

A Compilation of Methods and Datasets for Group and Crowd Action Recognition

Luis Felipe Borja, Universidad Central del Ecuador, Quito, Ecuador

Jorge Azorin-Lopez, Department of Computer Technology, University of Alicante, Alicante, Spain

Marcelo Saval-Calvo, Department of Computer Technology, University of Alicante, Alicante, Spain

ABSTRACT

The human behaviour analysis has been a subject of study in various fields of science (e.g. sociology, psychology, computer science). Specifically, the automated understanding of the behaviour of both individuals and groups remains a very challenging problem from the sensor systems to artificial intelligence techniques. Being aware of the extent of the topic, the objective of this paper is to review the state of the art focusing on machine learning techniques and computer vision as sensor system to the artificial intelligence techniques. Moreover, a lack of review comparing the level of abstraction in terms of activities duration is found in the literature. In this paper, a review of the methods and techniques based on machine learning to classify group behaviour in sequence of images is presented. The review takes into account the different levels of understanding and the number of people in the group.

KEYWORDS

Computer Vision, Crowd Automated Analysis, Human Behavior Analysis, Machine Learning, Motion Analysis, Trajectory Analysis

1. INTRODUCTION

Nowadays, video surveillance of people is a widely used tool because there are many cameras that facilitate the capture and storage of video. Most of these products are dependent on an operator to analyze the content of stored information. Knowing this limitation, it is necessary to provide systems of video surveillance that make possible the automatic identification of behavior. These types of system can be carried out using computer vision techniques, since they allow the identification of patterns of people behavior in an unsupervised manner as gestures, movements and activities among others. In general terms, machine learning, it is possible to model the behavior of people in open or closed spaces such as universities, shopping malls, parks or streets, and then analyze them using automatic learning methods.

There are currently many researches on Human Behavior Analysis such as, (Azorin-Lopez et al., 2015) that have resulted in the identification of various types of people's behavior in video sequences. These behaviors have been classified from the simplest to the most complex taking into account their duration, from seconds to hours. For these behaviors, a classification has been proposed in (Chaaroui et al., 2012).

The objective of this paper is to provide a classification of human behavior analysis proposals taking into account the size of the group or crowd, identifying the number of people that comprises

it, the type of behavior detected, the level of abstraction (from simple actions to complex behaviors) and the techniques used for its treatment and analysis. The most important public datasets are also reviewed which are used to test algorithms there exist several studies on the identification of human behaviors such as (Chaaroui et al., 2012), (Cardinaux et al., 2011), (Turaga et al., 2008), (Ryoo & Aggarwal, 2008). In (Mihaylova et al., 2014) a taxonomy of groups with fewer and more members is established, in addition the methods to analyze them are specified. There are works such as (Climent-Pérez et al., 2014), where it is proposed to analyze the behavior of crowds by classifying them into two levels, macro and micro. Despite research efforts to analyze behavior in groups and crowds, we still have many fronts on this subject for researchers.

The organization of the paper is as follow: Section 2. Aspects of human behavior, levels of semantics and datasets; Section 3. Classification of the state of the art proposals; finally, conclusions and possible future works.

2. ASPECTS OF HUMAN BEHAVIOR ANALYSIS

In this section, the main aspects of the human behavior analysis are explained. First, we will present the different levels of understanding and later the main datasets available for experimentation.

2.1. Description of Human Behavior Types and Semantics (Gesture, Motions, Activities, Behavior)

In order to identify human behavior according to the level of abstraction and understanding the data has to be classified depending on the meaning, duration and complexity of tasks performed by humans.

Classifications of activities have taking as its main reference the level of complexity of them, from the easiest to the most complex. The complexity factor is directly related to the time duration of the activity, generally, an activity is considered complex if it has a longer duration. In (Vishwakarma & Agrawal, 2013) four levels related to their semantics:

- **Level 1 (Gestures):** Basic movements of parts of the body that last a time. Examples of gestures can be movements of the hand, arm, foot or head among others.
- **Level 2 (Actions):** Also called atomic, consists of actions performed by a single person, their duration is larger than a gesture. An example of actions could be walking, running, jumping.
- **Level 3 (Interaction):** In this category human-human or human-object interaction activities are performed. Examples of these interactions can be two people dancing, kissing, running one behind another, children playing, people cycling.
- **Level 4 (Group Activity):** At this level of description it conforms to two or more groups of people, one or more objects can intervene in the scene. An athletic race, basketball team forwarding, pedestrians crossing a street, a football game, a fight in a stadium can be examples of group activities.

Another taxonomy of human behavior that classifies it according to the complexity and duration time is proposed in (Chaaroui et al., 2012). In this approach, the analysis is classified on the degree of semantics in four levels:

- **Level 1 (Motion):** Detection in seconds or frames.

- **Level 2 (Action):** Detection of simple tasks in terms of seconds. The human can interact with objects, or be sitting, standing, walking.
- **Level 3 (Activity):** These are tasks from of minutes to hours. They constitute the sequence of actions, such as cleaning a room, washing a vehicle.
- **Level 4 (Behavior):** This is the higher level of understanding since its duration time can be hours and days. Example behavior can be daily routines of a person, personal habits, and mix of two activities in logical sequence.

Both taxonomies described above are based on the daily activities of people, taking into account important factors such as the level of semantics, the duration and the activities composed of other simpler parts such as movements and actions. They described the levels/orders of behavior from the simple movements lasting seconds, to complex activities performed by people for several minutes, hours and even days. The aim of the researchers has been to propose a general classification human behavior. There are other classification, however, in this work we are going to base our proposal on these focused on group and crowd behavior classification

2.2. Specialized Datasets

In (Blunsden & Fisher, 2010) Blunsden and Fisher presented a set of datasets which include sequences for individual and group behavior which are part of the BEHAVE project and include some form of ground truth. Since this paper is focused on group and crowd analysis, the individual datasets are not studied, but authors refer to the original paper for further details.

