them are analyzed.

In [26] the focus is in the same area as the previous two papers, but the focal point is in methods for action recognition and event detection. Event abnormalities, suspect retrieval, and event and action detection, are briefly covered. Tracking-based methods for retrieval of subjects and objects are in the spotlight as well.

Closely related to the aforementioned studies is the one that we are going to talk about next, which deals with multi-view action recognition, another field less exploited in [27].

They present 3D methods that require a fixed number of cameras setup in order to train and test them, the most exploited approach is the fusion of the visual information received by the different angles and next proceed with the action representation and classification. In contrast 2D methods are regarded as the ones that use an arbitrary number of cameras, and approach the problem from two different directions. They either adopt a single-view, view-independent approach or a cross-view action recognition approach.

In [28] the authors conducted a comparison and analysis between various methods involved in the detection of human action recognition automatically. This kind of comparison reflects that the efficiency differs from each method. This paper shows that all the discussed methods are based on pixel-based approaches, which gives better resolution to the image.

III. EVALUATION METRIC

In this section, we present 8 aspects by which the selected Survey Papers are going to be evaluated. We assigned weights based on perspectives gained from professors, graduate students, postdocs, and researchers, who have a close relation with the field of HAR. A comparative evaluation of the presented survey papers described in previous sections is going to be done with the use of aspects and weights. Based on clustering similar positive and non-positive characteristics we create a maturity formula for testing methodologies with datasets, providing relations and contradictions between methodologies. The final scores obtained from this process are multiplied together and squared to provide a better evaluation. Accordingly, in readability, originality, usefulness and reliability are related, but dividing them with complexity and completeness, provides us with an unambiguous outcome.

Finally balancing efforts and outcomes of approaches and analysis of them, are added together since they are closely related features. We are going to explain all of the features that we used, in the table below and also provide detailed information about the pointing system of our evaluation metric. In order to provide results as trustful and truthful as possible, we provided thorough reports of the survey papers to a group of professionals. Since survey papers contain a lot of information we decided that all survey papers should be evaluated by at least seven different reviewers. For each system component or methodology presented in the sections below, we present a score for each aspect of the proposed approach and a score weighted by the perspective brought through our studied surveys. Each survey was given a score based on the degree to which it accounted for a given aspect: does not account for the aspect (0.5), somewhat accounts for the aspect (1), mostly accounts for the aspect (2), and fully accounts for the aspect (3).

The proposed evaluation metric is the following:

$$\mathbf{M} = [\frac{\mathbf{f7} + \mathbf{f5} + \mathbf{f10} + \mathbf{f8}}{\mathbf{f3} + \mathbf{f4}} + (\mathbf{f1} + \mathbf{f2}) + (\mathbf{f6} + \mathbf{f9})^{\mathbf{2}}]$$

TABLE I. EVALUATION METRIC FEATURES

Features	Features Analysis
F1 Analysis	How thoroughly studied an approached or methodology is.
F2 Balancing	If in each group of studied methodologies, the researchers in the end balances the effort and outcomes, of each one of them, and as a group of them.
F3 Complexity	It reflects how thoroughly a grouping of methodologies is done, why they did it, and how they help the reader with it.
F4 Completeness	Number of referenced paper that can be found across other surveys, here we used a standardized grading system.
F5 Originality	Whether the survey promotes original ideas and has really contributed to the better understanding of an approach or methodology, for the reader.
F6 Outcomes	Whether the survey contrasts methodologies, in order to find the pros and cons of each one of them, and between them. Also the area in which, each approach works the best should be referred.
F7 Readability	How well written a survey paper is.
F8 Reliability	Why an approach or methodology works more favorable towards a specific procedure, how is that justified.
F9 Robustness	Why should we rely when a writer says, that a methodology is robust, very accurate. Are results provided for these sayings?
F10 Usefulness	How the survey covered reader's needs, when he/she studied it.

Our next step is to depict the results of the aforementioned evaluation of the studies. When we examined the feedback from fellow PhD candidates, professors, postdocs and researcher, we came across some interesting findings.

- The ones that were fully informed about the area of human activity recognition, gave higher scores to the subcategories of comparison category.
- Graduate students, and PhD candidates that were in their early years of study, or weren't up to date with the advances in the area, gave good scores to the review category papers, mostly because they could understand the greater picture.
- A lot of papers that don't belong to Deep Evaluation or First Level Evaluation categories received low

scores in the following features balancing, completeness, outcomes and robustness.

