Gesture Recognition for Human-Robot Collaboration: A Review

Hongyi Liu[1], Lihui Wang[2]

KTH Royal Institute of Technology, Department of Production Engineering, Stockholm, Sweden
[1]hongyil@kth.se, [2]lihuiw@kth.se

Abstract

Recently, the concept of human-robot collaboration has raised many research interests. Instead of robots replace human workers in workplaces, human-robot collaboration is the direction that allows human workers and robots working together. Human-robot collaboration can release human workers from heavy tasks if effective communication channels between human workers and robots are established. Although the communication channels between human workers and robots are still limited, gesture recognition has been effectively applied as the interface between humans and computers for long time. Covering some of the most important technologies and algorithms of gesture recognition, this paper is intended to provide an overview of the gesture recognition research, and explore the possibility to apply gesture recognition in human-robot collaboration. In this paper, an overall model of gesture recognition for human-robot collaboration is also proposed. There are four essential technical components in the model of gesture recognition for human-robot collaboration: sensor technologies, gesture identification, gesture tracking and gesture classification. Reviewed approaches are classified according to the four essential technical components. Statistical analysis is presented after technical analysis. In the last part of the review, future research trends are outlined.

Keywords: human-robot collaboration, gesture, gesture recognition.

1. Introduction

1.1 Human-Robot Collaboration

Robotic systems have already become key components in various industrial sectors. Recently, the concept of Human-Robot Collaboration (HRC) has raised research interests. Literature suggested that human workers have incomparable problem-solving skills and sensory-motor capabilities, but are restricted in force and precision [1, 2]. Robotic systems however provide better fatigue, higher speed, higher repeatability and better productivity, but are restricted in flexibility. Jointly, HRC can release human workers from heavy tasks and establish communication channels between human workers and robots for better overall performance.

Ideally, an HRC team should work similarly as a human-human collaborative team. However, traditionally, time-separation or space-separation approaches are applied in HRC systems, which reduced productivity for both human workers and robots [1]. In order to build an efficient HRC team, human-human collaboration teams can be analysed as examples. In human teamwork and collaboration, there are two theories: joint intention theory and situated learning theory [3-6]. To apply the theories into an HRC team, there are three experiences that benefit HRC team:

- All team members in an HRC team should share the same plan of execution;
- All team members in an HRC team should be aware of the context of the collaboration environment; and
- An HRC team should have structured ways of communication.

This paper mainly focuses on the third experience, i.e. structured ways of communication.

1.2 Gesture Recognition

Gesture is one communication method. Head nodding, hand gestures and body postures are effective communication channels in human-human collaboration [2, 7]. Gestures can be categorised into three types [8]:

- Body gestures: full body actions or motions,
- Hand and arm gestures: arm poses, hand gestures, and
- Head and facial gestures: nodding or shaking head, winking lips.

Gesture recognition refers to the mathematical interpretation of human motions by a computing device. In order to collaborate with human workers, robots need to understand human gestures correctly and act based on the gestures efficiently. In HRC environment, a natural way of gesture communication between robots and humans should be available.

1.3 Gesture Recognition for Human-Robot Collaboration

To recognise gestures in the HRC context, it is beneficial to investigate into a generic and simplified human information processing model. As shown in Fig. 1, Parasuraman et al. [9] generalised human information processing into a four-stage model. Based on the generic model in Fig. 1, we propose a specific model for gesture recognition in HRC. As shown in Fig. 2, there are five essential parts related to gesture recognition for HRC: sensor data collection, gesture identification, gesture tracking, gesture classification and gesture mapping, as explained as follows.

Fig. 1. A four-stage model of human information processing [9].

- Sensor data collection: gesture raw data is captured by sensors.
- Gesture identification: in each frame, a gesture is located from the raw data.
- Gesture tracking: the located gesture is tracked during the gesture movement. For static gestures, gesture tracking is unnecessary.

- Gesture classification: tracked gesture movement is classified according to pre-defined gesture types.
- Gesture mapping: gesture recognition result is translated into robot commands and sent back to workers.

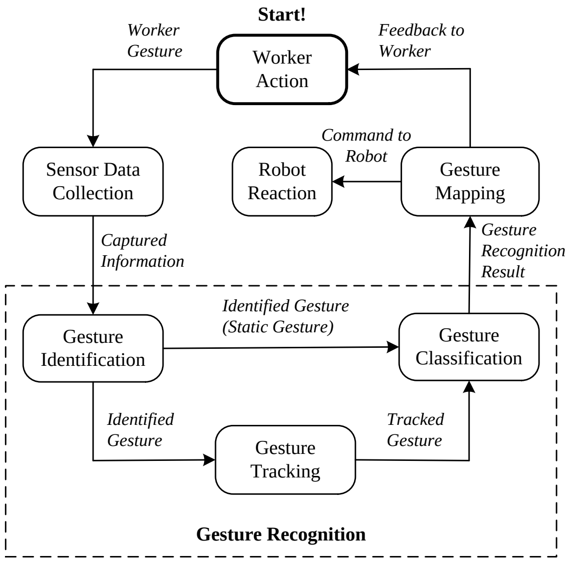

Fig. 2. A process model of gesture recognition for human-robot collaboration.

The remainder of this paper is organised as follows: Section 2 reviews enabling sensor technologies. Section 3 provides an overview of gesture identification methods. Section 4 discusses gesture tracking problems. Section 5 introduces gesture classification algorithms. Section 6 reveals statistical analysis of the reviewed papers. Section 7 concludes the paper with future research trends outlined.

2. Sensor Technologies

Before gesture recognition process started, gesture raw data need to be collected by sensors. In this section, different sensors in the literature are analysed based on different sensing technologies. As shown in Fig. 3, there are two basic categories of data acquisition: image based and non-image based approaches.

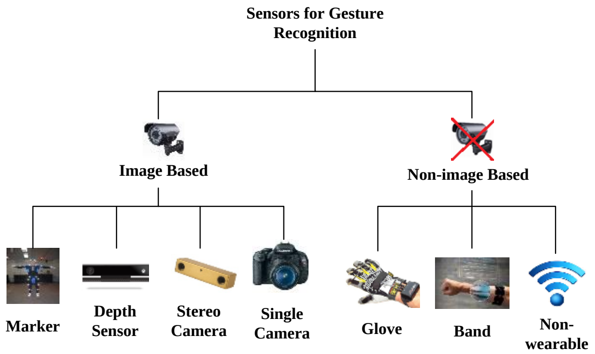

Fig. 3. Different types of gesture recognition sensors.