In group analysis, there are three datasets belonging to BEHAVE project: CAVIAR, CVBASE, ETISEO. Examples of behavior detected in these datasets are: InGroup (The people are in a group and not moving very much), Approach (Two people or groups with one (or both) approaching the other), WalkTogether (People walking together), Meet (Two or more people meeting one another), Split (Two or more people splitting from one another), Ignore (Ignoring of one another), Chase (One group chasing another), Fight (Two or more groups fighting), Run-Together (The group is running together), Following (Being followed).

In this paper we analyze the behavior of groups and crowds such as pedestrians, crowds in public places such as stadiums or squares, interactions of large and small groups, sport actions such as soccer and basketball, and others. The datasets used by the researchers are numerous, being the main ones the following: BEHAVE, BIWI, VSPETS, ETH, DGPI, UHD, HMDB, SportsVU, PETS, UNM, ViF, Bus STATIONS, Subway STATIONS, others. Also in some cases the researchers use their own datasets or videos obtained on YouTube. In (Chaquet et al., 2013) it is proposed a study and dataset classification taking into account the behaviors, number of people involved, techniques used to recognize behaviors, types of scene, year of publication, among other characteristics. From this study, an absence of RGB-D (Color and depth) datasets is shown.

With the objective of studying human behavior, in the last years several public datasets have been created. In these dataset, video sequences with contents of several activities in different scenarios and situations are stored. There are also sites dedicated to study particular activities such as a movement or action of a sport, identification of abandoned objects, or daily activities (ADL) such as having a cup of coffee, detection of falls of human, gait study, gesture analysis. These studies are directly related to public datasets, where tests of the algorithms and techniques used in each case are performed, in certain studies more than one dataset is used to check the accuracy of the recognition systems developed, in other cases it is used custom datasets or the researcher's own, video sequences obtained in public places like bus stations or trains, also of people who carry out activities in squares, streets and commercial centers of a city, are also used. There are very few studies that use YouTube as a source for video footage for research.

Performing video analysis studies require effort and time for researchers; thousands of man-hours are used for the labeling of the different situations that need to be identified in a video. Currently, in

cities, it is common to find camcorders capturing video that are later stored. However, all this large amount of information is not available for public access and experimentation.

The main datasets used in Table 1 as reference in the present work are described below, they specify dataset name, the type of behavior, the general and specific characteristics how frames per second, video size, actions (Boxing, Clapping, Waving, Walking, Jogging, Running). We can also specify if the scene is inside or outside, that identify the dataset:

3. CLASSIFICATION OF THE LEVEL OF UNDERSTANDING OF GROUPS

To analyze human behavior by using video surveillance cameras, a system based on computer vision requires following a series of ordered steps as suggested in (Banos et al., 2013). This paper aims to organize a classification of human behavior according to the number of people that make up a group or crowd, and the techniques, algorithms or frameworks used for analysis. The reviewed papers are presented in Table 2. The most relevant proposals in our opinion, regarding the number of references to the paper and its relation to crowd and group behavior study, are more deeply explained in this section.

Human behavior analysis (HBA) investigations have different applications: improving the quality of life of human beings, in aspects such as support in the health area to detect unusual behaviors, for example falls of elderly people in assisted living environments (AAL) (Bruckner et al., 2012), (Banos et al., 2013), (Cardinaux et al., 2011); surveillance of pedestrians, fights, people running, assaults, ingesting liquor in public places, for example.

The classification of tasks performed by humans described in the previous section is analyzed in (Cardinaux et al., 2011) according to the level of semantics (in ascending order according to the duration time of this is): Movement (seconds), actions (seconds, minutes), activity (minutes, hours), behavior (hours, days). Each of these tasks must be recognized and modeled, using different techniques, algorithms and other tools suitable for this task.

Turaga et al. (Turaga et al., 2008) proposed a scale of recognition of human activities from simple (actions) to complex (activities), for actions called simple uses (Non-Parametric, Volumetric, and Parametric), for activities called complex uses (Graphical Models, Syntactic, and Knowledge Based). Another organization proposal for recognition of activities is set out in (Banos et al., 2013), where it proposes the Chain of Activity Recognition. This approach divides the recognition process into different procedures, which are: Data Acquisition, Preprocessing, Segmentation, Characteristic extraction, Classification, Decision. Most current research focuses on the last two procedures of this proposal and is often referred to as the learning and decision phases.

In the studies about human behavior of groups and crowds analyzed, it was found that there are few works dealing with RGBD cameras and analysis of human behavior using 3D information. It is important to highlight the work of Wu et al. (Wu et al., 2015). They proposed the MoSIF method is combined with HMM (Wu et al., 2015) to analyze video sequences obtained from a Microsoft Kinect RGBD device. The accuracy obtained is 60% for 3600 video sequences. However, according to the authors, a better result could be obtained if they used more videos to improve learning.

The methods of classification can be supervised and not supervised, and can be used individually or combined using boosting techniques.

On the subject of behavior and trajectories of groups of people there are also some approaches that are based on (HBA) study individually, for example: to recognize activities of groups of people we use the Group Activity Descriptor Vector (GADV) Proposed in (Azorin-Lopez et al., 2016). This method has as its predecessor the Activity Description Vector (AVD) revised in (Azorin-Lopez et al., 2016), (Azorin-Lopez et al., 2014), and aims to recognize human behavior in advance.

3.1. Features of a Groups and Crowds

For example, Andrade et al. (Andrade et al., 2006a) detected behavior of a crowd in different scenarios considered unusual or an emergency, usually provoked by a minority of people in the

Table 1. Classification of datasets

Dataset	Type	General Feature	Specific Feature	In-Outdoor
ASLAN (Kliper-Gross, Hassner & Wolf, 2012)	Action Recognition	The Action Similarity Labeling (ASLAN) Challenge: a video database of actions and a full testing protocol for studying action similarity from videos.	Contains 3697 action samples from 1571 unique YouTube videos divided into 432 non-trivial action categories.	INDOOR OUTDOOR
BEHAVE (Fisher et al., 2011)	Action Recognition	Investigate two novel computer-based image analysis processes to prescreen video sequences for abnormal or crime-oriented behavior	The data is captured at 25 frames per second. The resolution is 640x480. The videos are available either as AVI's or as a numbered set of JPEG single image files	OUTDOOR
BIWI (Pellegrini et al., 2009)	Walking Pedestrians	Walking pedestrians in busy scenarios from a bird eye view. Manually annotated	Modeling Social Behavior for Multi-target Tracking	OUTDOOR
Broadcast Television Action (Rodríguez, Ahmed, & Shah, 2008)	Sports actions	Set of actions from various sports featured on broadcast television channels such as the BBC and ESPN	Diving, golf swinging, kicking, lifting, horseback riding, running, skating, swinging a baseball bat, and pole vaulting	OUTDOOR
CAVIAR (Aggarwal & Ryoo, 2011)	Action Recognition	For the CAVIAR project, a number of video clips were recorded acting out the different scenarios of interest	Walking alone, meeting with others, window shopping, entering and exiting shops, fighting and passing out and last, but not least, leaving a package in a public place.	INDOOR
COLLECTIVE (Choi, Shahid & Savarese, 2009)	Action Recognition	Contains 5 different collective activities in 44 short video sequences some of which were recorded by consumer handheld digital camera with varying viewpoint.	Crossing, Walking, Waiting, Talking, and Queueing	OUTDOOR
FIFA WC 2006 (Liu, Tong, Li et al., 2009)	Sports (Soccer)	The testing results on FIFA World Cup 2006 video	Test our algorithm on World Cup 2006 MPEG2 videos (image size: 720x576, 25 fps)	OUTDOOR
FIFA WC 2007 (Oskouie, Alipour & Eftekhari-Moghadam, 2014)	Sports (Soccer)	The testing results on FIFA World Cup 2007 video	Approach is used for goal, booking, penalty and corner detection and others.	OUTDOOR
HOHA (Laptev, Marszalek, Schmid et al., 2008)	Action Recognition	HOHA (Hollywood Human Actions) dataset includes 10 types of actions extracted from movies. Almost all of the sequences can be considered within the moving camera domain	Provides the video samples and annotations used in the experimental section of the paper "Learning realistic human actions from movies"	INDOOR OUTDOOR
INRIA (Poppe et al., 2010)	Action Recognition	This dataset was collected as part of research work on detection of upright people in images and video	Check watch, Cross arms, Scratch head, Sit down, Get up, Turn around, Walk, Wave, Punch, Kick, Point, Pick up, Throw (overhead), Throw (from bottom up).	OUTDOOR
IXMAS (Poppe et al., 2010)	Action Recognition	Dataset composed of the IXMAS actions, but performed by different actors, who could be partially occluded.	Wave hand, Kick, Punch, Check watch, Cross arms	INDOOR
KTH (Solmaz, Assari & Shah, 2013)	Action Recognition	Includes videos captured in a controlled setting of six action classes with 25 subjects for each class. In total it contains 598 video sequences	Boxing, Clapping, Waving, Walking, Jogging, Running	INDOOR OUTDOOR
Parking Lot 1 (Shah et al., 2012)	Groups of pedestrians	Consists of two video sequences collected in a parking lot using a static camera	Is a moderately crowded scene including groups of pedestrians walking in queues with parallel motion and similar appearance	OUTDOOR
Parking Lot 2 (Shah et al., 2012)	Groups of pedestrians	Consists of two video sequences collected in a parking lot using a static camera	Is a more challenging sequence due to the large amounts of pose variations and occlusions	OUTDOOR
PETS (Aggarwal & Ryoo, 2011)	People motion	Performance Evaluation of Tracking and Surveillance		
PETS 2004 (Aggarwal & Ryoo, 2011)	Action Recognition	Performance Evaluation of Tracking and Surveillance		
PETS 2009 (Ferryman et al., 2009)	Action Recognition	This is a relatively sparse scene including a few people walking in random directions		OUTDOOR
PETS-S2L1 (Dehghan et al, 2016)	Action Recognition	Includes 800 frames with a challenging scenario because of frequent changes in the directions of the pedestrians		OUTDOOR
PNNL Parking lot Dataset (Shah et al., 2012)	Groups of pedestrians	Consists of two video sequences collected in a parking lot using a static camera	Is a moderately crowded scene including groups of pedestrians walking in queues with parallel motion and similar appearance	OUTDOOR

continued on following page

Table 1. Continued

Dataset	Type	General Feature	Specific Feature	In-Outdoor
Skateboarding Dataset (Shu, Afshin & Mubarak, 2013)	Action Recognition	Consist of two video sequences captured by a hand-held camera	Due to the camera motion and severe pose changes	OUTDOOR
Sub-JHMDB (Jhuang, Gall, Zuffi et al., 2013)	Action Recognition	Contains 12 complex human actions	Catch, Climb stairs, Run, Jump, Swing basketball	INDOOR OUTDOOR
THUMOS13 (Jiang, Liu, Zamir et al., 2013)	Action Recognition	Is the largest and the most challenging trimmed action detection dataset with 24 complex human actions	Pole vault, Skiing, Ski-jet, Surfing, Fencing, Cricket bowling	OUTDOOR
Town Center (Bengio, Courville & Vincent, 2012)	Action Recognition	This is a semi-crowded sequence with rare long-term occlusions	The motion of pedestrians is often linear and predictable	OUTDOOR
UCF Sports Dataset (Khurram & Amir, 2008)	Sports actions	Is a challenging dataset with sequences mostly acquired by moving cameras. It includes 10 sports actions with a total of 150 sequences	Diving, Golf swinging, Kicking, Lifting, Horseback riding, Running, Skating, Baseball Swinging, Walking.	INDOOR OUTDOOR
UCF101 (Khurram & Amir, 2012)	Action Recognition	Action recognition on a subset of 12 HMDB51 categories with no additional training	Brush hair, Climb, Dive, Dive, Golf, Handstand, Pullup, Punch, Pushup, Ride bike, Shoot ball, Shoot bow, Throw	INDOOR OUTDOOR
WEIZMAN (Blank, Gorelick, Shechtman et al., 2005)	Action Recognition	Contains 10 actions performed by 9 different persons. There are total of 92 video sequences	Walk, Run, Jump, Gallop sideways, Bend, One-hand wave, Two-hands wave, Jump in place, Jumping jack, Skip.	OUTDOOR
YouTube Action Dataset (Liu, Yang, Saleemi et al., 2012)	Action Recognition	Consist of 11 categories with about 1160 videos	Walking, Jumping, Diving, Cycling, Shooting, Riding, Swinging, Spiking, Juggling	OUTDOOR

crowd. These behaviors are coded in Hidden Markov Models (HMM) with mixture of Gaussians output (MOGHMMs), detecting within the different scenes according to their density of people that conform it. It should be considered that the system must be previously trained to detect a type of behavior considered normal that usually have the majority of members of a crowd analyzed. Analyzing specifically the modeling of dense crowds is still an open problem of researchers.