- All the participants of the survey gave their highest scores for the studies that belong to the Deep Evaluation category, because approaches were analyzed thoroughly enough, as well as they provided facts regarding the recognition ratio, applicability of them and other related features.

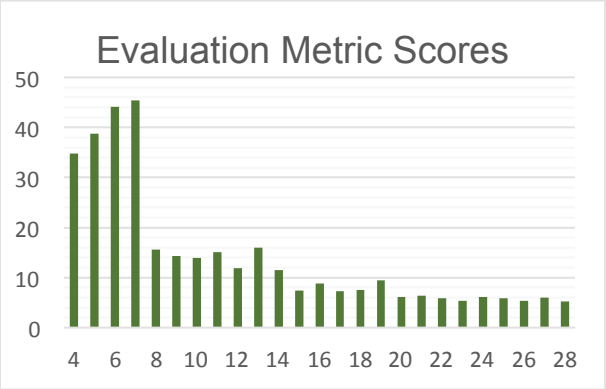

Fig. 2. Evalution Metric Scores

IV. CONCLUSSIONS

A systematic study was performed for categorizing survey papers using google scholar. We searched for survey papers or review papers, which were written for human activity recognition methodologies or approaches, although this study can be used in different areas as well. The main objective was to provide the reader with relevant information regarding already existing research work and some valuable criteria regarding his/her interests. Our search led to a selection of 69 journals and conference papers that were published from 2010-2015. From that set of papers we analyzed only 24 in this effort without limiting the scope of our effort. This survey might have limitations and we don't claim that is exhaustive since there might be not included related papers that were not identified, but the main goal was the categorization of survey papers and their importance.

Our research provided us with some very important pieces of information, regarding each proposed category of survey papers, as well as general inductions regarding the bigger picture of surveys that were written for the human activity recognition area. Although year by year the number of surveys increases only a very small number of them covers thoroughly the studied material in a comparative way. Also we have seeing that a lot of papers have wrong paper titles or authors name, this might belong to pretty typing, but sometimes it makes it difficult for the reader to find a specific paper that might interest him or her.

Regarding the conclusions from each one of the four categories, the findings are the following. Papers that belong to either Deep Evaluation or First Level Evaluation categories, offer valuable inside regarding each methodology and also for a group of methodologies that are used to implement the same or similar tasks. Specifically, the Deep Evaluation category can offer accurate information regarding the performance for most of the studied approaches, and sometimes hints about the robustness and the complexity of them. The First Level Evaluation category provides information regarding the maturity of each methodology towards completeness and its potential applications.

Moreover surveys that belong to the Descriptive Reviews category, feeds us with descriptive information, as the First Level Evaluation, but usually not as thoroughly and mainly without evaluation. Finally in this category it is not expect to come across a holistic view of category specific approaches, nor about the accuracy of them. Brief Reviews provides us with the general picture of each methodology and not specific information, implementation attributes, or reasoning of how they work.

In summary, our effort here was to evaluate survey papers regarding their usefulness and at the same time offer to future survey efforts, a point of reference on what type of survey is needed in the research field. It is also important to mention again that each survey is useful because it provides us with different type of information that is unlikely for someone to find them faster than reading a single survey study, and distinguish approaches that interest them. Although this kind of research should include as much information as possible in order to correctly assess the validity of methodologies that are presented, however we believe that our message here is the categorization and its importance to researchers in the field.

REFERENCES

[1] Morrison, Kevin A., Jonathan Culler, and Kevin Lamb. "Just being difficult? Academic writing in the public arena." (2006): 35-38.

[2] Furman, Rich, and Julie T. Kinn. Practical Tips for Publishing Scholarly Articles: Writing and Publishing in the Helping Professions. Lyceum Books, 2011.

[3] Murray, Rowena. Writing for academic journals. McGraw-Hill Education (UK), 2013.K. Elissa, "Title of paper if known," unpublished.

[4] Bulling, Andreas, Ulf Blanke, and Bernt Schiele. "A tutorial on human activity recognition using body-worn inertial sensors." ACM Computing Surveys (CSUR) 46.3 (2014): 33.

[5] Lim, Chern Hong, Ekta Vats, and Chee Seng Chan. "Fuzzy human motion analysis: A review." Pattern Recognition 48.5 (2015): 1773-1796.

[6] Dawn, Debapratim Das, and Soharab Hossain Shaikh. "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector." The Visual Computer (2015): 1-18.

[7] Holte, Michael B., et al. "Human pose estimation and activity recognition from multi-view videos: Comparative explorations of recent developments." Selected Topics in Signal Processing, IEEE Journal of 6.5 (2012): 538-552.

[8] Nguyen, Duc Thanh, Wanqing Li, and Philip O. Ogunbona. "Human detection from images and videos: A survey." Pattern Recognition 51 (2016): 148-175.

[9] Aggarwal, Jake K., and Michael S. Ryoo. "Human activity analysis: A review." ACM Computing Surveys (CSUR) 43.3 (2011): 16.

[10] Ye, Mao, et al. "A survey on human motion analysis from depth data." Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications. Springer Berlin Heidelberg, 2013. 149-187.

[11] Afsar, Palwasha, Paulo Cortez, and Henrique Santos. "Automatic visual detection of human behavior: a review from 2000 to 2014." Expert Systems with Applications 42.20 (2015): 6935-6956.

[12] Weinland, Daniel, Remi Ronfard, and Edmond Boyer. "A survey of vision-based methods for action representation, segmentation and recognition." Computer Vision and Image Understanding 115.2 (2011): 224-241.

[13] Tsitsoulis, A. and N. Bourbakis. "A first stage comparative survey on human activity recognition methodologies." International Journal on Artificial Intelligence Tools 22.06 (2013): 1350030.

[14] Attolico, C., et al. "A Review of Gesture Recognition Approaches for HRI." Workshop on Real-Time Gesture Recognition for Human Robot Interaction, in conjunction with IAS. 2014.

[15] Chen, Lulu, Hong Wei, and James Ferryman. "A survey of human motion analysis using depth imagery." Pattern Recognition Letters 34.15 (2013): 1995-2006.

[16] Lun, Roanna, and Wenbing Zhao. "A survey of applications and human motion recognition with Microsoft Kinect." International Journal of Pattern Recognition and Artificial Intelligence 29.05 (2015): 1555008.

[17] Vishwakarma, Sarvesh, and Anupam Agrawal. "A survey on activity recognition and behavior understanding in video surveillance." The Visual Computer 29.10 (2013): 983-1009.

[18] Aggarwal, Jake K., and Michael S. Ryoo. "Human activity analysis: A review." ACM Computing Surveys (CSUR) 43.3 (2011): 16.

[19] Vrigkas, M., C. Nikou, and I. A. Kakadiaris. "A Review of Human Activity Recognition Methods." Front. Robot. AI 2: 28. doi: 10.3389/frobt (2015).

[20] Wang, Jin, et al. "A review of vision-based gait recognition methods for human identification." Digital Image Computing: Techniques and Applications (DICTA), 2010 Int. Conference on. IEEE, 2010.

[21] Ji, Xiaofei, and Honghai Liu. "Advances in view-invariant human motion analysis: a review." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 40.1 (2010): 13-24.

[22] Alotaibi, Najm. "Human Motion Capture: New Innovations in the Field of Computer Vision." World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering 9.3 (2015): 694-697.

[23] Ann, Ong Chin, and Lau Bee Theng. "Human activity recognition: A review." Control System, Computing and Engineering (ICCSCE), 2014 IEEE International Conference on. IEEE, 2014.

[24] Hassan, Muhammad, et al. "A Review on Human Actions Recognition Using Vision Based Techniques." Journal of Image and Graphics 2.1 (2014): 28-32.

[25] Ke, Shian-Ru, et al. "A review on video-based human activity recognition." Computers 2.2 (2013): 88-131.

[26] Poppe, Ronald. "A survey on vision-based human action recognition." Image and vision computing 28.6 (2010): 976-990.

[27] Iosifidis, Alexandros, Anastasios Tefas, and Ioannis Pitas. "Multi-view human action recognition: A survey." Intelligent Information Hiding and Multimedia Signal Processing, 2013 Ninth International Conference on. IEEE, 2013.

[28] Dhivya, A., S. Saranya Devi, and M. Jayasudha. "A Survey on Action Recognition." (2014).