2.1 Image Based Approaches

Technologies are often inspired by nature. As human being, we use our eyes to recognise gestures. Therefore, for robots, it is reasonable to use cameras to "see" gestures. The image-based approaches are further divided into four categories.

- *Marker*

In marker-based approaches, the sensor is a normal optical camera. In most marker-based solutions, users need to wear obvious markers [8]. Today, we enjoy much faster graphical processing speed compared with twenty years ago. As a result, there are more gesture recognition sensors available.

- *Single Camera*

In the early 90th, researchers started to analyse gestures using a single camera [10, 11]. A drawback of single-camera approaches is the restriction of view angles, which affects a system's robustness [12]. Recent research, however, applied single-camera approaches in high-speed gesture recognition [13]. The system utilises speed image sensor and specially designed visual computing processor to achieve high-speed gesture recognition.

- *Stereo Camera*

In order to achieve robust gesture recognition, researchers suggested stereo camera based approaches to construct 3D vision. Here, we define stereo camera based approaches as applications that use two cameras (optical stereo camera) to construct 3D depth information. Many stereo camera based approaches followed a similar workflow [14, 15]. Although stereo camera systems improved robustness in outdoor environment, they still suffered from problems such as computational complexity and calibration difficulties [16].

- *Depth Sensor*

Recently, depth sensing technologies have emerged rapidly. We define a depth sensor as a non-stereo depth sensing device. Non-stereo depth sensors enjoy several advantages compared to the traditional stereo cameras. For example, the problems of setup calibration and illumination conditions can be prevented [17]. Moreover, the output of a depth sensor is 3D depth information. Compared with colour information, the 3D depth information simplifies the problem of gesture identification [8]. A comparison of gesture identification accuracy by using colour and depth information can be found in [18]. There are two types of common non-stereo depth sensors: Time-of-Flight (ToF) camera, and Microsoft Kinect (or similar IR sensors).

The basic principle of ToF cameras is to identify light travel time [19]. Various literature introduced hand gesture recognition applications based on ToF cameras [18, 20]. The advantage of ToF cameras is the higher frame rate. The limitation of a ToF camera is that the camera resolution highly depends on its light power and reflection [21].

Microsoft Kinect provides a cheap and easy solution for gesture recognition. Kinect is an IR light depth sensor. Similar sensors are ASUS Xtion Pro and Apple PrimeSense. Kinect has IR emitter, IR sensor and colour sensor. It is widely used in entertainment, education and research, which has introduced a large developer community [22-24]. With a large developer community, many open source tools and projects are available. Some researchers reported short distance Kinect-based hand gesture recognition systems [25-27]. Due to resolution restriction, currently, Kinect can be used for body gesture recognition and close distance hand

and arm gesture recognition. For hand and arm gesture recognition over 2 meters, other approaches are better used.

2.2 Non-Image Based Approaches

Gesture recognition has been dominated by image-based sensors for a long time. Recent developments in MEMS and sensor technology have greatly boosted non-image based gesture recognition technologies.

- *Glove*

Glove-based gestural interfaces are commonly used for gesture recognition. Usually, glove-based approaches require wire connection, accelerometers and gyroscopes. However, a cumbersome glove with load of cables can potentially cause problems in HRC [8, 28]. Glove-based approaches also introduced complex calibration and setup procedures [29].

- *Band*

Another contactless technology uses band-based sensors. Band-based sensors rely on a wristband or similar wearable devices. Band-based sensors enable wireless technology and electromyogram sensors, which avoids connecting cables. The sensor only needs to contact with wrist; user's hand and fingers can be released. An example of band-based sensor is Myo gesture control armband [30]. Recently, several band-based sensor gesture control systems were published [31-33].

- *Non-wearable*

The third type of non-image based technologies adopts non-wearable sensors. Non-wearable sensors can detect gestures without contacting human body. Google introduced Project Soli, a radio frequency spectrum (radar) based hand gesture tracking and recognition system [34]. The device is capable of recognising different hand gestures within a short distance. MIT has been leading non-wearable gesture recognition technology for years. Electric Field Sensing technology was pioneered by MIT [35]. A recent discovery by Adib et al. [36-38] from MIT introduced WiTrack and RF-Capture system that captures user motion by radio frequency signals reflected from human body. The systems are able to capture human gestures even from another room through a wall with a precision of 20 cm. In summary, non-wearable based technologies are promising and fast growing sensor technologies for gesture recognition.

2.3 Comparison of Sensor Technologies

A comparison of different sensor technologies is provided in Table 1. The advantages and disadvantages of different approaches are indicated. It is clear that there is no sensor fits all applications. Two observations of sensor technologies are provided based on the above analyses:

- With indoor applications, depth sensor approaches are the most promising image-based technologies. Depth sensors possess advantages of easy setup calibration and easy data processing. A large application development community exists, which provides ready solutions.
- Non-wearable approaches are the most promising technology among non-image based approaches. They avoid direct contact with users. Non-wearable sensing is also a fast-growing field.

Table 1
Advantages and disadvantages of different sensor technologies.

	Advantages	**Disadvantages**
Markers	Low computational workload	Markers on user body
Single camera	Easy setup	Low robustness
Stereo camera	Robust	Computational complexity, calibration difficulties
ToF camera	High frame rate	Resolution depends on light power and reflection
Microsoft Kinect	Fast emerging, software support for body gesture recognition	Cannot be used for hand gesture recognition over 2 meters
Glove	Fast response, precise tracking	Cumbersome device with load of cables
Band sensor	Fast response, large sensing area	Band needs to contact with human body
Non-wearable	Avoid contact with human body	Low resolution, technology not mature enough

3. Gesture Identification

Gesture identification is the first step in the gesture recognition process after raw data captured from sensors. Gesture identification means the detection of gestural information and segmentation of the corresponding gestural information from the raw data. Popular technologies to solve gesture identification problem are based on visual features, learning algorithms, and human models.

3.1. Visual Features

Human hands and body have unique visual features. In image-based gesture recognition, gestures consist of human hands or body. Therefore, it is straightforward to utilise such visual features in gesture identification.

- *Colour*

Colour is a simple visual feature to identify a gesture from background information. However, colour-based gesture recognition systems are easily influenced by illumination and shadows in a complex HRC environment [39]. Another common problem in skin colour detection is that human skin colour actually varies among human races. Due to the problems above, in recent approaches, skin colour is only considered to be one of many cues in gesture identification.

- *Local Features*

In image-based gesture recognition, illumination conditions greatly influence gesture identification quality. Therefore, many researchers utilise local features method that is not sensitive to illumination conditions. Local features approach is a detailed texture-based approach. It decomposes an image into smaller regions that are not corresponding to body parts [40]. As shown in Fig. 4, one of the most important local features is Scale Invariant Feature Transform (SIFT) [41]. The SIFT method is rotational, translational, scaling and partly illumination invariant. Several similar local feature approaches, for example, SURF and ORB are proposed in later years [42, 43]. Normally, local features approaches are also only considered as one of many cues in gesture identification. Several identification methods such as shape and contour methods, motion methods, and learning methods are based on local features.

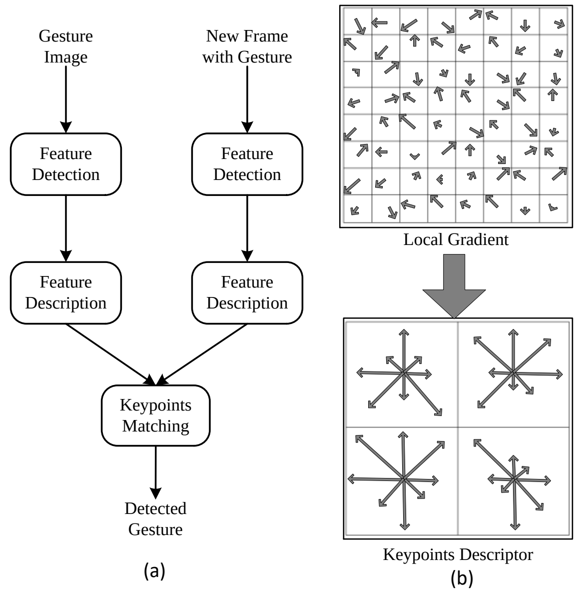

Fig. 4. SIFT algorithm: (a) SIFT algorithm for gesture identification; (b) SIFT feature description example [41].

- *Shape and Contour*

Another intuitive and simple way to identify gestures is to utilise the unique shape and contour of human body in HRC environment. A milestone for shape detection and matching is reported by Belongie et al. [44]. They introduced a shape context descriptor method. Shape context descriptor is used for detection of similar shapes in different images. The development of depth sensor provides opportunities to accurately measure surface shapes. The 3D models generated from the technologies enable highly detailed representation of human body shape [45].

- *Motion*

In certain HRC environment, a human worker is the only moving object in the raw data. Therefore, motion is a useful feature to detect human gestures. Optical flow is a key technology for motion-based gesture identification. Optical flow does not need background subtraction, which is an advantage compared to shape and contour based approaches. Several gesture recognition applications are implemented based on optical flow method [46, 47]. Dalal and Thurau [48] introduced the famous Histograms of Oriented Gradients (HOG) method. The HOG descriptors divide image frames into blocks. For each block, a histogram is computed. Among non-image based sensors, motion-based gesture identification is also a popular method [37, 49]. Usually, thresholding and filtering are applied to raw sensor data to identify human gestures.

3.2. *Learning Algorithms*

A recent trend of gesture identification is to apply learning algorithms, especially for static gesture detection that can be represented in a single frame. The visual feature methods are based on various visual features, while learning algorithms utilise machine learning algorithms to identify gestures from raw sensor data. Although some algorithms are based on the visual feature methods, image background removal is not always necessary for learning algorithms. Learning algorithms such as Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Random Decision Forests (RDF) are widely applied in gesture recognition systems [50-52].

3.3. *Human Model*

Different from the aforementioned approaches, human model approach uses an explicit human body model to recover human body poses. Therefore, human model approach is also called generative approach. Since human model based gesture identification provides an advantage of simplifying gesture classification process, human model approach has become a popular solution for depth sensors [53]. Depending on different model targets, recent research in human model based gesture identification can be divided into two types: hand model identification and body skeleton model identification.

- *Hand Model*

There are three types of hand model approaches for hand model detection: shape model, 3D model and hand skeleton model. Shape model based approach matches a pre-constructed hand shape model and shape features from observation [54]. 3D model approach interprets the hand detection problem as an optimisation problem that minimises the differences between 3D hand model and the observation [55]. Hand skeleton model approach is similar to 3D model approach [27].

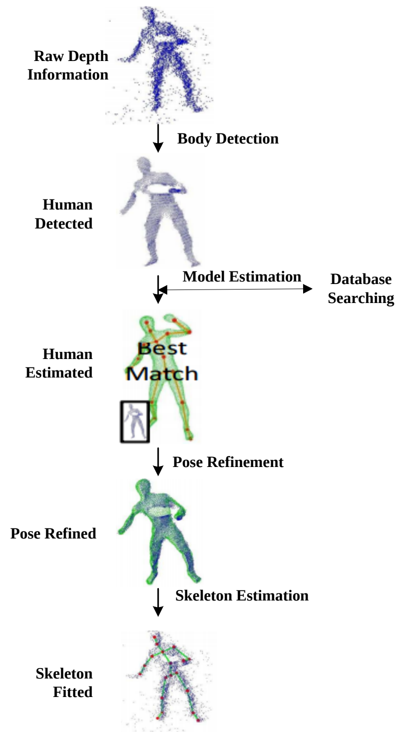

Fig. 5. Example of body skeleton identification [58].

- *Body Skeleton Model*

To identify body gestures, a detailed human model is useless. Therefore, in the field of body gesture identification, body skeleton model is a popular approach. Most of the literature regarding body skeleton processing pipeline are based on depth information gathered from depth sensors. As shown in Fig. 5, body skeleton model is a simplified human body model that preserves body joints information. Since only the most useful information is preserved from a human body, in body gesture recognition research, more and more body skeleton methods are applied [53, 56, 57].

3.4. Summary of Gesture Identification Approaches

In Table 2, the most suitable gesture identification method for each popular sensor is summarised. Due to the nature of HRC, human workers are the most important members of a HRC team. Despite understanding human body gestures, the human body model approach will also monitor human movements, which provides a secure environment for human-robot team. As mentioned earlier, body skeleton model simplifies human body, while body joint information is well preserved. Moreover, skeleton model approaches can simplify subsequent gesture classification process. Therefore, currently, body skeleton model approach is an appropriate solution for gesture recognition in HRC systems.

Table 2
Gesture identification methods for different sensors.

Sensor	Gesture identification method
Single camera	In single camera based systems, visual features method and learning algorithms method can be implemented. To achieve robust performance, learning algorithms should be applied. To achieve faster image processing, visual features method should be applied [13].
Depth sensor	Since human model method utilises and simplifies point cloud information, human model method is a better option for depth sensor based systems [53].
Band sensor	No visual based methods can be applied on band sensor based systems. Usually, the collected data need basic filtering in gesture identification, and learning algorithms will be implemented in later gesture classification [31, 59].
Non-wearable	The non-wearable sensors also receive signal data instead of images. Due to the fact that RF signals contain noises [37, 38], advanced filtering and processing solutions need to be implemented in non-wearable sensor based systems.

4. Gesture Tracking

In gesture recognition, the notion of tracking is used differently in different literatures. We define the notion of tracking as the process of finding temporal correspondences between frames. Specifically, we focus on gesture tracking problem that associates the identified gesture in the previous frames with the current frame. As for static gestures which can be represented by a single frame, gesture tracking is unnecessary. An example of gesture tracking is shown in Fig. 6.

4.1. Single Hypothesis Tracking

Single hypothesis tracking refers to a best-fit estimation with minimum-error matching. Therefore, in single hypothesis tracking, gesture is represented by only one hypothesis. Most of the advanced tracking algorithms are based on the single hypothesis tracking technologies.

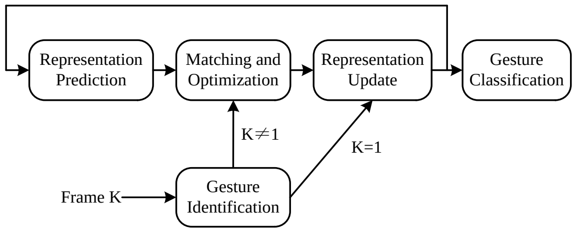

Fig. 6. Gesture tracking example.

- *Mean Shift*

Mean shift tracker is a basic tracking technology. Mean shift tracker performs matching with RGB-colour histograms [60]. For each new frame, mean shift tracker compares the Bhattacharyya distance between the target window histograms of the new frame with those of the old frame. A complete mathematical explanation can be found in [60].

- *Kalman Filter*

Kalman Filter (KF) is a real-time recursive algorithm used to optimally estimate the underlying states of a series of noisy and inaccurate measurement results observed over time. The process flow of KF is shown in Fig. 7. A complete KF mathematical derivation can be found in [61, 62]. Nowadays, KF is evolved and applied in different fields such as aerospace, robotics, and economics.

- *Kalman Filter Extensions*

KF is given a prerequisite that the state vector is a linear model. Extend Kalman Filter (EKF) is a functional tracking algorithm even if the model is nonlinear [63]. Another algorithm that solves the same problem from a different angle is Unscented Kalman Filter (UKF) [64]. UKF solves the problem by applying a deterministic weighted sampling approach. The state distribution is represented using a minimal set of chosen sample points.

4.2. Multiple Hypotheses Tracking

In many HRC scenarios, multiple human workers are working in the same work station at the same time [1]. To track multiple workers' gesture simultaneously, multiple hypotheses tracking technologies should be applied.

- *Particle Filter*

Particle filter (PF) is a popular technology in robotics problems. Different from KF, PF does not make assumption on posterior model. The PF representation is a nonparametric approximation which can represent a broader space of distribution. Therefore, PF satisfies multiple hypotheses tracking requirement [65]. An example of PF is shown in Fig. 8. Several advanced tracking algorithms also apply PF to scan probability density function [66-68].

- *Particle Filter Extensions*

Many researchers attempted to combine PF with other algorithms. Researchers have combined PF with mean shift tracker, Genetic Algorithm, PSO, Ant Colony Optimisation, and other machine learning algorithms to solve the sample degeneracy and impoverishment problem [70]. Some other researchers also improved PF resampling strategy [71, 72].

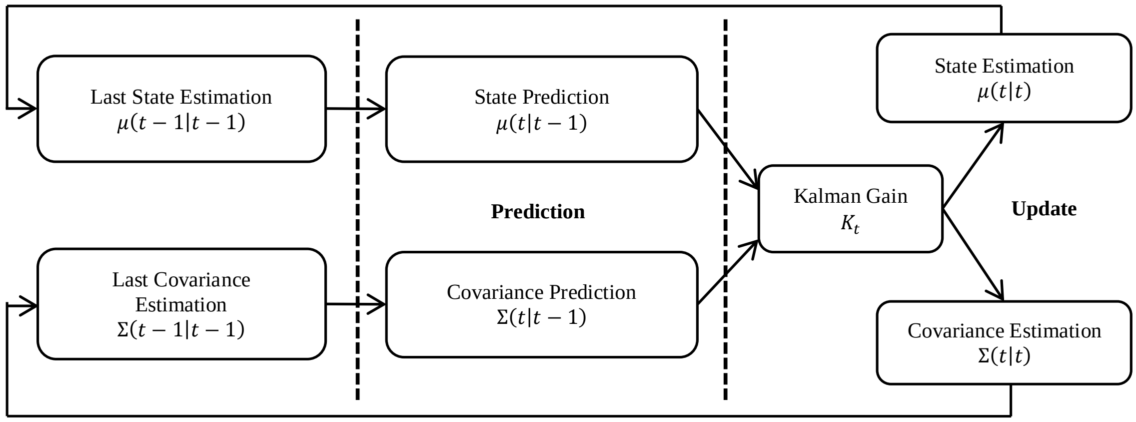

Fig. 7. Kalman filter process flow [69].

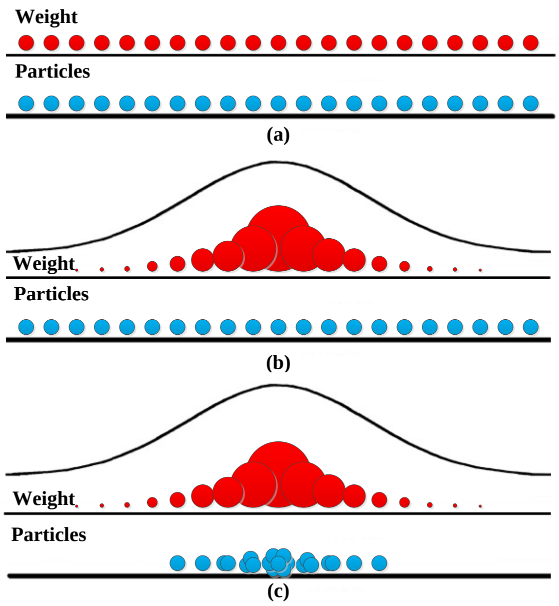

Fig. 8. Particles and weight factors: (a) after particles initialisation; (b) after weight factor calculation; (c) after resampling [69].

4.3. Advanced Tracking Methods

Recently, there are many advanced tracking methods introduced. Some of these advanced methods utilised part of the tracking algorithms mentioned above. Other methods improved tracking performance by detection or learning algorithms.

- *Extended Model Tracking*

For long-term tracking problems, many tracking algorithms fail because target maintains fixed models. Extended model tracking saves target behaviour or appearance from the past few image frames. Therefore, more target information is reserved for target estimation. Incremental Visual Tracker uses extended model to preserve more information for tracking process [73]. Kwon et al. [67] presented Tracking by Sampling Tracker. The extended model is preserved by a sampling process. The tracker samples many trackers and accordingly the appropriate tracker is selected.

- *Tracking by Detection*

Another kind of tracking algorithms is built together with the gesture identification learning algorithms introduced in the earlier sections. For these tracking algorithms, a classifier or detector is applied in image frames to identify gesture from the background information [68]. One representative approach is Tracking, Learning and Detection Tracker [74]. The approach integrates the result of an object detector with an optical flow tracker. Another typical tracking-by-detection technology is to apply Multiple Instance Learning [75]. The learning algorithm can increase tracker robustness and decrease parameter tweaks.

4.4. Comparison of Different Gesture Tracking Approaches

Smeulders et al. [76] presented a test result of different gesture tracking algorithms. The result score is normalised F-score. F-score provides us an insight of the average coverage of the tracked object bounding box and the ground truth bounding box. Therefore, the higher the F-score, the better the tracking quality. In Fig. 9, the test results in different video conditions are presented. *Kalman Appearance Tracker* and *Mean Shift Tracker* belong to the single hypothesis tracker. *Tracking by Sampling Tacker* and *Incremental Visual Tracker* belong to the extended model tracker. *Multiple Instance Learning Tracker* and *Tracking, Learning and Detection Tracker* belong to the tracking-by-detection method. It is easy to observe that the single hypothesis trackers perform lower than the others. A summary of different gesture tracking approaches is presented in Table 3.

Table 3
Summary of tracking approaches.

Approach	Summary
Single hypothesis	Fast and simple algorithm. Suitable for one gesture tracking in controlled environment.
Multiple hypotheses	Capable of tracking multiple targets at the same time. Suitable for multiple gestures tracking in controlled environment.
Extended model tracking	Target history is saved and available for target estimation. Suitable for long time gesture tracking task.
Tracking by detection	Learning algorithm increases robustness and reduces noise. This combined approach has the preferred performance in test. Suitable for gesture tracking in complex environment.

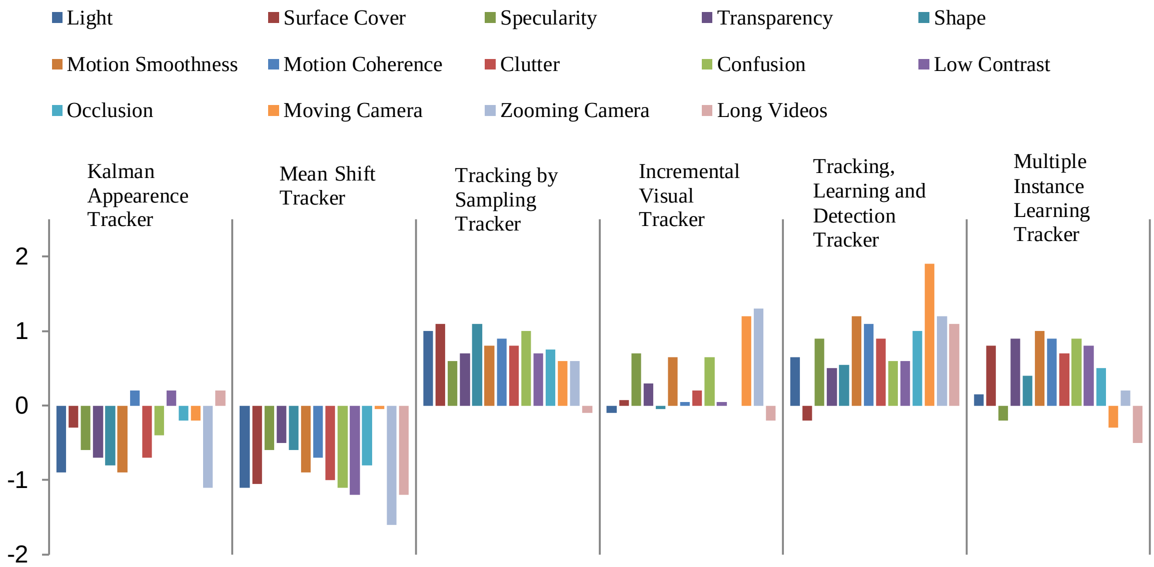

Fig. 9. Tracking algorithms test result in different video conditions [76].

5. Gesture Classification

Gesture classification is the last and the most important step in gesture recognition. Most of human gestures are dynamic gestures. One dynamic gesture always consists of several frames. To classify dynamic gestures, gesture classification has to be done after or together with gesture tracking.

5.1. K-Nearest Neighbours

K-Nearest Neighbours (KNN) algorithm is a fundamental and basic gesture classification algorithm that classifies input data according to the closest training examples [77]. Application of KNN in gesture classification can be found in [77].

5.2. Hidden Markov Model

Hidden Markov Model (HMM) is a popular gesture classification algorithm. The HMM is a combination of an un-observable Markov Chain and a stochastic process. An example of HMM is shown in Fig. 10, the un-observable Markov Chain consists of states X and state transition probabilities a. The stochastic process consists of possible observations O and output possibilities b. Gesture recognition is the problem that given observation sequence O, identify the most likely state sequence X [78, 79]. To solve the problem, Expectation Maximisation (EM) algorithm is applied [79]. There are many papers discussing HMM gesture recognition applications [80-82]. Some papers combined HMM with other classification approaches [81]. Others extended HMM algorithm into wider range of applications [82].

5.3. Support Vector Machine

As shown in Fig. 11, Support Vector Machine (SVM) is a discriminative classifier defined by a separating hyperplane [84, 85]. Classification decision boundaries are identified by maximising a margin distance. The optimal separation hyperplane maximises the margin of training data. The training examples closest to the optimal hyperplane are called support vectors. A common problem for SVM is that the number of support vectors grows linearly with the size of training set. Some researchers proposed Relevance Vector Machine (RVM) to solve the problem [86]. SVM kernel trick was introduced by Scholkopf [87]. SVM kernel trick enables linear SVM in non-linear problems. SVM kernel transforms low-dimensional training data into high-dimensional feature space with non-linear method [88]. There are also many papers that combined SVM with other classification methods to improve gesture classification performance [89-91].

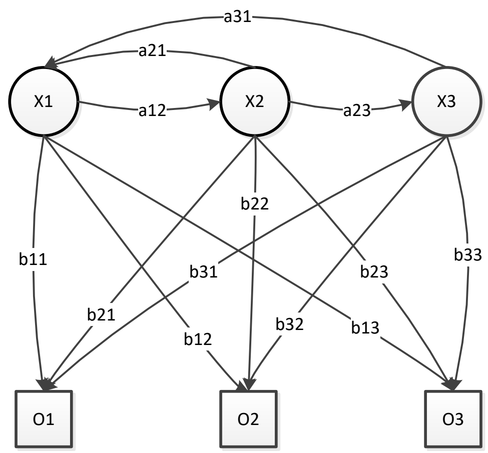

Fig. 10. Example of Hidden Markov Model [83].

5.4. Ensemble Method

Ensemble method is another type of widely-used gesture classification algorithm. The main assumption of ensemble method is that ensembles are more accurate than weak individual classifiers. One of the famous ensemble method is Boosting by Schapire et al. [92, 93]. Boosting algorithm starts with several weak classifiers. The weak classifiers are repeatedly applied. In a training iteration, part of training samples is used as input data. After the training iteration, a new classification boundary is generated. After all iterations, the boosting algorithm combines these boundaries and merges into one final prediction boundary. As shown in Fig. 12, another well-known ensemble method is Adaboost algorithm. A significant advantage of Adaboost algorithm is that Adaboost does not need many training data. Several papers applied Adaboost algorithm in gesture identification and gesture classification applications [94, 95]. The classification performs better when Adaboost algorithm is applied.

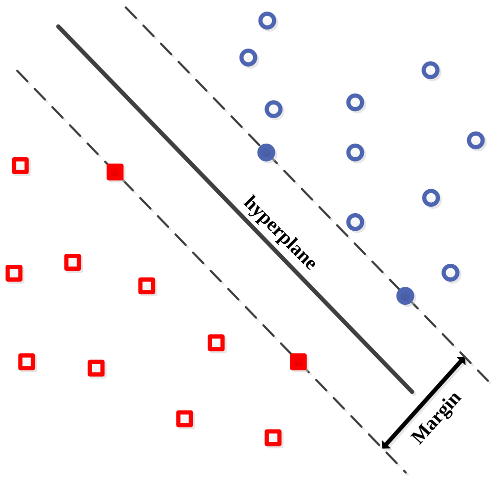

Fig. 11. Example of linear Support Vector Machine [84, 85].

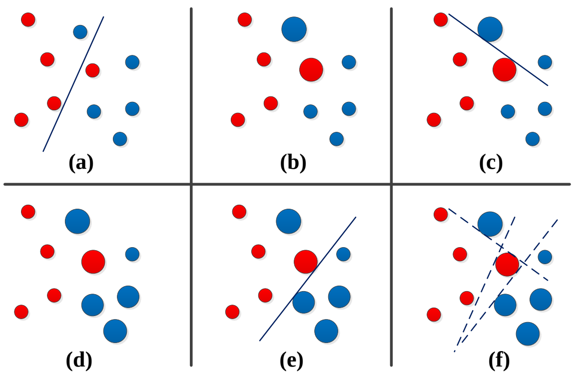

Fig. 12. Example of Adaboost: (a) weak classifier 1, (b) weights increase, (c) weak classifier 2, (d) weights increase, (e) weak classifier 3, (f) final decision boundary [93].

5.5. Dynamic Time Warping

Dynamic time warping (DTW) is an optimal alignment algorithm for two sequences. DTW generates a cumulative distance matrix that warps the sequences in a nonlinear way to match each other. Originally, DTW is used for speech recognition. Recently, there are many DTW applications in gesture recognition as well [96, 97]. Some papers also introduced Derivative Dynamic Time Warping (DDTW) as an extension of DTW [98, 99].

5.6. Artificial Neural Network

Artificial Neural Network (ANN) is a family of information processing models inspired by biological neural networks [100]. ANN consists of many interconnected processing unions (neurons) that work in parallel. Each union (neuron) receives input data, processes input data and gives output data. ANN can be used to estimate functions that depend on a large number of input data. Recently, there are many researchers who have utilised ANN for gesture recognition [101-103]. Several papers also presented gesture recognition systems that combined ANN with other classification methods [104-106].

5.7. Deep Learning

Deep learning is an emerging and fast-growing branch of machine learning. Deep learning enables computer to model data with high-level abstractions by using multiple processing layer neural network. Moreover, different from traditional learning algorithms, deep learning needs little engineering by hands, which enables the possibility to take advantages of exponentially increasing available data and computational power [107]. Nowadays, deep learning is applied in image recognition, speech recognition, particle accelerator data analysis etc. [108]. Especially, deep learning is employed for solving the problem of human action recognition in real-time video monitoring, which contains a large number of data [109, 110]. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are two popular deep learning architectures [107]. Several gesture recognition systems have applied above deep learning algorithms recently [111, 112].

Table 4
Advantages and disadvantages of gesture classification approaches.

Approach	Advantages	Disadvantages
K-Nearest Neighbours	Simple	K needs to be chosen carefully
Hidden Markov Model	Flexibility of training and verification, model transparency [113]	Many free parameters need to be adjusted [113]
Support Vector Machine	Different Kernel function can be applied [87]	Number of support vectors grows linearly with the size of training set [86]
Ensemble Method	Do not need large number of training data	Overfit easily, sensitive to noise and outliers
Dynamic Time Warping	Reliable non-linear alignment between patterns [114]	Time and space complexity [115]
Artificial Neural Network	Can detect complex nonlinear relationships between variables [116]	“Black box” nature and cannot be used for small training data set [116]
Deep Learning	Do not need good design of features, outperform other machine learning methods [107]	Need large number of training data and computationally expensive.

5.8. Comparison of Gesture Classification Approaches

Table 4 lists the advantages and disadvantages of the gesture classification approaches. One of the trends is deep learning approach. The main constrains of deep learning is the limited computational power. However, the exponentially increasing computation power can solve the problem easily. The number of deep learning based gesture classification applications is increasing rapidly. Another trend is to combine different classification algorithms. Every classification algorithm has own advantages and disadvantages. To utilise that, different classifiers can be combined to achieve better performance. We also observed that it is important to coordinate gesture classification algorithms with gesture identification and gesture tracking algorithms.

6. Statistical Analysis

This section presents a brief statistical analysis of gesture recognition technologies. As shown in Fig. 13, we selected 9 different journals and conferences that each published more than 5 papers within the 285 reviewed papers related to gesture recognition. In terms of number of papers, 65% are conferences papers, 35% are journal papers. Note that it is a common practice in computer science to publish extensively in conference proceedings. The most popular conference that published gesture recognition related papers is Conference on Computer Vision and Pattern Recognition. Fig. 14 shows the yearly distributions of the reviewed gesture recognition papers. It is clear that the number of gesture recognition papers increased rapidly after 1994, indicating the growing interests in this field.

Fig. 15 shows the yearly development trend of the four technical components in the field of gesture recognition. The horizontal axis represents the percentage of papers cited in a certain period of time as compared with all the papers reviewed on this technology. It is clear to observe:

- Depth sensor is the rapidly developing technology among sensor technologies. Band approach and non-wearable approach are also developing quickly.
- Human model method is the most promising technology among gesture identification approaches.
- Advanced tracking methods are emerging compared with other gesture tracking methods.
- Deep learning and ANN are growing fast among gesture classification technologies. Other technologies such as Hidden Markov Model and Support Vector Machine are frequently used, recently.

The statistical analysis confirmed our results of technical analyses in the earlier sections.

7. Conclusions and Future Trends

Although the above sections provide a general picture of gesture recognition for HRC, it is never easy to summarise such an interdisciplinary and fast developing field in any capacity. Sensor related technologies normally started from hardware. Software technologies and algorithms are designed to utilise the performance of hardware. Therefore, we would introduce some of predicted future trend start with sensor technologies.

- Depth sensor and human model based gesture recognition: due to the nature of HRC, human workers are the most important members of any HRC team. Despite understanding human body gestures, depth sensor together with body model approach will monitor human movements, which provides a safer environment for HRC. Moreover, body skeleton model will simplify gesture classification process. Therefore, simpler gesture tracking and classification method can be applied.
- Non-wearable sensor and deep learning based gesture recognition: although non-wearable sensor technologies are not ready, it is still the most promising non-image based sensor. In HRC, human workers should be able to communicate with robots naturally. For this very purpose, nothing should be attached to workers' body. Non-wearable sensors still suffer from low gesture identification and classification quality. The problem can potentially be solved using deep learning methods.
- Hard real-time gesture recognition system: one of the most important requirements in industry is the real-time requirement. Especially, in HRC system, the safety of human workers is always the first priority. Therefore, real-time gesture recognition system is another future direction. Currently, band and glove sensors provide the fastest response. Moreover, high speed single camera gesture recognition system is also emerging recently [13]. In gesture identification, tracking and classification, fast and effective methods can be applied.

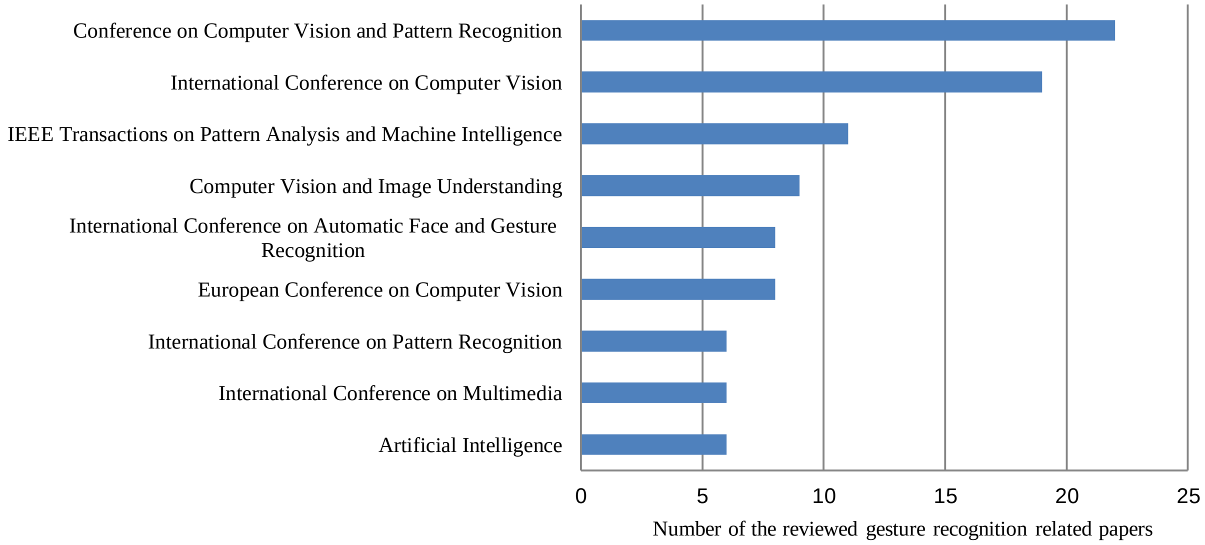

Fig. 13. Journals and conferences that publishing most gesture recognition related papers.

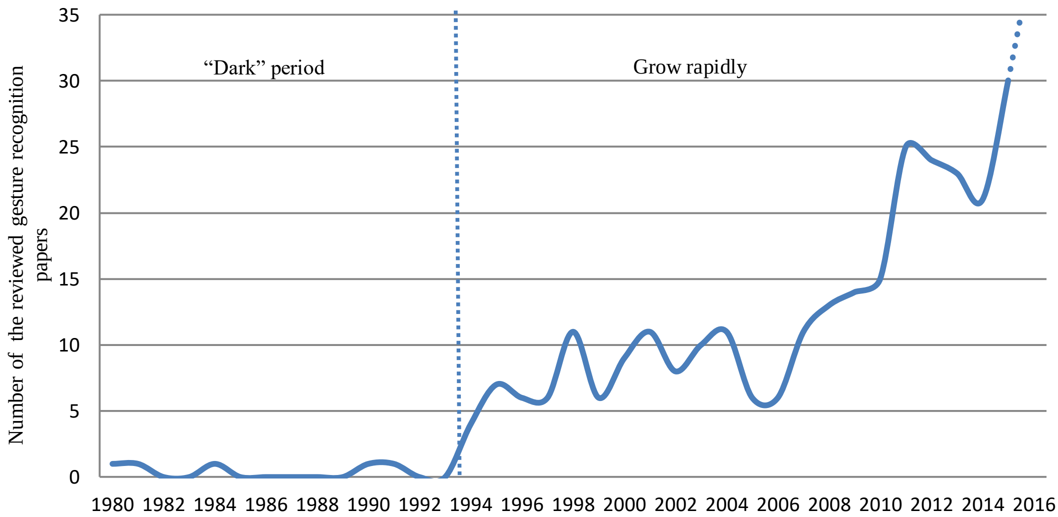

Fig. 14. Yearly distributions of the reviewed gesture recognition related papers.

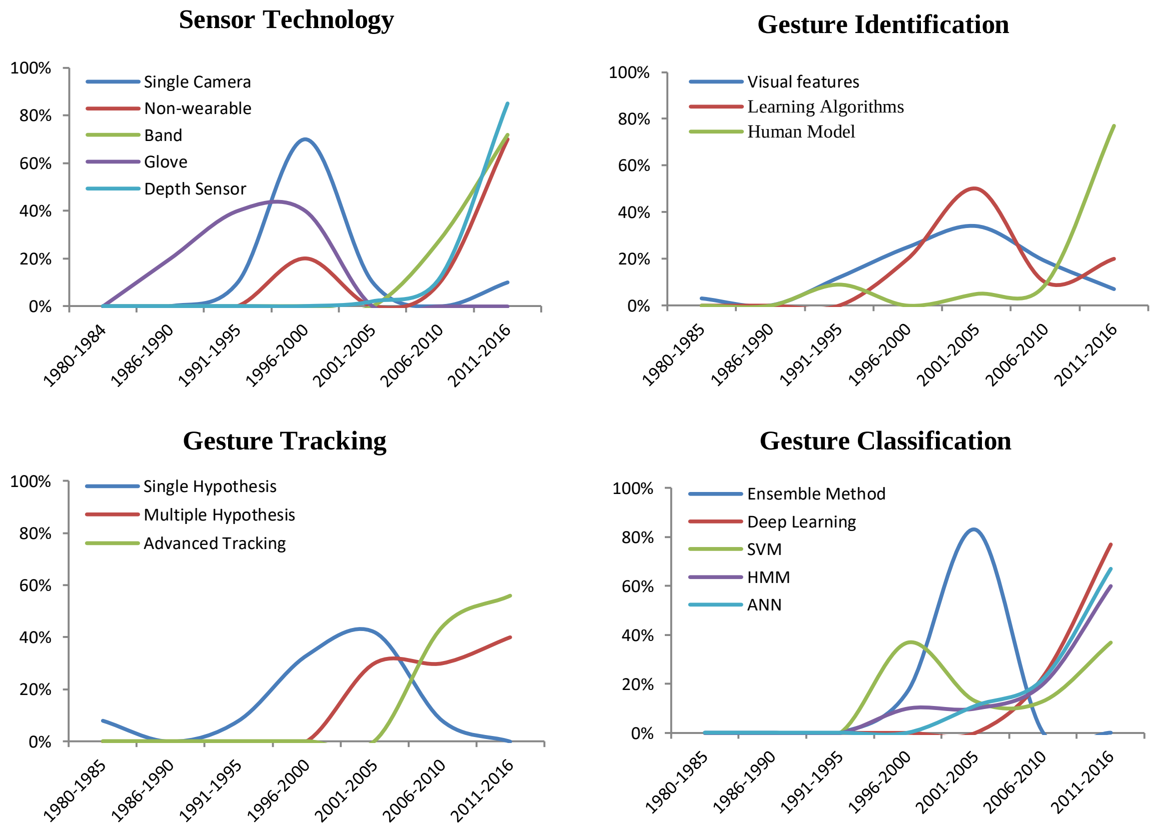

Fig. 15. Statistical analysis of papers in four gesture recognition technical components.

- Multi-sensors gesture recognition system: all the sensors have advantages and disadvantages. For instance, band sensor has large sensing area; Kinect has good performance in body gesture recognition. To best utilise the system performance, different gesture recognition sensors can be used in the same system.
- Algorithms combination approach: similar with sensors, different gesture classification algorithms also have their advantages and disadvantages. As we mentioned in gesture classification section, combination of algorithms improves efficiency.

References

[1] J. Krüger, T. Lien, A. Verl, Cooperation of human and machines in assembly lines, CIRP Annals-Manufacturing Technology, 58 (2009) 628-646.

[2] S.A. Green, M. Billinghurst, X. Chen, G. Chase, Human-robot collaboration: A literature review and augmented reality approach in design, International Journal of Advanced Robotic Systems, (2008) 1-18.
[3] P.R. Cohen, H.J. Levesque, Teamwork, Nous, (1991) 487-512.
[4] L.S. Vygotsky, Mind in society: The development of higher psychological processes, Harvard university press, 1980.
[5] P.R. Cohen, H.J. Levesque, Persistence, intention, and commitment, Reasoning about actions and plans, (1990) 297-340.
[6] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, D. Mulanda, Humanoid robots as cooperative partners for people, Int. Journal of Humanoid Robots, 1 (2004) 1-34.
[7] A. Bauer, D. Wollherr, M. Buss, Human–robot collaboration: a survey, International Journal of Humanoid Robotics, 5 (2008) 47-66.
[8] S. Mitra, T. Acharya, Gesture recognition: A survey, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 37 (2007) 311-324.
[9] R. Parasuraman, T.B. Sheridan, C.D. Wickens, A model for types and levels of human interaction with automation, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 30 (2000) 286-297.
[10] T.E. Starner, Visual Recognition of American Sign Language Using Hidden Markov Models, (1995).
[11] T. Starner, J. Weaver, A. Pentland, Real-time american sign language recognition using desk and wearable computer based video, in: Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1998, pp. 1371-1375.
[12] N.R. Howe, M.E. