A Survey of Vision-Based Transfer Learning in Human Activity Recognition

David Ada Adama *, **Ahmad Lotfi** and **Robert Ranson**

Department of Computer Science, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK; ahmad.lotfi@ntu.ac.uk (A.L.); robert.ranson@ntu.ac.uk (R.R.)
* Correspondence: david.adama@ntu.ac.uk; Tel.: +44-(0)-115-848-3339

Abstract: Human activity recognition (HAR) and transfer learning (TL) are two broad areas widely studied in computational intelligence (CI) and artificial intelligence (AI) applications. Much effort has been put into developing suitable solutions to advance the current performance of existing systems. However, challenges are facing the existing methods of HAR. In HAR, the variations in data required in HAR systems pose challenges to many existing solutions. The type of sensory information used could play an important role in overcoming some of these challenges. Vision-based information in 3D acquired using RGB-D cameras is one type. Furthermore, with the successes encountered in TL, HAR stands to benefit from TL to address challenges to existing methods. Therefore, it is important to review the current state-of-the-art related to both areas. This paper presents a comprehensive survey of vision-based HAR using different methods with a focus on the incorporation of TL in HAR methods. It also discusses the limitations, challenges and possible future directions for more research.

Keywords: human activity recognition; transfer learning; vision; ambient assisted living

1. Introduction

Understanding the process of learning in humans has been an area of interest for decades. This has attracted interest from different areas of study which use different approaches, such as computational intelligence (CI), biology, and psychology, amongst many others. One key aspect of the learning process that has been challenging to researchers in the artificial intelligence (AI) community is designing systems which leverage knowledge gained from solving a task into improved performance when solving similar or dissimilar problems. This is where the concept of transfer learning (TL) comes in. The importance of TL cannot be overemphasised; time spent learning new tasks is reduced, more situations can be handled effectively and the information required by human experts is also reduced.

Human activities are diverse in nature. There is imprecision, vagueness, ambiguity and uncertainty in information about the way activities are performed. Thus, variabilities are encountered when developing models for activity classification/recognition—for example, in human–robot or robot-robot interaction requiring a robot to recognise and learn activities from another robot or human as illustrated in Figure 1. This can affect the correct classification of human activities, which is relevant in improving the amount of knowledge that can be used by a robot in learning. Considering human activity information obtained using visual sensors, the quality of information obtained of activities is a vital component of any recognition system. Early approaches to vision-based HAR utilised 2D information obtained using sensors such as RGB cameras [1]. However, it would be extremely hard to understand and interpret activities using such sensors which provide 2D visual data. These sensors provide limited information for an activity performed in a real-world environment. However, recent developments in RGB-depth (RGB-D) sensors showed that they are better devices for observing human activities [1]. These sensors provide a means of better observing the world to detect human poses, to build activity recognition systems [2,3]. They also

provide a platform for exploiting depth maps, body shape and skeleton joint detection of humans in 3D space, which are used for developing sophisticated recognition algorithms.
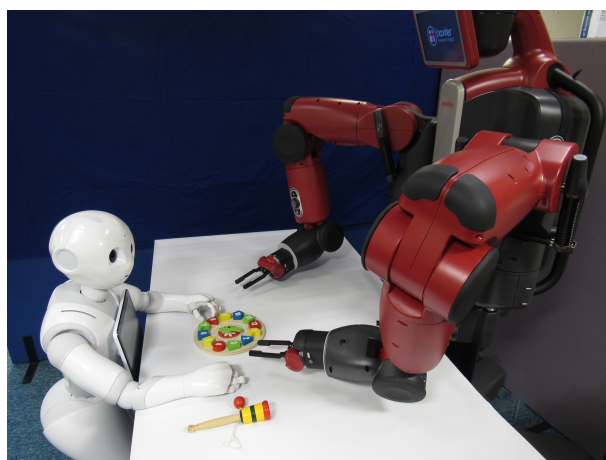

Figure 1. An illustration of transfer learning of an activity from one robot to another robot.

In the case of learning human activities, classifying them correctly from some given set of data is key to understanding how these activities are learnt and how one experience relates to another. The use of machine learning (ML) approaches to tackle these constraints is limited due to the necessity of the training and test data to come from the same feature space and data distribution. This limits its ability in situations where there are differences in data distribution between the training and test data, which can result in the predictive learner being degraded [4]. Obtaining training data to match the feature space and predicted data distribution of test data is often times expensive and difficult [5].

Many CI techniques have been applied to TL, amongst which deep learning approach has been widely researched [6]. This is one of the most popular CI techniques that has been significantly applied to the domain of TL [7–10]. However, it requires a large amount of training data, and it also works as a *black box* (learning only the relationship between input and output without providing knowledge of the relationship that is key in making decisions) due to its computational framework.

From recent surveys [5,6], the key challenge in TL has been defining the evaluation metrics related to *what to transfer*, *how to transfer* and *when to transfer*. This is mainly because there are various possible measurement options and/or algorithms. The algorithms used so far focus on three main steps: First, given a target task, select an appropriate source task or sets of tasks from which to transfer. Second, learn the relationship between the target task and source task(s). Third, transfer knowledge effectively from source task(s) to target task. The work in [11] focused on learning inter-task relations which were modelled using a three-way restricted Boltzmann machine (RBM). This model captures the similarity between samples from the source task and target task. The method, however, is computationally complex, since it requires a large amount of training data and it also does not capture the uncertainties associated with the task constraint. In [12], a TL technique was employed to speed up learning robot models using local procrustes analysis, but this method requires correspondence between datasets to be provided and requires a large amount of training data.

In the remaining sections, vision-based HAR methods, transfer learning of human activities and the challenges in these areas are discussed. Specifically, Section 2 presents HAR using 3D visual information and the limitations of the current methods applied in HAR. Section 3 presents the ontology of TL of human activities and CI techniques applied in TL. Section 4 identifies the gaps in existing research around TL in HAR. In Section 5, the challenges and future directions for research in this area are discussed. Finally, in Section 6 conclusions to the review are discussed.

2. Human Activity Recognition with 3D Vision Sensors

The learning and classification of human activities using some CI techniques is often referred to as HAR [13,14]. Over the last few decades, the study of HAR has been carried out to detect, recognise and/or classify activities of humans. HAR has seen many applications in several domains, such as security, health care, manufacturing and gaming, amongst many others. Owing to this, several approaches have been investigated. An integral component of HAR is how information of activities are obtained or observed. Based on the published literature, HAR approaches are divided into two main categories: visual sensor-based and non-visual sensor-based HAR. Observing activities through the use of visual [1–3,15] or non-visual sensors [16] makes it a lot easier to obtain information of human activities in an environment. Non-visual sensor-based approaches utilise information such as environmental conditions—e.g., temperature, motion detection or ambient light; location; and information from wearable devices. A comprehensive review of HAR using non-visual sensors can be found in [17] and more recently in [18]. Although this information has some advantages, they are sometimes invasive and burdensome. On the other hand, HAR using visual sensory information mainly relies on the interpretation of images to predict activities [1,19,20].

One of the main objectives of HAR is extracting descriptive information (i.e., features) from human activities to be able to distinctly characterise and classify one activity from another. Visual sensor-based approaches are mainly based on 2D or 3D information obtained from the sensor devices. However, it would be extremely difficult to understand and interpret activities using regular visual sensors such as RGB cameras which provide 2D visual information [1]. These sensors provide limited information for an activity performed in a real-world environment. Recently, most research on HAR based on visual sensors has employed RGB-D sensors, which have proven to be better devices for observing human activities [1,2,21,22]. These RGB-D sensors provide a means of better observing the world to detect human poses used to build HAR systems [3]. They provide a platform for exploiting depth maps and body shape, and detecting skeletal joints of humans in 3D space which are used in developing sophisticated recognition algorithms. Furthermore, among the many approaches to human representation based on 3D information [1,23–25], the majority of the existing methods can be generally grouped into local feature-based representations [26] and skeleton-based representations [3,27,28].

Figure 2 summarises the categorisation of HAR based on the grouping of activity information employed. Representations based on local features identify relevant points in space-time dimensions, interpret patches at the points as features and encode them into representations which can locate notable regions. However, local feature-based representation methods do not take into consideration the spatial relationships between features. As a result, they are unable to represent multiple humans in the same scene. The local features-based methods can also be computationally expensive due to the complexity involved in the extraction process. However, skeleton-based representations have shown promising performance in real-world applications, including gaming and assisted living [1]. These methods consider the spatial relationships among features, which enables the modelling of human joints' relationships for encoding a whole body's structure. Furthermore, skeleton-based representations are robust to variations in illumination, scale, view and motion speed. Due to these advantages, such representations are used in real-time applications, and many researchers [2,23] have introduced techniques to facilitate different applications.
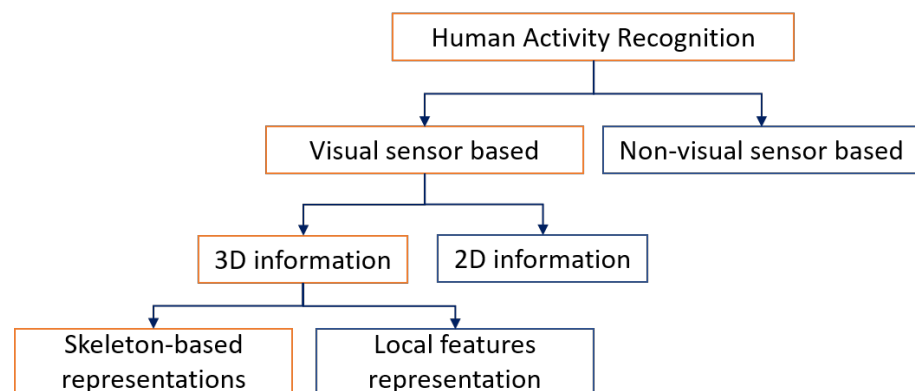
Figure 2. Classification of approaches widely used in human activity recognition based on the source of information.

2.1. Background and Challenge of 3D Vision-Based HAR

Over the past few decades, research on HAR has seen much improvement, with technological advances in the field leading to the availability of low cost, small and low power consumption sensors. Sensory devices used to obtain human activity information have become less intrusive, as they can be incorporated into an AAL environment without being noticed. The sensor networks have not been left out of the advancements as well. Wireless technologies [29] used in sensor networks have enabled unobtrusive recognition of activities with information accessible from any location. The benefits of these advancements cannot be over-emphasised: remote monitoring, individual profiling, intrusion detection, abnormality detection and so much more.

In the field of computer vision, HAR with vision-based methods is one of the most studied areas. The goal is usually to automatically detect and analyse human activities from a sequence of images captured using camera sensors or other vision sensing modalities. These activities take on different forms, which range from elementary actions to complex activities, depending on the environment. Ref. [23] categorised such activities into four groups: atomic actions, activities containing sequences of distinct actions, activities including person-object and person-person interactions, and lastly, group activities. The most difficult of all the categories mentioned are group activities. Research in this area has encountered several limitations which could be the results of the difficulties in collecting the data required or the limitations of existing vision-based sensors.

Here, the challenges of vision-based HAR systems are discussed. From the review of past studies on vision-based HAR, four main challenges were identified. First, the low-level challenges encountered from occlusions, shadows, varying illuminations and cluttered backgrounds [20,30]. These types of challenges are encountered in most cases when using visual sensors. They create difficulties in motion segmentation which alter the forms in which actions are observed. Ref. [31] proposed a technique used in filtering background clutter, occlusions and unstable camera motions for recognising human activities. The technique uses a combination of a multiple-instance formulation and a Markov model in a framework to select elementary actions for encoding movements of local parts. This technique allowed for long-range temporal information of actions in video sequences to be encoded. Ref. [32] also attempted to address the challenge of identifying human actions using conditional random fields (CRFs) to differentiate between unknown movements and intentional actions which may occur in a scene through the ordering of video regions and identifying the actors for actions. Furthermore, 3D sensor information [23] has been introduced as a solution to mitigate the low-level difficulties due to their ability to provide structure information from a scene.

The second challenge has to do with changes because of an activity [23,24,33,34]. Two sets of information of the same human action can generate different representations depending on the perspective from which such information is obtained. This poses a challenge

when using stand-alone cameras in acquiring activity information. To tackle this challenge with a single camera is an extremely challenging task. Solutions proposed to address this challenge have adopted multiple synchronised cameras, although implementing such cameras in applications can be a daunting task. One of such solutions is the introduction of 3D motion capture systems (MoCap) [23] which have enabled recognition algorithms to alleviate this challenge. The use of depth information from such MoCap systems to obtain skeletal joint information of a human can be used in constructing view-invariant information for algorithms used in HAR [35].

The third challenge identified with vision-based HAR is scale variance [23,33], which occurs when a subject or different subjects appear to be different sizes when viewed from differing distances to the camera. A solution to this when using 2D information is extracting features at multiple scales. Furthermore, using 3D information solves this challenge since the depth information of a subject is easily known and can be adjusted through the activity sequence.

Finally, there is the challenge of inter-class similarity and intra-class variability of actions [36]. This occurs as a result of the uncertainties in the way actions are performed by humans. A single action can be carried out by individuals in different directions with varying characteristics of body movements, and similarly, two actions may only be differentiated by subtle spatio-temporal information [23]. This poses a challenge for real-world applications of vision-based HAR, and to date, it remains a difficult problem for recognition algorithms using the different modalities of visual data.

To achieve recognition of human activities, three main steps are involved. These steps correspond to data input, processing and classification. The data input step is the acquisition of human activity data with the means of a sensory device and the data are then processed, which entails stages of feature extraction, feature reduction, standardisation, etc. The processing step prepares the data for fitting in the model which will be used in identifying activities. Different methods proposed by researchers that have been applied in human activity data acquisition, processing and recognition and classification to build HAR systems are discussed in the following sections. In Table 1, a general taxonomy of vision-based HAR based on the categorisation in Figure 2 is provided. The main features of both the 2D and 3D vision-based approaches are highlighted; examples of the sensors used and their main advantages and disadvantages are given.

Table 1. Taxonomy of vision-based HAR based on the grouping of information used.

Grouping	Summary	Benefits	Short-Comings	Example Sensors
2D information	Infer human activities from 2D points extracted from images.	Processing does not require as much computational resources as 3D information	- Information obtained is limited. - Not robust to variations in the scale of subjects.	RGB cameras. E.g. Webcams.
3D information	Identifies human activities from point clouds of changes in human movement or RF signals.	- Overcomes the scale variance problem. - Provides more information on human activities. - Are robust to view changes in activities.	- Usually require more computational resources. - MoCap systems require the installation of multiple sensors.	Motion Capture systems, RF devices [37], RGB-D sensors. E.g. Microsoft Kinect [38].

2.2. Data Collection of Human Activities in 3D Skeletal Data Space

Data obtained from RGB-D sensors give information relevant for a robot to understand the activity. By exploring human pose detection using RGB-D sensors, activity recognition has advanced recently [2,15]. RGB-D sensors extract 3D skeleton data from depth images and body silhouettes for feature generation. In [2], a RGB-D sensor was used to generate a human 3D skeleton model with matching of body parts linked by its joints. They extracted positions of individual joints from the skeleton in a 3D form *x, y, z*. Ref. [14] used a similar RGB-D sensor to obtain depth silhouettes of human activities from which body point information was extracted for the activity recognition system. Ref. [39] also used an RGB-D sensor to capture human skeleton information as part of a system for controlling a mobile robot using human gestures, which is also similar to an application proposed by [40]. Another approach was shown in [41], where the RGB-D sensor was used to obtain an orientation-based human representation of each joint relative to the human centroid in 3D space. Furthermore, radio frequency (RF) devices have been used for acquiring 3D human information. Ref. [37] used a radio device to obtain RF signals used in estimating 3D poses which is robust to occlusions in the environment. These researchers [39–41] used different devices for the acquisition of data. In the following section, methods of acquisition of 3D human skeletal data are discussed.

2.2.1. Direct Acquisition of 3D Human Skeletal Data from Sensors

Direct methods of acquisition of 3D skeletal data of human activities are carried out using different devices that are commercially available, which include MoCap systems [23], structured-light cameras and time-of-flight sensors. These devices detect the kinematics of human body models in order to identify the relevant joints in the body. Figure 3 shows a representation of tracked skeletal joints obtained from a Microsoft Kinect v2 RGB-D sensor [38].
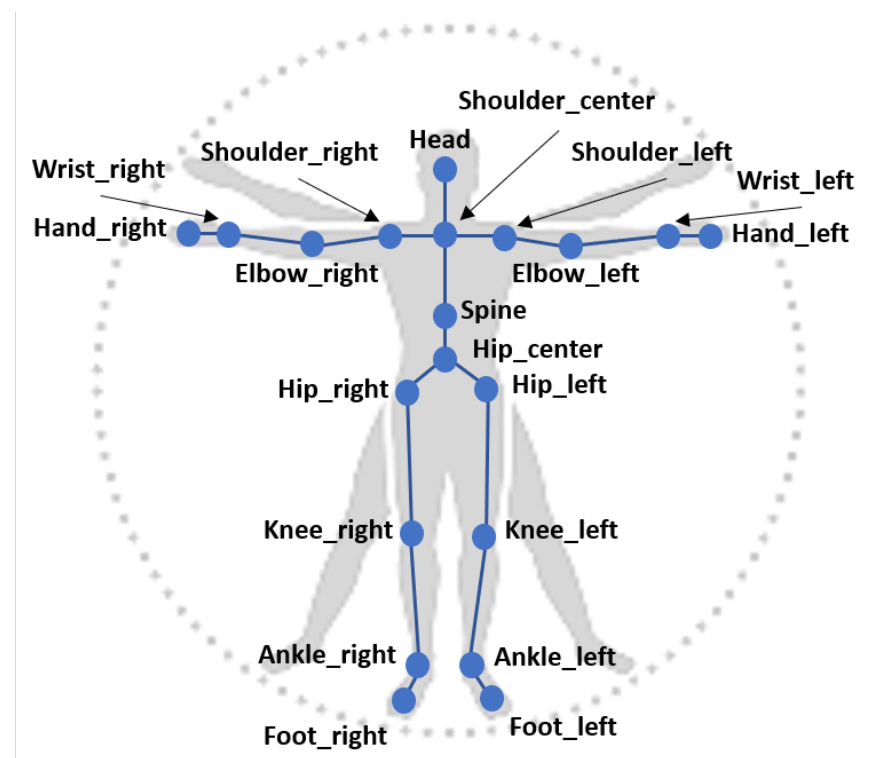

Figure 3. An example of a skeletal body model obtained from a Microsoft Kinect device. This shows the 20 tracked joints.

MoCap systems obtain 3D skeletal information by tracking markers placed on a human in its scene [1]. These systems either utilise multiple cameras at different positions

around a subject to track reflective markers that are attached to a subject's body, or 3-axis inertial sensors that estimate body part rotations with reference to a fixed point. It should be noted that the inertial sensor-based MoCap systems can obtain the skeletal data without any visual cameras. The existing MoCap systems have the software to enable the collection of the 3D skeletal data with a high degree of accuracy. However, most of the systems can only be used in controlled environments and are typically expensive.

Structured-light cameras, which are types of camera devices that utilise infrared light to capture depth information, are also used in the direct acquisition of 3D skeleton data [38]. Light is projected through the infrared sensor in a known pattern and the distortion observed in the pattern when it meets a subject allows the device to decide the depth. The RGB image of the scene observed can also be acquired. Most of the RGB-D sensors are inexpensive, which makes them available for use in most applications. This source has been popularly used in research for HAR [2,22,42,43].

Time-of-flight sensors [1] acquire 3D information by emitting light and measuring the time it takes for the light to be returned. Some examples of such sensing technologies are radar and light detection and ranging (LiDAR) [44]. These sensors acquire 3D information accurately at high frame rates. Among all three methods of direct acquisition of 3D skeletal information, the RGB-D sensors are the most affordable and can be installed in any environment. Furthermore, they provide additional RGB data which can be accessed and processed with depth information.

2.2.2. 3D Skeleton Construction from Pose Estimation

3D skeletal information can also be acquired through human pose estimation and construction of a skeleton [3,14,20,45]. A number of approaches have been proposed to estimate human joints and pose recognition from the knowledge of available data. Such approaches take advantage of depth images or extra information accessible from the visual sensing device. The majority of the methods are based on the identification of body parts which are fitted to models which extract specific locations of the identified parts. This section provides a review of such methods of human skeleton construction based on visual data.

The first approach considered is the construction of 3D human activity information from depth images. The human skeleton can be constructed from a single observed depth image or from acquired sequences of depth images. This approach is widely used in the acquisition of activity information due to the additional geometric information depth images provide. Ref. [14] introduced a vision-based life logging system using depth images to track human body points and location. Their method identifies 15 joints from a depth silhouette, and an additional eight centre points of limbs joints are constructed using the Gaussian contours mechanism. The work was further extended in [46] using temporal depth motion identification to obtain depth human silhouettes from other objects within the scene. Recently, in [47] another model for human body part estimation and detection was proposed using depth imagery. A colour space transformation based on the heuristic thresholding segmentation technique [48] is used to obtain salient regions, and then skin tone detection through foreground segmentation of silhouettes. Afterwards, the body parts are estimated using a proposed body parts model through pixel-wise searching and computation of the distance from the top to the bottom of the silhouette. A novel approach for pose estimation from a single depth image called model-based recursive matching (MRM) was introduced in [49]. This approach combines a depth image and a 3D point cloud of the input to create a human skeleton model with customised parameters based on T-pose to fit different body types. The results reported in [49] show that the proposed method is able to give accurate estimations in cases where there are occlusions obstructing pose recognition. The method uses a MoCap for depth image acquisition which is able to handle occlusions better than a single RGB-D sensor device. The downside to the use of depth images for pose estimation is that most of the systems are computationally complex to set up.

Another approach widely employed in human skeleton construction is using traditional RGB images. Typically, most of the methods using RGB images extract visual features using deep learning (DL) architectures and other methods to match poses of segmented silhouettes for identifying body parts. Deep neural networks (DNN) have demonstrated their ability in construction of human skeletons from RGB images [35,50,51]. Ref. [51] applied DNNs in an approach to estimate human poses called "DeepPose." They formulated the pose estimation as a regression problem by proposing a cascade of DNN regressors for high precision estimates. Ref. [50] adopted a dual-source deep convolutional neural network (DS-CNN) approach for both joint detection and localisation from a single RGB image. The approach takes image patches as inputs and learns the appearance of each body part by considering each in the context of the full body. Refs. [37,52] in a similar approach proposed a novel CNN architecture that performs high-dimensional convolutions on RF signals obtained using radio devices by decomposing them into low-dimensional operations which are used for estimating 3D human poses.

Apart from DNNs, other methods have also been used for human body part estimation from RGB images. For example, [53] proposed an algorithm for estimating sequences of upper-body parts in unconstrained videos. They used a two-step approach in which a spatial model was constructed to capture relationships between adjacent parts, and then a method to select the best out of different pose configurations. Furthermore, a general parametrisation of a body pose method to estimate 3D human poses from 2D joint locations was shown in [54]. The method uses priors that are learned from joint limits in poses. Multiple images acquired using multiple cameras from different angles can be used when observing humans, and then image processing techniques can be employed for estimating human depth maps from the combined images. After obtaining depth maps, human skeleton models can be composed using some of the methods already described. Although existing solutions use depth maps constructed from multiple images to construct human skeletons, such solutions are usually slow and encounter problems related to noisy depth data and correspondence search failures.

2.2.3. Benchmark 3D Skeleton Human Activity Datasets

There are several datasets that have been generated which contain 3D skeletal human activity information. In the review presented by [1], benchmark datasets between 2007 and 2016 are presented. Therefore, this section presents a review of the benchmark 3D skeletal human activity datasets made available to the public in the last five years. A comprehensive list of these datasets is given in Table 2, along with information of the types of devices used and modality of the data.

Table 2. Benchmark public datasets of human activity containing 3D skeleton human information.

Year	Dataset	Modality	Acquisition Device
2021	UAV-Human [55]	RGB + Depth + 3D joints	Azure DK
2020	Bimanual Actions Dataset [56]	RGB + Depth	PrimeSense
2020	NTU RGB+D 120 [57]	RGB + Depth + 3D joints	Kinect v2
2019	AMASS [58]	Depth + 3D Mesh	MoCap
2018	3DPW [59]	RGB	Camera + IMU
2018	MuPoTs-3D [60]	RGB + Depth	MoCap
2017	TotalCapture [61]	RGB + 3D joints	MoCap + IMU
2017	PKU-MMD [62]	RGB + Depth + 3D joints	Kinect v2
2017	CLAD [63]	RGB + Depth + Point cloud	Kinect v1
2016	NTU RGB+D [64]	RGB + Depth	Kinect v2

Recent 3D skeleton datasets have attempted to obtain information for complex HAR settings/tasks, such as outdoor settings and concurrent activities. Most recent is the UAV-Human dataset [55], which is a large dataset containing information of human behaviour,

and was collected by unmanned aerial vehicles (UAV). The dataset contains 155 different activity categories for action recognition collected using three sensors, an Azure DK, a fisheye camera and a night-vision camera. It is important to note that this dataset was collected using a UAV in urban and rural districts for challenging HAR problems in outdoor environments. Bimanual actions dataset [56] is another dataset collected for analysis of bimanual human actions. The dataset comprises 540 videos of six subjects performing nine tasks and was collected using a PrimSense RGB-D camera at 30 *fps*.

MoCap systems are also used extensively for collecting 3D human activity information. The AMASS dataset [58] is one of the largest recently generated datasets using MoCap systems. Fifteen different optical marker-based MoCap datasets are unified together into realistic 3D meshes. The dataset contains more than 40 hours of motion data of over 300 subjects performing more than $11,000$ actions. MoCap systems were also combined with other sensors, such as inertial measurement units (IMUs) to accurately estimate human 3D poses. The MuPoTs-D3 [60] dataset is another dataset that made use of a marker-less MoCap system to obtain over 8000 frames of 3D poses of eight subjects performing a variety of activities in both indoor and outdoor settings. The 3DPW TotalCapture [61] dataset combined MoCaps and IMUs. The dataset consists of five subjects performing several activities. The MoCap system comprised eight calibrated static cameras, and 13 IMUs were attached to different body parts to obtain the ground-truth poses.

Other modern datasets containing 3D skeletal human activities used RGB-D sensors to acquire information. The NTU RGB+D 120 [57], PKU-MMD [62], CLAD [63] and NTU RGB+D [64] datasets made use of the Kinect sensor to obtain 3D information. These datasets contain information of subjects performing activities in different settings, gathering information on 3D skeleton joint coordinates, RGB and depth. Besides the datasets collected using MoCaps and RGB-D sensors, some datasets have been collected using standalone RGB cameras. The 3DPW [59] dataset is one such dataset collected using a single camera combined with IMUs attached to the subjects' limbs for estimating 3D human poses. Seven subjects with 17 IMUs attached per subject were used to collect over $51,000$ frames of different activities.

2.3. Feature Extraction in HAR from 3D Skeletal Human Activities Data

Feature extraction is a vital component of any HAR system. The goal of feature extraction is to find recognisable characteristics of human activity data that can be used to accurately differentiate between activities. Due to the importance of the process of feature extraction and the role features play in an HAR system, the performance of any HAR system is largely attributed to the quality of features obtained from the available data.

Following the acquisition of human activity data using one of the methods reviewed in Sections 2.2.1 and 2.2.2, the raw data obtained from these sensors have to be preprocessed prior to feature extraction. This process is carried out to reduce redundancy in data to better represent the features of an activity. Most of the works [22,65] employing 3D joint coordinates data of skeleton used a preprocessing step to offset the data centroids (usually obtained with reference to the sensor origin) to the human centroid as the origin. This makes the data scale-invariant and makes it easier for recognition algorithms to attain good performances.

According to [66], approaches to HAR using RGB-D information fall into two categories: feature-based and model-based. Feature-based techniques such as histogram of oriented gradients (HOG) and the subspace clustering-based approach (SCAR) are used to extract features for recognising human activity from data acquired using sensors. Ref. [67] applied statistical covariance of 3D joints (Cov3DJ) as features, as shown in Figure 4, to encode the skeleton data of joint positions, which were then used as input to an SVM model for activity recognition. Another approach applied by [68] used a sequence of joint trajectories and applied wavelets to encode each temporal sequence of joints into features used in activity classification. Model-based techniques have to do with the construction of a human model for recognition either as a 2D, 3D or skeletal model. Ref. [69] constructed

models using a kinematic approach that extract features from frame sequences for human structure representations. Ref. [70] used a neural network technique to propose an end-to-end hierarchical recurrent neural network (RNN) for representing skeleton-based construction. They made use of the raw positions of human joints as input to the RNN. A combination of both feature-based and model-based approaches for the classification of activities was shown in [3]. The authors used a maximum entropy Markov model (MEMM) for the classification of activities using features from skeleton tracking combined HOG.
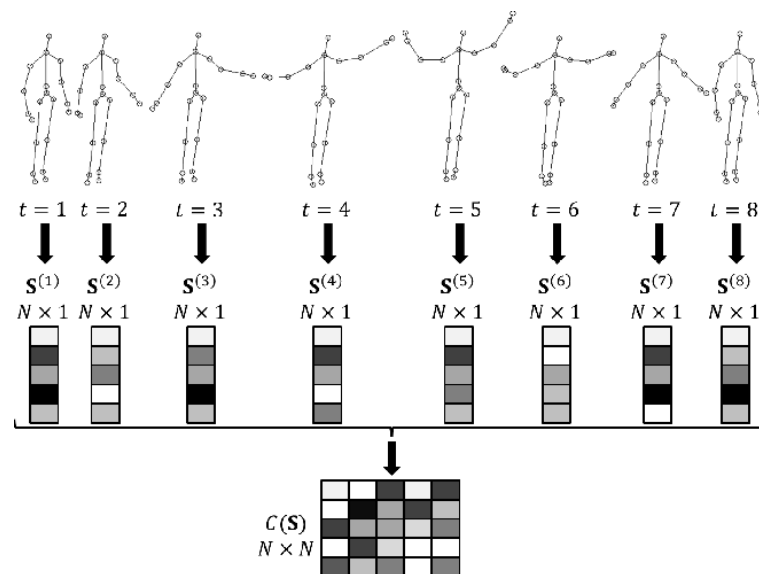

Figure 4. Representations of human activity features based on the statistical covariance of 3D joints (Cov3DJ) [67].

2.4. Recognition and Classification of 3D Skeletal Human Activity

Following the extraction of features from 3D skeletal human activity data, the processed features are used in a classification step for learning/recognition of human activities. There are many techniques to use, which range from statistical to CI methods, in the recognition process. The classification process involves grouping activities from observed sequences based on the similarities identified from features.

2.4.1. Classification with Statistical and Classical Machine Learning Algorithms

Statistical and classical ML techniques such as support vector machines (SVM), k-nearest neighbour (KNN), naive Bayesian and latent Dirichlet allocation (LDA) are some of the commonest methods applied in HAR using 3D human skeleton data [71]. Classification of human activities is carried out by extracting relevant features from data obtained using RGB-D sensors. The work in [72] proposed a method for activity recognition using RGB-D data. The 3D joint position information extracted from the sensor was transformed into feature vectors by applying selected soft computing techniques to group key postures of an activity. The posture features were used as input to a learning algorithm for the classification of human activities. SVM and KNN algorithms were used separately in classifying activities, and the results were compared. The SVM algorithm used in classifying 3D human activity skeleton data [20,42,72] works by finding the optimal hyper-plane which allows separation between distinct classes in an observed feature space. It uses a kernel function ϕ that allows the transformation of activity feature spaces to a higher dimensional space where the data are separable. Ref. [22] applied random forest (RF) in a framework using max-min features from human activity skeleton data. They proposed an extension to the traditional RF which combines a DE meta-heuristic algorithm with RF to optimise recognition performance.

2.4.2. Recognition of Human Activities Using Probabilistic Models

In the work presented in [2], the authors proposed using a probabilistic classification method in a framework that combines multiple classifiers to form a dynamic Bayesian mixture model (DBMM) for characterising activities from features obtained from distances between different parts of the body. The Bayesian mixture model was integrated into a dynamic process that took into consideration the temporal information of activities. The use of non-parametric approaches which are capable of dealing with a large number of classes and the problem of overfitting has been proposed as a solution for HAR using 3D skeleton data. For example, [73] proposed a naive Bayes nearest neighbour (NBNN) approach to recognise human actions from the accumulated motion energy computed from 3D human skeleton joints. Such methods require no learning process. Other techniques have been applied for sequence-based classification of human activities using 3D skeleton information, among which are dynamic time warping (DTW) and Markov models [65,74]. Markov models such as HMMs are very useful for modelling activity sequences, and thus they are very resourceful in the recognition of activities. By defining the elements of an HMM which are given to be the prior distribution for initial states, the emission matrix and the transition matrix, an HMM can be used to calculate the probability of an action for a given activity sequence consisting of observed human key poses.

2.4.3. Recognition of Human Activities Using Fuzzy Systems

Regarding 3D skeleton data HAR, CI methods have also been extensively studied by researchers. CI is a collection of nature-inspired computational models that are used to solve complex real-world problems which traditional statistical or ML techniques struggle with. Due to the uncertainties inherent in the problems, such problems might be too complex for mathematical inference or may be stochastic in structure. Human activity actions when observed through 3D visual information can be complicated with a lot of uncertainties when distinguishing one activity from a set of related activities. CI methods such as fuzzy logic [13,75,76] are suited for such recognition applications.

Ref. [75] used a fuzzy logic model for human behaviour recognition. Silhouette slices and movement speed from human silhouettes were used as input to the fuzzy system. A fuzzy c-means clustering algorithm was used to learn the fuzzy membership functions, and the human behaviour was then identified via selecting the behaviour category with the highest membership degree. Similarly, the work in [76] employed fuzzy logic in a view-invariant HAR system using a single camera. They used a fuzzy qualitative Poisson human model to extract a fuzzy qualitative human contour descriptor for human viewpoint analysis. Clustering algorithms were then applied to classify the viewpoints. These methods achieved reasonable performance in HAR. Other variations of fuzzy systems, such as evolving fuzzy systems in [13], have also been used. Fuzzy models are good at handling uncertainties in human activity data, which makes them a good tool in HAR.

2.4.4. Recognition of Human Activities Using Artificial Neural Networks

Artificial neural networks (ANNs) is another branch of CI techniques applied in HAR [10,35,50]. They have been applied extensively in 3D human skeleton-based activity recognition. [43] employed an ANN model in their work on HAR. They extracted pose and motion features from video sequences of activities and applied a clustering technique for grouping actions in prototypical pose-motion trajectories. The classification model consisted of self-organising growing when required (SOGWR) networks to obtain continuous representations of inputs and determine the latent spatio-temporal dependencies. Other works using neural networks [10,51] took advantage of its ability to model complex and non-linear relationships which occur in human actions to attain high accuracies. Furthermore, ANNs, unlike other ML techniques, do not impose restrictions on input data due to their ability to learn hidden relationships in data, which makes them good in prediction scenarios.

With the recent evolution in technology, DL models [35] have also more recently been applied to activity recognition problems. The results show the robustness of the method in activity recognition. [70] proposed an end-to-end hierarchical RNN human skeleton recognition model that models the long-term contextual information of temporal activity sequences. Figure 5 shows an illustration of their proposed system for action recognition from human skeleton information. DL models are good at automatically learning the features from any dataset, and this makes them suitable for large and complex applications. Ref. [35] applied extreme learning machines for the classification of features obtained using a convolutional neural network (CNN). The method was tested on five human activity datasets and achieved high performance. Other works have used different variations of CNN for HAR [77,78]. In the study presented in [79], a sequence-to-sequence model based on DL was used to recognise ADLs, taking advantage of activity state representations. Many other applications using DL architectures in HAR can be seen in [10,50,51,78]. However, DL models require a large amount of data to achieve concise predictions of activities, and in most cases more resources, such as time and reliable processing architectures. Furthermore, using DL limits the flexibility of defining the features to be used in the classification stage. Implementing such DL architectures requires high processing power with a huge amount of computational resources for training, as some architectures take several days or weeks to train.
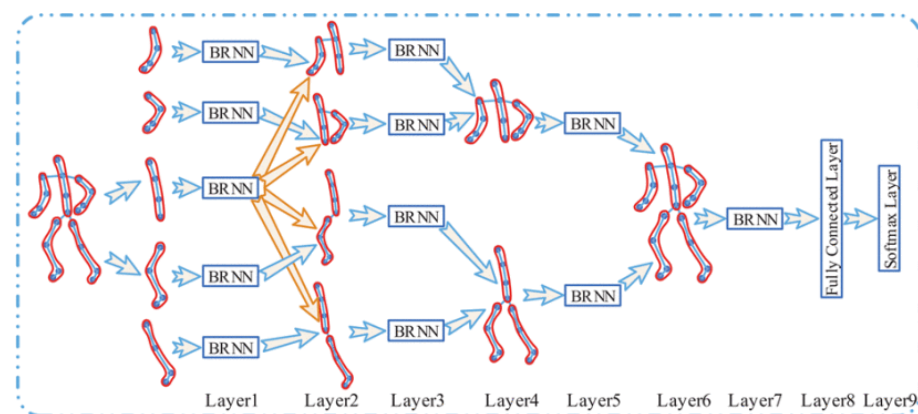

Figure 5. An illustration a hierarchical recurrent neural network system for action recognition [70].

2.4.5. Benchmark Performance of Different HAR Approaches

This section presents a performance comparison of the recognition and classification approaches used on different datasets. The performances were evaluated using three key metrics applied in HAR, precision, recall and accuracy. Table 3 presents the state-of-the-art performances reported by other works using the datasets. It can be observed that there is not a particular HAR approach that can be used to generalise and guarantee the optimal performance on all datasets. This is because each HAR approach varies in the setup, information modality and computational complexity when applied to datasets. In addition, some other datasets mentioned in Table 2 have not been included in the performance comparison in Table 3, as most research employing those datasets [58–61,63] has used them in the context of human pose estimation, which is beyond the scope of the review presented in this paper.

Table 3. State-of-the-art performances of different classification/recognition approaches on benchmark datasets.

Dataset	Approach	Modality	Prec. %	Rec. %	Acc. %
UAV-Human	Shift-GCN [55]	3D joints	-	-	37.98
Bimanual Actions Dataset	Graph network [56]	RGB + Depth	89.00	89.00	-
NTU RGB+D 120	Efficient-GCN [80]	3D joints	-	-	88.30
PKU-MMD	RF-Action [81]	3D joints	94.4	-	-
NTU RGB+D	Graph Convolution Network [82]	3D joints	-	-	96.5

2.5. Limitations of Vision-Based HAR

From the review presented, it is evident that HAR is a well-studied area with applications in many disciplines, hence the need for further research into solutions to improve current HAR systems. Although there have been many successes recorded in vision-based HAR, the complexities associated with occlusions, varying illuminations, changes in view, scale variance and activity similarity remain challenging in many applications. These have effects on the computational requirements of many systems. The limitations from the review on HAR presented are outlined as follows:

- Suitable data for HAR systems must be obtained, as the data have a defining impact on any system. In addition, the algorithms used for recognition should be investigated and selected based on the performances obtained with the information modality and other relevant factors.
- Most research focuses on activity classification using one person; however, action detection and activity pattern discovery require more investigation to provide a better understanding of the nature of activities.

3. Human Activities and Transfer Learning

Transfer learning (TL) has been well studied in different fields, include psychology, education, biology, computational intelligence (CI) and many others [83]. In CI, TL involves developing computational models which are capable of mimicking humans' ability to learn and reuse knowledge in different but related tasks. For example, the knowledge acquired while learning to eat with a spoon can be applied in learning to use chopsticks. This is a situation of TL from one task to another within the same information distribution. Another situation considered is when the source and target data are drawn from different information distributions—for example, transfer learning of activity from a human to a robot. The latter situation is very challenging to traditional ML techniques, and TL models seek to address this challenge by applying the knowledge learned from one information distribution in a different distribution [83].

The purpose of TL is to learn information from a reference source, and transfer the information to improve the performance achievable by the target. We focused on that goal in this review of TL in the context of human activity recognition (HAR). In HAR using 3D human information, TL is important in applications requiring knowledge transfer from a human information distribution to a different information distribution, such as a robot learning an activity, as illustrated in Figure 6. The entire processes of data acquisition, processing and activity recognition embodies TL, as there is the need for understanding the activity and both information distributions for the effective transfer to be implemented on a robot [84].

The following section discusses the ontology of the transfer learning of human activities and TL methods which apply such CI techniques as solutions to learning problems.

Figure 6. An illustration of the application of transfer learning in human activity recognition with different information distributions—a robot learning from a human.

3.1. Ontology of the Transfer Learning of Human Activities

Assisted living environments incorporate different technological solutions to improve quality of life and well-being. In recent years, there has been a growing interest in the research community in how to develop evolving solutions to aid assisted living, especially in areas of human activity recognition and learning. Different techniques have been studied to address the need for technological systems which are intelligent enough to evolve their knowledge to solve tasks which have not been previously encountered.

TL has recently attracted interest in recent years, due to the potential benefits it offers in artificial intelligence applications, including assisted living [83], computer vision [85] and robotics [86]. It has not recorded as much success as the long existing traditional machine learning (ML) methods, partly due to the challenges which yet remain unresolved in the research community [87], although it has the potential to become a fundamental driver for the success of ML in the coming years. The challenges facing TL implementations, as shown in Figure 7, depend on defining the metrics associated with the following aspects: *what to transfer*, *how to transfer* and *when to transfer*. Providing solutions to address these three aspects has been the focus of many researches, thereby motivating proposals of different TL algorithms.
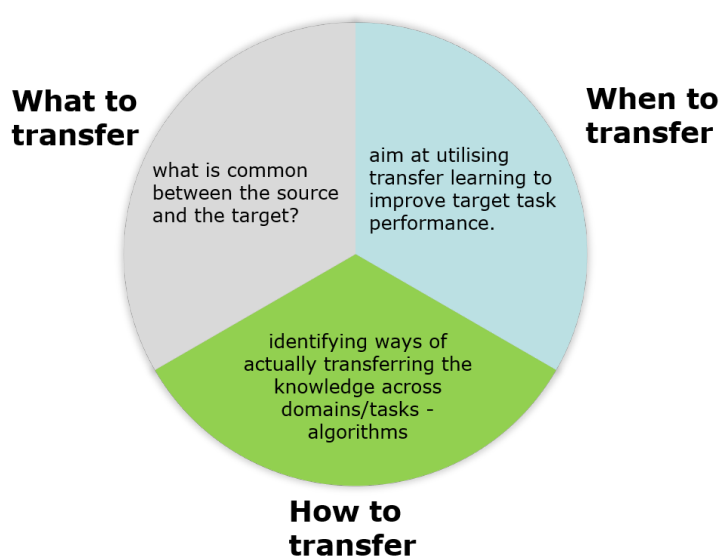

Figure 7. Challenges of transfer learning.

In relation to assisted living, different applications of TL have been studied. Ref. [83] proposed a model called fuzzy TL, which was applied in an intelligent environment. Data

from the source domain were learned by constructing a fuzzy inference system from the generated fuzzy rules. The constructed fuzzy inference system was then applied to a new domain referred to as the target domain through stages of adaptation of the generated fuzzy rules with the target data. Results from the model tested on real datasets from two intelligent environments (source and target environments), which were different but related, showed the model can achieve better performance in the target with the transfer of knowledge when compared to performance attained without transfer.

Ref. [88] proposed a method for improving robot learning manipulation tasks from data obtained from the robot performing other tasks or from similar robot architectures. Their method made an attempt to address the challenge of *how to transfer* by considering two steps, which include dimensionality reduction of data obtained from the robot to a low-dimensional space and manifold alignment of the source and target robot dimensions through a transformation function. The work in [86] also followed a similar approach of finding *how to transfer* between multi-robots. Even though these methods achieved impressive performances, the challenges of *what to transfer* and *when to transfer* still prove to be difficult in TL applications. Addressing these challenges will require consideration of properties related to spatial and temporal occurrences of both source/target domains.

TL methods usually employ various computational techniques as training models, such as neural networks [7], support vector machines [89] and rule-based models [83,90]. Therefore, it is important to review TL works that apply key computational intelligence techniques to solving problems.

3.2. Neural Network Transfer Learning Methods

Neural network architectures have been used in TL applications over the years, with results demonstrating superior performance compared to statistical models. However, most applications of neural networks in TL apply deep ANN architectures to propose solutions often referred to as *deep transfer learning* (DTL) solutions. A representation of the common configuration used in DTL is given in Figure 8 in which data from both source and target domains are mapped to a new data space which is then passed into a deep neural network. In a recent survey by [10], DTL was defined as a case of learning a target task where the objective predictive function, $f(\cdot)$, is a non-linear function that reflects a deep neural network. The effectiveness of deep neural networks in TL relies on the flexibility of its architecture when extracting high-level features which are transferable. This is possible due to the multiple hidden layers which can capture sophisticated non-linear representations in a dataset. In [91], a TL approach using deep neural networks was proposed for vehicle classification. The authors investigated the possibility of the TL of a pre-trained CNN model's parameters for classifying truck images generated from 3D point cloud data from LiDAR. Furthermore, in [92] four strategies of TL based on different configurations of CNN models were proposed for plant classification applications.

In relation to HAR, an important type of network applied in TL is the generative adversarial network (GAN) [93]. GANs are suited for TL due to their ability to learn generative models of data. The structure of the GAN features a generator and a discriminator, which makes use of a CNN and makes it possible to generate and augment several artificial samples of a training set that are very similar to the real data in the target domain. Several works have recently applied GANs for TL in domains including image processing [94] and HAR using wearable sensors [95]. The success of the many applications DTL has been applied to can mostly be attributed to the accessibility of DL architectures such as AlexNet [9], GoogleNet [96], VGG [97] and others, all of which can be pre-trained and configured to suit a variety of applications. Other methods of TL using neural networks for various applications can be found in [6].
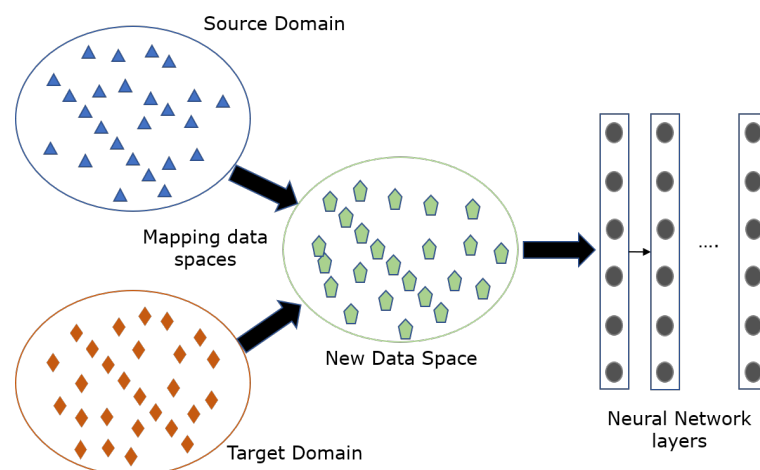
Figure 8. Mapping of data spaces in DTL. The source and target domains are mapped and the new data space used as input to a deep neural network.

3.3. Genetic Algorithm Transfer Learning Methods

Genetic algorithms (GA) are evolutionary computational methods inspired by natural selection to handle optimisation and global search problems. The algorithms are based on biological evolution operators, such as selection, mutation and crossover. Initially, GAs were used to solve complex non-linear optimisation problems, and later, they were used in hybrid techniques with other CI methods (such as fuzzy logic and neural networks) to solve classification and clustering problems. The authors in [98] proposed a genetic TL model which used two similar fitness functions to predict solutions for source and target tasks. The model aimed at maximising both functions by choosing the best samples and label variables. The results showed that the transfer of inter-task mappings was able to reduce the time required to learn a more complex task. However, there have not been many studies focusing on the application of GAs to TL.

3.4. Fuzzy Logic Transfer Learning Methods

Attempts to learn activities when there is little information available are often plagued with concerns of imprecision, vagueness, approximation and ambiguity of information. Therefore, it can be said that the level of certainty in any activity learning system and the availability of information are co-dependent. This is the reason many researches have incorporated fuzzy logic techniques into TL [99,100]. Incorporating fuzzy logic allows for approximation and expression of uncertainty encountered in the transfer of knowledge.

The concept of fuzzy logic was introduced in [101] as fuzzy set theory, and further expanded to include other aspects such as fuzzy rules [102]. The major elements of fuzzy logic are the if–then rules and the linguistic variable which captures imprecisions in a way similar to humans, which makes it relevant in TL. A fuzzy-based transductive TL model for predicting long-term bank failure was developed in [99,103]. The model applied a fuzzy similarity measure to refine predicted labels for samples in a target domain. Afterwards, the authors improved on the model by proposing a fuzzy refinement domain adaptation method which considers the similarity and dissimilarity in the refinement stage [104]. Ref. [100] proposed a framework for fuzzy transfer learning (FTL) for prediction in intelligent environments. The framework introduced the use of a transferable fuzzy inference system from a source domain that is adapted to a target domain. The method was applied in two simulated intelligent environments, and the experimental results indicated the proposed FTL framework outperformed classical prediction models, although the model was not compared with other TL models.

4. Research Opportunities

The opportunities and shortcomings identified from current research, as discussed in the review, are highlighted in this section.

A different approach to HAR uses non-visual sensory information due to some sensors, such as passive infrared (PIR), temperature and pressure sensors, being non-intrusive. However, other non-visual sensors, such as wearable sensors, can be intrusive, and as such may not be the best fit for HAR. Furthermore, people often find them uncomfortable and may forget to wear them while carrying out activities. Furthermore, as human activities differ in nature and sequence of occurrence, non-visual sensors are limited in the information they provide. It is often difficult to understand the nature of human actions, such as the positions/orientations of different parts of the human body during an activity using the information from non-visual sensors. This results in limitations when effectively creating models for human activity. On the other hand, vision-based approaches to HAR offer rich information (for example, depth, heat maps, coloured images and many others) from which a range of features can be extracted for high performance activity modelling and recognition algorithms.

Previous approaches to vision-based in HAR mostly focused on the technical aspects (a system's ability to accurately recognise activities) of the proposed systems [2,22]. These studies have been directed towards evaluating an algorithm's/model's ability to attain good performances in AR. However, not much research has been directed towards the practical applications of HAR.

TL has been studied in many contexts and applications. Most successful applications have been in object recognition from images [96,105]. Other applications in activity recognition [87,106] and robotics [86] have not achieved much success due to the complexities of TL. The work in [86] considered a multi-robot TL system. The work addressed TL from a control systems perspective by evaluating the performances of controllers. Ref. [87] proposed TL through feature space remapping with tests on activity recognition datasets. However, they only considered the case of a feature-rich dataset but did not address situations with sparse data.

5. Challenges and Future Directions

Developing solutions to help with assisted living is an ever growing field of interest in the research community. It involves the incorporation of a range of technological solutions in assisted living environments to enhance the quality of life and well-being. The rapid evolution of artificial intelligence techniques which are used to learn and model real-world behaviours has left the classical ML methods behind in terms of the performance levels obtainable. The classical learning models usually rely on situations where similar distributions of data are used in training and testing [83]. When there are changes in data distribution, such models fail. The models need to be retrained from scratch, which is a slow process and teaching a new model requires many data, which are always not readily available.

The differences in data distributions can be observed in many applications which involve AAL—for example, in assistive care for monitoring a person living independently [107], detecting changes/abnormality in an AAL environment [108] or learning daily routine activities of a person through an assistive agent. These and many more applications are increasingly encountered in pervasive technologies developed for assisted living. A solution to learning the differences in (or lack of sufficient) data distributions is TL. TL applies the knowledge acquired from one domain in a different but related domain to reduce the time needed for training the models from scratch and performance improvement [109]. This method has seen many applications in assisted living [83,106].

The relationship between the feature spaces in which TL has targeted influences the approach applied to achieve transfer of knowledge. This relationship can be either homogeneous or heterogeneous [109]. In the case of homogeneous TL, the feature spaces of the data in both source and target domains are equal. Situations involving homogeneous

TL are much simpler to accomplish when compared to heterogeneous transfer. The work proposed in [110] attempts TL by proposing a method of transfer component analysis (TCA) for domain adaptation. This method entails a dimensionality reduction framework for reducing the distance between domains in a latent space with similar features. The authors in [83] proposed a method of FTL for knowledge transfer. The approach considered the case of applying fuzzy logic to learn and transfer knowledge in intelligent environments. The authors showed that the performance achieved using the proposed FTL framework was comparable to those of other conventional methods of TL. Although the method in [83] performed well, it considered a situation in which labelled data were only present in the source domain and did not focus on the case of differing feature spaces.

Heterogeneous TL, on the other hand, is more challenging due to the fact that the feature spaces in both domains are drawn from different distributions of data [109]. The work in [90] proposed a method for a fuzzy rule-based approach to TL in both homogeneous and heterogeneous spaces. Furthermore, a heterogeneous TL method was shown in [111]. The incorporation of a fuzzy computational technique in [83,90] showed its advantage when applied to the transfer of knowledge to a target domain where critical information is inadequate. The benefits of heterogeneous TL enable it to be applied in many real-world applications [105,106].

The works reviewed in this paper have used different approaches to TL. Although these works achieved impressive performances when used in their respective applications, not much attention has been given to applications in activities of daily living, especially regarding human activity recognition in assisted living environments. The application of TL would be of great use in driving technological advancements.

6. Conclusions

This paper presented a review of the state-of-the-art research related to HAR and TL. The review presented HAR research based on visual sensory information. Different techniques to recognise activities were investigated. In assisted living, HAR plays a major role in the development of technological solutions to meet the needs of independent living. Although there are still failures in the practical implementations of such systems, their importance cannot be overemphasised.

TL, as an alternative to traditional learning methods, exists to aid in the transfer of knowledge across different but related situations of learning, so as to reuse knowledge and avoid having to train models from scratch, which is the case with traditional learning methods. By incorporating this concept in HAR, systems such as assistive robots can adapt to situations which require learning of activities by knowledge transfer from a human to robot space, etc. From the literature review, it was seen that simple, low-cost RGB-D sensors can be used to obtain rich information (which is relevant to any computational system) about activities.

Author Contributions: The work presented here is the result of the research conducted by D.A.A., which was conducted during his PhD study at Nottingham Trent University. The main supervisor, A.L., supervised the project. The final review was conducted by D.A.A., A.L. and R.R. All authors have read and agreed to the published version of the manuscript.

References

1. Han, F.; Reily, B.; Hoff, W.; Zhang, H. Space-time representation of people based on 3D skeletal data: A review. *Comput. Vis. Image Underst.* **2017**, *158*, 85–105. [CrossRef]
2. Faria, D.R.; Premebida, C.; Nunes, U. A Probabilistic Approach for Human Everyday Activities Recognition using Body Motion from RGB-D Images. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN, Edinburgh, England, 25–29 August 2014; pp. 732–737.

3. Sung, J.; Ponce, C.; Selman, B.; Saxena, A. Unstructured human activity detection from RGBD images. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MI, USA, 14–18 May 2012; pp. 842–849.

4. Shimodaira, H. Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Stat. Plan. Inference* **2000**, *90*, 227–244. [CrossRef]

5. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1–40. [CrossRef]

6. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl.-Based Syst.* **2015**, *80*, 14–23. [CrossRef]

7. Long, M.; Wang, J.; Cao, Y.; Sun, J.; Yu, P.S. Deep Learning of Transferable Representation for Scalable Domain Adaptation. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 2027–2040. [CrossRef]

8. Tan, S.; Sim, K.C.; Gales, M. Improving the interpretability of deep neural networks with stimulated learning. In Proceedings of the Workshop on Automatic Speech Recognition and Understanding (ASRU), Scottsdale, AZ, USA, 13–14 December 2015; pp. 617–623.

9. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—NIPS'12, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.

10. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In *Artificial Neural Networks and Machine Learning—ICANN 2018*; Springer International Publishing: Rhodes, Greece, 2018; pp. 270–279.

11. Wang, Y.; Han, X.; Liu, Z.; Luo, D.; Wu, X. Modelling inter-task relations to transfer robot skills with three-way RBMs. In Proceedings of the 2015 IEEE International Conference on Mechatronics and Automation (ICMA), Beijing, China, 2–5 August 2015; pp. 1276–1282.

12. Makondo, N.; Rosman, B.; Hasegawa, O. Knowledge transfer for learning robot models via Local Procrustes Analysis. In Proceedings of the 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), Seoul, Korea, 3–5 November 2015; pp. 1075–1082.

13. Iglesias, J.A.; Angelov, P.; Ledezma, A.; Sanchis, A. Human activity recognition based on Evolving Fuzzy Systems. *Int. J. Neural Syst.* **2010**, *20*, 355–364. [CrossRef]

14. Jalal, A.; Kamal, S. Real-time life logging via a depth silhouette-based human activity recognition system for smart home services. In Proceedings of the 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014, Seoul, Korea, 26–29 August 2014; pp. 74–80.

15. Sung, J.; Ponce, C.; Selman, B.; Saxena, A. Human Activity Detection from RGBD Images. In *Proceedings of the 16th AAAI Conference on Plan, Activity, and Intent Recognition*; AAAIWS'11-16; AAAI Press: Palo Alto, CA, USA, 2011; pp. 47–55.

16. Capela, N.A.; Lemaire, E.D.; Baddour, N. Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able bodied, Elderly, and Stroke Patients. *PLoS ONE* **2015**, *10*, 1–18. [CrossRef]

17. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Commun. Surv. Tutorials* **2013**, *15*, 1192–1209. [CrossRef]

18. Wang, Y.; Cang, S.; Yu, H. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **2019**, *137*, 167–190. [CrossRef]

19. Gupta, R.; Chia, A.Y.S.; Rajan, D. Human Activities Recognition Using Depth Images. In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; pp. 283–292.

20. Ni, B.; Pei, Y.; Moulin, P.; Yan, S. Multilevel Depth and Image Fusion for Human Activity Detection. *IEEE Trans. Cybern.* **2013**, *43*, 1383–1394.

21. Manzi, A.; Fiorini, L.; Limosani, R.; Dario, P.; Cavallo, F. Two-person activity recognition using skeleton data. *IET Comput. Vis.* **2018**, *12*, 27–35. [CrossRef]

22. Nunes, U.M.; Faria, D.R.; Peixoto, P. A human activity recognition framework using max-min features and key poses with differential evolution random forests classifier. *Pattern Recognit. Lett.* **2017**, *99*, 21–31. [CrossRef]

23. Aggarwal, J.; Xia, L. Human activity recognition from 3D data: A review. *Pattern Recognit. Lett.* **2014**, *48*, 70–80. [CrossRef]

24. Belagiannis, V.; Amin, S.; Andriluka, M.; Schiele, B.; Navab, N.; Ilic, S. 3D Pictorial Structures for Multiple Human Pose Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.

25. Burenius, M.; Sullivan, J.; Carlsson, S. 3D Pictorial Structures for Multiple View Articulated Pose Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013.

26. Zhang, H.; Parker, L.E. 4-dimensional local spatio-temporal features for human activity recognition. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, Francisco, CA, USA, 25–30 September 2011; pp. 2044–2049.

27. Han, F.; Yang, X.; Reardon, C.; Zhang, Y.; Zhang, H. Simultaneous Feature and Body-Part Learning for real-time robot awareness of human behaviors. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 2621–2628.

28. Sun, X.; Wei, Y.; Liang, S.; Tang, X.; Sun, J. Cascaded Hand Pose Regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.

29. Wang, Y.; Jiang, X.; Cao, R.; Wang, X. Robust Indoor Human Activity Recognition Using Wireless Signals. *Sensors* **2015**, *15*, 17195–17208. [CrossRef] [PubMed]

30. Chen, L.; Hsieh, J.; Chuang, C.; Huang, C.; Chen, D.Y. Occluded human action analysis using dynamic manifold model. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 1245–1248.

31. Zhou, W.; Zhang, Z. Human Action Recognition With Multiple-Instance Markov Model. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1581–1591. [CrossRef]

32. Chen, W.; Xiong, C.; Xu, R.; Corso, J.J. Actionness Ranking with Lattice Conditional Ordinal Random Fields. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–26 June 2014; pp. 748–755.

33. Xia, L.; Chen, C.C.; Aggarwal, J.K. View invariant human action recognition using histograms of 3D joints. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; pp. 20–27.

34. Belagiannis, V.; Amin, S.; Andriluka, M.; Schiele, B.; Navab, N.; Ilic, S. 3D Pictorial Structures Revisited: Multiple Human Pose Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1929–1942. [CrossRef] [PubMed]

35. Ijjina, E.P.; Chalavadi, K.M. Human action recognition in RGB-D videos using motion sequence information and deep learning. *Pattern Recognit.* **2017**, *72*, 504–516. [CrossRef]

36. Poppe, R. A Survey on Vision-based Human Action Recognition. *Image Vis. Comput.* **2010**, *28*, 976–990. [CrossRef]

37. Zhao, M.; Tian, Y.; Zhao, H.; Alsheikh, M.A.; Li, T.; Hristov, R.; Kabelac, Z.; Katabi, D.; Torralba, A. RF-Based 3D Skeletons. In Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, SIGCOMM '18, Budapest, Hungary, 20–25 August 2018.

38. Microsoft. Developing with Kinect for Windows. 2017. Available online: https://developer.microsoft.com/en-us/windows/kinect/develop (accessed on 28 February 2017).

39. Zhou, D.; Shi, M.; Chao, F.; Lin, C.M.; Yang, L.; Shang, C.; Zhou, C. Use of human gestures for controlling a mobile robot via adaptive CMAC network and fuzzy logic controller. *Neurocomputing* **2018**, *282*, 218–231. [CrossRef]

40. Chao, F.; Huang, Y.; Zhang, X.; Shang, C.; Yang, L.; Zhou, C.; Hu, H.; Lin, C.M. A robot calligraphy system: From simple to complex writing by human gestures. *Eng. Appl. Artif. Intell.* **2017**, *59*, 1–14. [CrossRef]

41. Gu, Y.; Do, H.; Ou, Y.; Sheng, W. Human gesture recognition through a kinect sensor. In Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO), Guangzhou, China, 11–14 December 2012; pp. 1379–1384.

42. Gaglio, S.; Re, G.L.; Morana, M. Human Activity Recognition Process Using 3-D Posture Data. *IEEE Trans. -Hum.-Mach. Syst.* **2015**, *45*, 586–597. [CrossRef]

43. Parisi, G.; Weber, C.; S, W. Self-Organizing Neural Integration of Pose-Motion Features for Human Action Recognition. *Front. Neurobotics* **2015**, *9*, 1–14

44. Benedek, C.; Gálai, B.; Nagy, B.; Jankó, Z. Lidar-Based Gait Analysis and Activity Recognition in a 4D Surveillance System. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 101–113. [CrossRef]

45. Zhu, Y.; Chen, W.; Guo, G. Evaluating spatiotemporal interest point features for depth-based action recognition. *Image Vis. Comput.* **2014**, *32*, 453–464. [CrossRef]

46. Jalal, A.; Kamal, S.; Kim, D. Depth silhouettes context: A new robust feature for human tracking and activity recognition based on embedded HMMs. In Proceedings of the 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Goyang City, Korea, 28–30 October 2015; pp. 294–299.

47. Jalal, A.; Nadeem, A.; Bobasu, S. Human Body Parts Estimation and Detection for Physical Sports Movements. In Proceedings of the 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), Islamabad, Pakistan, 6–7 March 2019; pp. 104–109.

48. Anh, D.N. Detection of lesion region in skin images by moment of patch. In Proceedings of the 2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), Hanoi, Vietnam, 7–9 November 2016; pp. 217–222.

49. Wu, Q.; Xu, G.; Li, M.; Chen, L.; Zhang, X.; Xie, J. Human pose estimation method based on single depth image. *Iet Comput. Vis.* **2018**, *12*, 919–924. [CrossRef]

50. Fan, X.; Zheng, K.; Lin, Y.; Wang, S. Combining local appearance and holistic view: Dual-Source Deep Neural Networks for human pose estimation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1347–1355.

51. Toshev, A.; Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1653–1660.

52. Zhao, M.; Li, T.; Alsheikh, M.A.; Tian, Y.; Zhao, H.; Torralba, A.; Katabi, D. Through-Wall Human Pose Estimation Using Radio Signals. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7356–7365. [CrossRef]

53. Li, Q.; He, F.; Wang, T.; Zhou, L.; Xi, S. Human Pose Estimation by Exploiting Spatial and Temporal Constraints in Body-Part Configurations. *IEEE Access* **2017**, *5*, 443–454. [CrossRef]

54. Akhter, I.; Black, M.J. Pose-conditioned joint angle limits for 3D human pose reconstruction. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1446–1455.

55. Li, T.; Liu, J.; Zhang, W.; Ni, Y.; Wang, W.; Li, Z. UAV-Human: A Large Benchmark for Human Behavior Understanding With Unmanned Aerial Vehicles. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 16266–16275.

56. Dreher, C.R.G.; Wächter, M.; Asfour, T. Learning Object-Action Relations from Bimanual Human Demonstration Using Graph Networks. *IEEE Robot. Autom. Lett.* **2020**, *5*, 187–194. [CrossRef]

57. Liu, J.; Shahroudy, A.; Perez, M.; Wang, G.; Duan, L.Y.; Kot, A.C. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2684–2701. [CrossRef] [PubMed]

58. Mahmood, N.; Ghorbani, N.; Troje, N.F.; Pons-Moll, G.; Black, M.J. AMASS: Archive of Motion Capture as Surface Shapes. In Proceedings of the International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 5442–5451.

59. von Marcard, T.; Henschel, R.; Black, M.; Rosenhahn, B.; Pons-Moll, G. Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

60. Mehta, D.; Sotnychenko, O.; Mueller, F.; Xu, W.; Sridhar, S.; Pons-Moll, G.; Theobalt, C. Single-Shot Multi-Person 3D Pose Estimation From Monocular RGB. In Proceedings of the 3D Vision (3DV), 2018 Sixth International Conference, Verona, Italy, 5–8 September 2018.

61. Trumble, M.; Gilbert, A.; Malleson, C.; Hilton, A.; Collomosse, J. Total Capture: 3D Human Pose Estimation Fusing Video and Inertial Sensors. In Proceedings of the 2017 British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017.

62. Liu, C.; Hu, Y.; Li, Y.; Song, S.; Liu, J. PKU-MMD: A Large Scale Benchmark for Skeleton-Based Human Action Understanding. In Proceedings of the Workshop on Visual Analysis in Smart and Connected Communities, VSCC '17, Mountain View, CA, USA, 23 October 2017; pp. 1–8.

63. Tayyub, J.; Hawasly, M.; Hogg, D.C.; Cohn, A. CLAD: A Complex and Long Activities Dataset with Rich Crowdsourced Annotations. *arXiv* **2017**, arXiv:1709.03456.

64. Shahroudy, A.; Liu, J.; Ng, T.; Wang, G. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1010–1019.

65. Shan, J.; Akella, S. 3D human action segmentation and recognition using pose kinetic energy. In Proceedings of the 2014 IEEE International Workshop on Advanced Robotics and its Social Impacts, Evanston, IL, USA, 11–13 September 2014; pp. 69–75.

66. Subetha, T.; Chitrakala, S. A survey on human activity recognition from videos. In Proceedings of the 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 25–26 February 2016; pp. 1–7.

67. Hussein, M.E.; Torki, M.; Gowayyed, M.A.; El-Saban, M. Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*; AAAI Press: Beijing, China, 2013; pp. 2466–2472.

68. Wei, P.; Zheng, N.; Zhao, Y.; Zhu, S.C. Concurrent action detection with structural prediction. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 3136–3143.

69. Vemulapalli, R.; Arrate, F.; Chellappa, R. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 588–595.

70. Du, Y.; Wang, W.; Wang, L. Hierarchical recurrent neural network for skeleton based action recognition. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1110–1118.

71. Torrey, L.; Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2009; pp. 242–264.

72. Cippitelli, E.; Gasparrini, S.; Gambi, E.; Spinsante, S. A Human Activity Recognition System Using Skeleton Data from RGBD Sensors. *Comput. Intell. Neurosci.* **2016**, *2016*. [CrossRef] [PubMed]

73. Yang, X.; Tian, Y. Effective 3D action recognition using EigenJoints. *J. Vis. Commun. Image Represent.* **2014**, *25*, 2–11. [CrossRef]

74. Piyathilaka, L.; Kodagoda, S. Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features. In Proceedings of the 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), Melbourne, Australia, 19–21 June 2013; pp. 567–572.

75. Yao, B.; Hagras, H.; Alhaddad, M.J.; Alghazzawi, D. A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments. *Soft Comput.* **2015**, *19*, 499–506. [CrossRef]

76. Lim, C.H.; Chan, C.S. Fuzzy qualitative human model for viewpoint identification. *Neural Comput. Appl.* **2016**, *27*, 845–856. [CrossRef]

77. Zhang, F.; Niu, K.; Xiong, J.; Jin, B.; Gu, T.; Jiang, Y.; Zhang, D. Towards a Diffraction-Based Sensing Approach on Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–25. [CrossRef]

78. Tang, C.I.; Perez-Pozuelo, I.; Spathis, D.; Brage, S.; Wareham, N.; Mascolo, C. SelfHAR: Improving Human Activity Recognition through Self-Training with Unlabeled Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–30. [CrossRef]

79. Zhu, H.; Chen, H.; Brown, R. A sequence-to-sequence model-based deep learning approach for recognizing activity of daily living for senior care. *J. Biomed. Informatics* **2018**, *84*, 148–158. [CrossRef] [PubMed]

80. Song, Y.F.; Zhang, Z.; Shan, C.; Wang, L. Stronger, Faster and More Explainable: A Graph Convolutional Baseline for Skeleton-Based Action Recognition. In Proceedings of the 28th ACM International Conference on Multimedia (ACMMM), Seattle, WA, USA, 12–16 October 2020; pp. 1625–1633.

81. Li, T.; Fan, L.; Zhao, M.; Liu, Y.; Katabi, D. Making the Invisible Visible: Action Recognition Through Walls and Occlusions. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 872–881.

82. Obinata, Y.; Yamamoto, T. Temporal Extension Module for Skeleton-Based Action Recognition. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 534–540.

83. Shell, J.; Coupland, S. Fuzzy Transfer Learning: Methodology and application. *Inf. Sci.* **2015**, *293*, 59–79. [CrossRef]

84. Adama, D.A.; Lotfi, A.; Ranson, R.; Trindade, P. Transfer Learning in Assistive Robotics: From Human to Robot Domain. In Proceedings of the 2nd UK-RAS Conference on Embedded Intelligence (UK-RAS19), Loughborough, UK, 24 January 2019; pp. 60–63.

85. MathWorks Inc. Transfer Learning Using AlexNet 2018. Available online: https://www.mathworks.com/help/deeplearning/examples/transfer-learning-using-alexnet.html (accessed on 30 December 2018).

86. Helwa, M.K.; Schoellig, A.P. Multi-robot transfer learning: A dynamical system perspective. In Proceedings of the International Conference on Intelligent Robots and Systems (IROS). IEEE/RSJ, Vancouver, BC, Canada, 24–28 September 2017; pp. 4702–4708.

87. Feuz, K.D.; Cook, D.J. Transfer Learning Across Feature-Rich Heterogeneous Feature Spaces via Feature-Space Remapping (FSR). *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 3:1–3:27. [CrossRef]

88. Bócsi, B.; Csató, L.; Peters, J. Alignment-based transfer learning for robot models. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–7.

89. Tommasi, T.; Orabona, F.; Caputo, B. Learning Categories From Few Examples With Multi Model Knowledge Transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 928–941. [CrossRef]

90. Zuo, H.; Lu, J.; Zhang, G.; Pedrycz, W. Fuzzy Rule-Based Domain Adaptation in Homogeneous and Heterogeneous Spaces. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 348–361. [CrossRef]

91. Vatani Nezafat, R.; Sahin, O.; Cetin, M. Transfer Learning Using Deep Neural Networks for Classification of Truck Body Types Based on Side-Fire Lidar Data. *J. Big Data Anal. Transp.* **2019**, *1*, 71–82. [CrossRef]

92. Kaya, A.; Keceli, A.S.; Catal, C.; Yalic, H.Y.; Temucin, H.; Tekinerdogan, B. Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* **2019**, *158*, 20–29. [CrossRef]

93. Soleimani, E.; Nazerfard, E. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing* **2021**, *426*, 26–34. [CrossRef]

94. Wang, Y.; Wu, C.; Herranz, L.; van de Weijer, J.; Gonzalez-Garcia, A.; Raducanu, B. Transferring GANs: Generating Images from Limited Data. In Prooceedings of the the European Conference on Computer Vision (ECCV) 2018, Munich, Germany, 8–14 September 2018.

95. Saeedi, R.; Sasani, K.; Norgaard, S.; Gebremedhin, A.H. Personalized Human Activity Recognition using Wearables: A Manifold Learning-based Knowledge Transfer. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 1193–1196.

96. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.

97. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

98. Koçer, B.; Arslan, A. Genetic Transfer Learning. *Expert Syst. Appl.* **2010**, *37*, 6997–7002. [CrossRef]

99. Behbood, V. Fuzzy Transfer Learning for Financial Early Warning System. Ph.D. Thesis, University of Technology, Sydney, Australia, 2013.

100. Shell, J. Fuzzy Transfer Learning. Ph.D. Thesis, De Montfort University, Leicester, UK, 2013.

101. Zadeh, L. Fuzzy sets. *Inf. Control.* **1965**, *8*, 338–353. [CrossRef]

102. Bellman, R.E.; Zadeh, L.A. Decision-Making in a Fuzzy Environment. *Manag. Sci.* **1970**, *17*, B141–B164. [CrossRef]

103. Behbood, V.; Lu, J.; Zhang, G. Fuzzy bridged refinement domain adaptation: Long-term bank failure prediction. *Int. J. Comput. Intell. Appl.* **2013**, *12*, 1350003. [CrossRef]

104. Behbood, V.; Lu, J.; Zhang, G. Fuzzy Refinement Domain Adaptation for Long Term Prediction in Banking Ecosystem. *IEEE Trans. Ind. Informatics* **2014**, *10*, 1637–1646. [CrossRef]

105. Day, O.; Khoshgoftaar, T.M. A survey on heterogeneous transfer learning. *J. Big Data* **2017**, *4*, 29. [CrossRef]

106. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556. [CrossRef] [PubMed]

107. Ortega-Anderez, D.; Lotfi, A.; Langensiepen, C.; Appiah, K. A multi-level refinement approach towards the classification of quotidian activities using accelerometer data. *J. Ambient. Intell. Humaniz. Comput.* **2018**, *10*, 4319–4330. [CrossRef]

108. Elbayoudi, A.; Lotfi, A.; Langensiepen, C. The human behaviour indicator: A measure of behavioural evolution. *Expert Syst. Appl.* **2019**, *118*, 493 – 505. [CrossRef]

109. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]

110. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain Adaptation via Transfer Component Analysis. *IEEE Trans. Neural Netw.* **2011**, 22, 199–210. [CrossRef] [PubMed]
111. Liu, F.; Zhang, G.; Lu, H.; Lu, J. Heterogeneous Unsupervised Cross-domain Transfer Learning. *arXiv* **2017**, arXiv:1701.02511.