In a public space, where there are a lot of people, the behavior could be analyzed by two variables: actions and duration.

A general trend could be noticed and described as the actions considered normal ones have an extended duration ... a general trend that would be described as that the behaviors considered normal ones have an extended duration, in which most people make up the crowd, while the behaviors considered abnormal are caused by few people in the crowd and in short times of duration. For the study of these types of behaviors, Hu et al. (Hu et al., 2013) proposed to use a statistical exploration method analyzing the video in a separate way as sliding windows in which the behaviors considered anomalous are detected, taking into account that the algorithm used in this technique requires monitoring.

As we have previously described in order to understand the behavior of crowd, we must take into account the social behavior of the masses, since in this one can observe patterns of behavior that can be modeled by computer studying their structure and special characteristics as proposed in (Ge et al., 2009). This study analyzes the human activity of medium level in the granularity, that is to say in the number of people that conform it based on algorithms for the detection of pedestrians and tracking of several moving objects. A particular fact is that the study considers small groups of people traveling together considering the hierarchy of smaller to larger size of the group. It takes into account the proximity of pairs of people and their speed when walking in a particular scene. According to (Ge et al., 2009), a group is formed from two people, in addition it must meet other parameters such as: if they are within 2,13 meters of each other and not separated by another in the middle, have the same speed up to within 0,15 meters per second, and is traveling in the same direction within 3 degrees. When a member of the group stops fulfilling these characteristics or complies with them, it can be

said that he or she is inside or outside the group. The conditions for a group to become crowded have not been precisely defined by researchers, however what if it is clear is how a crowd is analyzed, among the main features of analysis are; Treat it as a single mass and detect movement considered abnormal, count the number of heads, studies about crowd are; (Rodriguez, Ahmed & Shah, 2008), (Junior, Musse & Jung, 2010), (Li, Mahadevan & Vasconcelos, 2014), (Zhan, Monekosso, Remagnino et al., 2008), (Zhou, Tang & Wang, 2013).

The datasets can be chosen by the researchers according to their criteria, taking into account the suitability for their objective. The data are grouped into two categories the heterogeneous, referring to the general activities and the specific when these actions have a special treatment. A third category is included in (Chaquet et al., 2013), which specify techniques for motion capture such as the use of infrared, thermal and motion capture (MOCAP).

3.2. Behavior of Groups and Crowds

This paper shows in Table 2 a classification of the group size according to the number of members and the activities that each type of group performs, besides specifying the methods, algorithms and forms of recognition that can be used for their study, the order of the table is made according to the level of semantics and the name of the author. We can see the following analyzed fields: Ref = Reference to the article, CL = Classification (G = Group (number of people) if exist, and C = Crowd), TE = Technique, D = Dataset, LA = Level Abstraction. In the column LA = Level Abstraction we show three levels of abstraction: Mot = Motion, Act = Action, Actv = Activity, also two automatic tasks, CP = Count-People and Tra = Tracking.

The classification of the papers analyzed according to the number of people that make up the groupings is GROUP and CROWD. Group is defined as the rapprochement of two or more people in a given site and performing an action or activity, Crowd is grouping a number of people over a group with large groupings that usually performs simultaneous activities. The types of behaviors analyzed using video surveillance are limited and specific in relation to the whole universe of behaviors that a group or multitude of people can have in a real environment. The behaviors analyzed most frequently in the papers are the following:

Tracking, trajectories, bicyclist, pedestrian, skateboarders, count people in a group or crowd, street fights, interaction objects-people, motions or actions in sports, human actions (walking, jogging, running, boxing, hand waving and hand clapping). The DATASETS frequently used for the tests of the techniques, algorithms and systems developed to analyze the behavior of groups and crowds are the following: BEHAVE, BIWI, CAVIAR, VSPETS, ETH, DGPI, UHD, HMDB, SportsVU, PETS, UNM, ViF, Bus STATIONS, Subway STATIONS, others. Also in some cases the researchers use their own DATASETS or videos obtained on YouTube. Based on the information analyzed in the papers, it is possible to propose a classification according to the level of abstraction of the analyzed human behavior of groups and crowds according to the case, in order of shortest to longest duration of behavior we propose five levels of abstraction:

Motion, Action, Activity, Count-People and Tracking. The techniques or methods frequently used to analyze human behavior of groups and multitudes using video surveillance are as follows: Bag of Words, Deep Neural Networks, Hidden Markov Models, Monte Carlo, Gaussian Mixture Model, Multiple Human Tracking, and Support Vector Machines. Also, many of the authors adopt custom names for their methods and techniques used in research, another option used by researchers to name a particular method or technique is the combination of one or more algorithms. The topics covered in this paper are mainly groups of small and large people (between 2 up to 50 people), pedestrians, crowds (over 100 people), sports teams (basketball and soccer), people walking in parks, metro stations and buses. Most researchers perform functionality tests of their human behavior recognition algorithms, methods or systems in one or more specialized DATASETS, few jobs that can display real-time video sequence analysis results.

Table 2. Classification of proposals reviewed

Reference	CI	Technique	Dataset	La
(Andrade et al., 2006a)	C	Hidden Markov Model (HMM)	N	Mot
(Brostow and Cipolla, 2006)	C	Unsupervised Bayesian Clustering Framework(UBCF)	N	Mot
(Cao and Huang, 2013)	C	Accumulated Mosaic Image Difference (AMID) OpticalFlow+BackgroundModel (OFBM) Markov Random Fields (MRF) Support Vector Machine (SVM)	Subway Station Bus Station Plaza	Mot
(Climent-Perez et al., 2014)	G(20)	Bag of words modelling (BoW)	Novel Dataset	Mot
(Gning et al., 2011)	G	Evolving Networks (EN) Monte Carlo (MC)	N	Mot
(Jacques et al., 2007)	G	Voronoi Diagrams Model(VDM)	N	Mot
(Li et al., 2014)	C	Model Dynamic Textures Temporal (MDT-temp) Local Motion Histogram (LMH) Model Dynamic Textures Spatial (MDT-spat)	UNM UCSD	Mot
(Li et al., 2014)	G	Hidden Markov Models (HMM) Dynamic Probabilistic Networks (DPN)	HMDB BEHAVE	Mot
(Pellegrini et al., 2010)	G	Linear Trajectory Avoidance (LTA)	N	Mot
(Ryoo & Aggarwal, 2008)	G	Dynamic Probabilistic Networks (DPNs) Dynamically Multi-Linked (DML) Hidden Markov Model(HMM)	PETS 2004 YouTube	Mot
(Shao et al., 2014)	C	Collective Transition priors (CT) Mixture of dynamic texture (DTM) Hierarchical clustering (HC) Coherent filtering (CF)	CUHK	Mot
(Yi et al., 2015a)	C	Pedestrian Simulation(PS) Person re-identification (PT) Pedestrian tracking(MPF)	NY Station Shanghai- ExpoNY Station	Mot
(Yi et al., 2015b)	C	Motion Pattern Features(MDA)	N	Mot
(Zhang et al., 2012)	C	Bag of Words (BoW) Locality-constrained Linear Coding (LLC) Vector Quantization (VQ)	BEHAVE	Mot
(Chang et al., 2010)	C	Fast Corner Detect(FAST) Support Vector Machine (SVM)	BEHAVE	Act
(Fradi & Dugelay, 2016)	C	Hidden Markov Models (HMM) Support Vector Machine (SVM) Robust Local Optical Flow (RLOF)	PETS UMN	Act
(Gong et al., 2014)	G	Cumulative Match Characteristic (CMC) Synthetic Disambiguation Rate (SDR) Center Rectangular Ring Ratio-Occurrence (CRRO) Block based Ratio-Occurrence (BRO)	2008 i-LIDS MCTS	Act
(Rabiee, Haddadnia, Mousavi et al., 2016)	C	Histogram of Oriented Gradients (HOG) Histogram of Optical Flow (HOF) Motion Boundary Histogram (MBH)	UCSD CUHK PETS2009 ViF Rodriguez UCF UMN	Act
(Liao et al., 2011)	C	Support Vector Machine (SVM) Library for Support Vector Machines (LIBSVM) Basis Radial Function (BRF) Block Matching Algorithm (BMA)	UMN	Act
(Mehran et al., 2009)	C	Social Force Model (SFM) Pure Optical Flow(POF)	UNM	Act
(Schuld et al., 2004)	G(25)	Support Vector Machines (SVM)	N	Act
(Vishwakarma & Agrawal, 2013)	G	Neural Network(NN)	N	Act
(Azorin-Lopez et al., 2014)	G	Self-Organizing Map (SOM) Supervised Self-Organizing Map (SSOM) Neural GAS (NGAS) Linear Discriminant Analysis (LDA) k-Nearest Neighbour (kNN) Multiclassifier (MC)	CAVIAR	Actv
(Goel & Robicquet, 2015)	G	Convolutional Neural Networks (CNN) Long Short-Term Memory (LSTM)	UAV	Actv
(Harmon et al., 2016)	G	Convolutional Neural Network (CNN) Feed Forward Network (FFN)	SportsVU	Actv
(Lin et al., 2008)	G	Category Feature Vectors (CFVs) Gaussian Mixture Models (GMM) Recognizing algorithm (CFR)	N	Actv

continued on following page

Table 2. Continued

Reference	CI	Technique	Dataset	La
(Liu et al., 2007)	G(50)	Inter-Relation Pattern Matrix (IRPM) Game-Theoretic Conversational Groups (GTCG) Spectral Clustering (R-GTCG SC)	DGPI	Actv
(Maksai et al., 2016)	G	Multiple Object Tracking Accuracy (MOTA) K-Shortest Paths Optimization (KSP) Markov Decision Process (MDP) Recurrent Neural Networks (RNN)	TOWN ETH HOTEL STATIONTOWN	Actv
(Perse et al., 2009)	G	Gaussian Mixture Model (GMM) EM algorithm	N	Actv
(Vascon et al., 2016)	G	Stability Features (HDP)	BEHAVE	Actv
(Yin et al., 2013)	G	Minimum Description Length (MDL)	COLLECTIVE ACTIVITY BEHAVE	Actv
(Ge & Collins, 2009)	C	Bayesian Marked Point Process (MPP)	CAVIAR VSPETS SOCCER	CP
(Kilambi et al., 2008)	G(90)	Heuristic learned (HL)	N	CP
(Kong et al., 2006)	C	Linear Fitting (LF) Unsupervised Neural Network (UNN)	N	CP
(Camplani et al., 2016)	G	Multiple Human Tracking (MHT) Correct Detected Tracks (CDT) False Alarm Tracks (FAT) Track Detection Failure (TDF)	ETH UHD	Tra
(Cupillard et al., 2002)	C	Detection of moving regions	METRO	Tra
(Ge et al., 2012)	G	Hidden Markov Models (HMM) Dynamic Bayes Networks (DBN)	BIWI	Tra
(Ge et al., 2012)	G	Sampling Importance Resampling (SIR) Discrete Choice Model (DCM) Multi Hypothesis Tracking (MHT) Statistical Shape Modeling (SSM)	BIWI	Tra
(Lau et al., 2010)	G(20)	Multi-model MHT	Own	Tra
(Liu et al., 2007)	G(25)	Markov Chain Monte Carlo (MCMC) Gaussian Mixture Model (GMM)	FIFA WC 2006	Tra

4. CONCLUSION AND FUTURE DIRECTIONS

In this work, the human behavior of groups and crowds has been approached taking into account the degree of semantics and especially the size of people that integrate the group or crowd, in addition has been considered behaviors like; Sports teams of soccer and basketball, pedestrians, groups of people in metro and bus stations, people grouped in parks and squares. We propose a classification of behavior of groups and crowds according to degree of semantics has been carried out in three types; Motion, Action, Activity, also two automatic tasks, Count-People and Tracking. It has included techniques and algorithms that researchers use for analysis, and has included the datasets used, which in most of the investigations are traditional and in a few cases custom datasets or YouTube videos are used.

In the next works is important to address the issue of video sequences with RGBD cameras, as this type of technology is currently in increasing use.

REFERENCES

- Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys*, 43(3), 16. doi:10.1145/1922649.1922653
- Al-Raziqi, A., & Denzler, J. (2016). Unsupervised Framework for Interactions Modeling between Multiple Objects. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (pp. 509–516). doi:10.5220/0005680705090516
- Andrade, E. L., Blunsden, S., & Fisher, R. B. (2006, August). Hidden markov models for optical flow analysis in crowds. In *Proceedings of the 18th International Conference on Pattern Recognition ICPR '06* (Vol. 1, pp. 460-463). IEEE. doi:10.1109/ICPR.2006.621
- Andrade, E. L., Blunsden, S., & Fisher, R. B. (2006, August). Modelling crowd scenes for event detection. In *Proceedings of the 18th International Conference on Pattern Recognition ICPR '06* (Vol. 1, pp. 175-178). IEEE. doi:10.1109/ICPR.2006.806
- Azorin-Lopez, J., Saval-Calvo, M., Fuster-Guillo, A., Garcia-Rodriguez, J., Cazorla, M., & Signes-Pont, M. T. (2016, July). Group activity description and recognition based on trajectory analysis and neural networks. In *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 1585-1592). IEEE. doi:10.1109/IJCNN.2016.7727387
- Azorin-Lopez, J., Saval-Calvo, M., Fuster-Guillo, A., Garcia-Rodriguez, J., & Orts-Escolano, S. (2015, July). Self-Organizing Activity Description Map to represent and classify human behaviour. In *Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE. doi:10.1109/IJCNN.2015.7280784
- Azorin-Lopez, J., Saval-Calvo, M., Fuster-Guillo, A., & Oliver-Albert, A. (2014, July). A predictive model for recognizing human behaviour based on trajectory representation. In *Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 1494-1501). IEEE. doi:10.1109/IJCNN.2014.6889883
- Banos, O., Damas, M., Pomares, H., Rojas, F., Delgado-Marquez, B., & Valenzuela, O. (2013). Human activity recognition based on a sensor weighting hierarchical classifier. *Soft Computing*, 17(2), 333–343. doi:10.1007/s00500-012-0896-3
- Bengio, Y., Courville, A. C., & Vincent, P. (2012). Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR*, abs/1206.5538, 1
- Blank, M., Gorelick, L., Shechtman, E., Irani, M., & Basri, R. (2005, October). Actions as space-time shapes. In *Proceedings of the Tenth IEEE International Conference on Computer Vision ICCV '05* (Vol. 2, pp. 1395-1402). IEEE. doi:10.1109/ICCV.2005.28
- Blunsden, S., & Fisher, R. B. (2010). The BEHAVE video dataset: ground truthed video for multi-person behavior classification. *Annals of the BMVA*, 4(1-12), 4.
- Brostow, G. J., & Cipolla, R. (2006, June). Unsupervised bayesian detection of independent motion in crowds. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 1, pp. 594-601). IEEE. doi:10.1109/CVPR.2006.320
- Bruckner, D., Yin, G. Q., & Faltinger, A. (2012). Relieved commissioning and human behavior detection in Ambient Assisted Living Systems. *Elektrotechnik und Informationstechnik*, 129(4), 293-298.
- Camplani, M., Paiement, A., Mirmehdi, M., Damen, D., Hannuna, S., Burghardt, T., & Tao, L. (2016). Multiple human tracking in RGB-D Data: A Survey. arXiv:1606.04450
- Cardinaux, F., Bhowmik, D., Abhayaratne, C., & Hawley, M. S. (2011). Video based technology for ambient assisted living: A review of the literature. *Journal of Ambient Intelligence and Smart Environments*, 3(3), 253–269.
- Charaoui, A. A., Climent-Pérez, P., & Flórez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. *Expert Systems with Applications*, 39(12), 10873–10888. doi:10.1016/j.eswa.2012.03.005
- Chang, M. C., Krahnstoeber, N., Lim, S., & Yu, T. (2010, August). Group level activity recognition in crowded environments across multiple cameras. In *Proceedings of the 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 56-63). IEEE. doi:10.1109/AVSS.2010.65

- Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, 117(6), 633–659. doi:10.1016/j.cviu.2013.01.013
- Choi, W., Shahid, K., & Savarese, S. (2009, September). What are they doing?: Collective activity classification using spatio-temporal relationship among people. In *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 1282-1289). IEEE.
- Climent-Pérez, P., Maudit, A., Monekosso, D. N., & Remagnino, P. (2014, January). Detecting events in crowded scenes using tracklet plots. In *Proceedings of the 2014 International Conference on Computer Vision Theory and Applications (VISAPP)* (Vol. 2, pp. 174-181). IEEE.
- Cupillard, F., Bremond, F., & Thonnat, M. (2002). Tracking groups of people for video surveillance. In *Video-Based Surveillance Systems* (pp. 89-100). Springer US. doi:10.1007/978-1-4615-0913-4_7
- Dehghan, A. (2016). *Global Data Association for Multiple Pedestrian Tracking*.
- Ferryman, J. (2009). PETS 2009 Benchmark Data. Retrieved 19 March 2017 from <http://www.cvg.reading.ac.uk/PETS2009/a.html>
- Fisher, R. (2011). Computer-Assisted Prescreening of Video Streams for Unusual Activities. BEHAVE. Retrieved 19 March 2017 from <http://homepages.inf.ed.ac.uk/rbf/BEHAVE/>
- Fradi, H., & Dugelay, J. L. (2016). Spatial and temporal variations of feature tracks for crowd behavior analysis. *Journal on Multimodal User Interfaces*, 10(4), 307–317. doi:10.1007/s12193-015-0179-2
- Ge, W., & Collins, R. T. (2009, June). Marked point processes for crowd counting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR '09* (pp. 2913-2920). IEEE. doi:10.1109/CVPR.2009.5206621
- Ge, W., Collins, R. T., & Ruback, B. (2009, December). Automatically detecting the small group structure of a crowd. In *Proceedings of the 2009 workshop on Applications of computer vision (wacv)* (pp. 1–8). IEEE. doi:10.1109/WACV.2009.5403123
- Ge, W., Collins, R. T., & Ruback, R. B. (2012). Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5), 1003–1016. doi:10.1109/TPAMI.2011.176 PMID:21844622
- Gning, A., Mihaylova, L., Maskell, S., Pang, S. K., & Godsill, S. (2011). Group object structure and state estimation with evolving networks and Monte Carlo methods. *IEEE Transactions on Signal Processing*, 59(4), 1383–1396. doi:10.1109/TSP.2010.2103062
- Goel, K., & Robicquet, A. (n. d.). *Learning Causalities behind Human Trajectories*.
- Gong, S., Cristani, M., Yan, S., & Loy, C. C. (2014). *Person re-identification* (Vol. 1). London: Springer. doi:10.1007/978-1-4471-6296-4
- Harmon, M., Lucey, P., & Klabjan, D. (2016). Predicting Shot Making in Basketball using Convolutional Neural Networks Learnt from Adversarial Multiagent Trajectories. arXiv:1609.04849
- Hu, Y., Zhang, Y., & Davis, L. (2013). Unsupervised abnormal crowd activity detection using semiparametric scan statistic. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 767-774). doi:10.1109/CVPRW.2013.115
- Jacques, J. C. S. Jr, Braun, A., Soldera, J., Musse, S. R., & Jung, C. R. (2007). Understanding people motion in video sequences using Voronoi diagrams. *Pattern Analysis & Applications*, 10(4), 321–332. doi:10.1007/s10044-007-0070-1
- Jhuang, H., Gall, J., Zuffi, S., Schmid, C., & Black, M. J. (2013). Towards understanding action recognition. In *Proceedings of the IEEE international conference on computer vision* (pp. 3192-3199).
- Jiang, Y. G., Liu, J., Zamir, A. R., Laptev, I., Piccardi, M., Shah, M., & Sukthankar, R. (2013). THUMOS: The first international workshop on action recognition with a large number of classes.

- Kilambi, P., Ribnick, E., Joshi, A. J., Masoud, O., & Papanikolopoulos, N. (2008). Estimating pedestrian counts in groups. *Computer Vision and Image Understanding*, 110(1), 43–59. doi:10.1016/j.cviu.2007.02.003
- Kliper-Gross, O., Hassner, T., & Wolf, L. (2012). The action similarity labeling challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3), 615–621. doi:10.1109/TPAMI.2011.209 PMID:22262724
- Kong, D., Gray, D., & Tao, H. (2006, August). A viewpoint invariant approach for crowd counting. In *Proceedings of the 18th International Conference on Pattern Recognition ICPR '06* (Vol. 3, pp. 1187-1190). IEEE. doi:10.1109/ICPR.2006.197
- Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008, June). Learning realistic human actions from movies. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR '08* (pp. 1-8). IEEE. doi:10.1109/CVPR.2008.4587756
- Lau, B., Arras, K. O., & Burgard, W. (2010). Multi-model hypothesis group tracking and group size estimation. *International Journal of Social Robotics*, 2(1), 19–30. doi:10.1007/s12369-009-0036-0
- Li, W., Mahadevan, V., & Vasconcelos, N. (2014). Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1), 18–32. doi:10.1109/TPAMI.2013.111 PMID:24231863
- Liao, H., Xiang, J., Sun, W., Feng, Q., & Dai, J. (2011, August). An abnormal event recognition in crowd scene. In *Proceedings of the 2011 Sixth International Conference on Image and Graphics (ICIG)* (pp. 731-736). IEEE. doi:10.1109/ICIG.2011.66
- Lijun, C., & Kaiqi, H. (2013). Video-based crowd density estimation and prediction system for wide-area surveillance. *China Communications*, 10(5), 79–88. doi:10.1109/CC.2013.6520940
- Lin, W., Sun, M. T., Poovandran, R., & Zhang, Z. (2008, May). Human activity recognition for video surveillance. In *Proceedings of the IEEE International Symposium on Circuits and Systems ISCAS '08* (pp. 2737-2740). IEEE.
- Liu, J., Tong, X., Li, W., Wang, T., Zhang, Y., & Wang, H. (2009). Automatic player detection, labeling and tracking in broadcast soccer video. *Pattern Recognition Letters*, 30(2), 103–113. doi:10.1016/j.patrec.2008.02.011
- Liu, J., Yang, Y., Saleemi, I., & Shah, M. (2012). Learning semantic features for action recognition via diffusion maps. *Computer Vision and Image Understanding*, 116(3), 361–377. doi:10.1016/j.cviu.2011.08.010
- Maksai, A., Wang, X., Fleuret, F., & Fua, P. (2016). Globally Consistent Multi-People Tracking using Motion Patterns. arXiv:1612.00604
- Mehran, R., Oyama, A., & Shah, M. (2009, June). Abnormal crowd behavior detection using social force model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR '09* (pp. 935-942). IEEE. doi:10.1109/CVPR.2009.5206641
- Mihaylova, L., Carmi, A. Y., Septier, F., Gning, A., Pang, S. K., & Godsill, S. (2014). Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking. *Digital Signal Processing*, 25, 1–16. doi:10.1016/j.dsp.2013.11.006
- Oskouie, P., Alipour, S., & Eftekhari-Moghadam, A. M. (2014). Multimodal feature extraction and fusion for semantic mining of soccer video: A survey. *Artificial Intelligence Review*, 42(2), 173–210. doi:10.1007/s10462-012-9332-4
- Pellegrini, S. (2009). BIWI Walking Pedestrians Dataset. Retrieved 20 March 2017 from <http://www.vision.ee.ethz.ch/datasets/>
- Pellegrini, S., Ess, A., Tanaskovic, M., & Van Gool, L. (2010, June). Wrong turn-no dead end: a stochastic pedestrian motion model. In *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 15-22). IEEE. doi:10.1109/CVPRW.2010.5543166
- Perše, M., Kristan, M., Kovačič, S., Vučkovič, G., & Perš, J. (2009). A trajectory-based analysis of coordinated team activity in a basketball game. *Computer Vision and Image Understanding*, 113(5), 612–621. doi:10.1016/j.cviu.2008.03.001
- Rabiee, H., Haddadnia, J., Mousavi, H., Nabi, M., Murino, V., & Sebe, N. (2016). Emotion-Based Crowd Representation for Abnormality Detection. arXiv:1607.07646

- Rodriguez, M. D., Ahmed, J., & Shah, M. (2008, June). Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR '08* (pp. 1-8). IEEE. doi:10.1109/CVPR.2008.4587727
- Ryoo, M. S., & Aggarwal, J. K. (2008, January). Recognition of high-level group activities based on activities of individual members. In *Proceedings of the IEEE Workshop on Motion and video Computing WMVC '08* (pp. 1-8). IEEE. doi:10.1109/WMVC.2008.4544065
- Schuldt, C., Laptev, I., & Caputo, B. (2004, August). Recognizing human actions: A local SVM approach. In *Proceedings of the 17th International Conference on Pattern Recognition ICPR '04* (Vol. 3, pp. 32-36). IEEE. doi:10.1109/ICPR.2004.1334462
- Shah, M. (2012). PNNL Parking Lot. Retrieved 31 March 2017 from <http://csrcv.ucf.edu/data/ParkingLOT/>
- Shao, J., Change Loy, C., & Wang, X. (2014). Scene-independent group profiling in crowd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2219-2226).
- Shen, C., Xie, R., Zhang, L., & Song, L. (2015, June). Small group people behavior analysis based on temporal recursive trajectory identification. In *Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.
- Shu, G., Dehghan, A., & Shah, M. (2013). Improving an Object Detector and Extracting Regions Using Superpixels. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3721-3727). doi:10.1109/CVPR.2013.477
- Solmaz, B., Assari, S. M., & Shah, M. (2013). Classifying web videos using a global video descriptor. *Machine Vision and Applications*, 24(7), 1473-1485. doi:10.1007/s00138-012-0449-x
- Soomro, K., & Zamir, A. R. (2008). Sports Action Data Set. UCF. Retrieved 1 January 2017 from http://csrcv.ucf.edu/data/UCF_Sports_Action.php
- Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101 - Action Recognition Data Set. Retrieved 31 March 2017 from http://csrcv.ucf.edu/data/UCF101.php#Results_on_UCF101
- Turaga, P., Chellappa, R., Subrahmanian, V. S., & Udreă, O. (2008). Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11), 1473-1488. doi:10.1109/TCSVT.2008.2005594
- Vascon, S., Mequanint, E. Z., Cristani, M., Hung, H., Pelillo, M., & Murino, V. (2016). Detecting conversational groups in images and sequences: A robust game-theoretic approach. *Computer Vision and Image Understanding*, 143, 11-24. doi:10.1016/j.cviu.2015.09.012
- Vishwakarma, S., & Agrawal, A. (2013). A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, 29(10), 983-1009. doi:10.1007/s00371-012-0752-6
- Wickramaratna, K., Chen, M., Chen, S. C., & Shyu, M. L. (2005, December). Neural network based framework for goal event detection in soccer videos. In *Proceedings of the Seventh IEEE International Symposium on Multimedia*. IEEE. doi:10.1109/ISM.2005.83
- Wu, Y., Jia, Z., Ming, Y., Sun, J., & Cao, L. (2016). Human behavior recognition based on 3D features and hidden Markov models. *Signal. Image and Video Processing*, 10(3), 495-502. doi:10.1007/s11760-015-0756-6
- Yi, S., Li, H., & Wang, X. (2015). Pedestrian travel time estimation in crowded scenes. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3137-3145). doi:10.1109/ICCV.2015.359
- Yi, S., Li, H., & Wang, X. (2015). Understanding pedestrian behaviors from stationary crowd groups. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3488-3496). doi:10.1109/CVPR.2015.7298971
- Yin, Y., Yang, G., & Man, H. (2013, September). Small human group detection and event representation based on cognitive semantics. In *Proceedings of the 2013 IEEE Seventh International Conference on Semantic Computing (ICSC)* (pp. 64-69). IEEE. doi:10.1109/ICSC.2013.20
- Zhang, C., Yang, X., Lin, W., & Zhu, J. (2012, December). Recognizing human group behaviors with multi-group causalities. In *Proceedings of the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (Vol. 3, pp. 44-48). IEEE. doi:10.1109/WI-IAT.2012.162

Jorge Azorin-Lopez received a degree in Computer Engineering in 2001 and a PhD degree in Computer Science at the University of Alicante (Spain) in 2007. Since 2001, he has been a faculty member of the Department of Computer Science Technology and Computation at the same university, where he is currently an Associate Professor and the Deputy Director of Research. He was awarded the Post-Doctoral Research Fellowship "Automatic visual inspection of shape defects on specular surfaces. Methods to compensate low sensitivity of 3D image acquisition and reconstruction techniques" by the Spanish Ministry of Science and Education for research at University of Edinburgh. He has worked in 14 research projects and has published more than 40 papers on computer vision and computer architecture in several journals, conferences and book chapters. He has served as a reviewer to numerous scientific journals and international conferences.

Marcelo Saval-Calvo obtained a PhD in Computer Science in 2015, as well as Computer engineer at University of Alicante in 2010 and Master of Science in 2011. His interests include computer vision problem in human behaviour understanding, 3D shapes changes analysis and modelling, and 3D sensing and mapping. Moreover, he has expertise in CUDA programming for GPGPUs. Marcelo has published various paper in these areas and has attended international conferences and specialist courses. He worked as a senior researcher in a National Funded project in Advanced Driver Assistance Systems accelerating computer vision algorithms using heterogeneous platforms. He is a postdoc visitor in the School of Informatics, University of Edinburgh, in the computer vision group led by Prof. Robert B. Fisher.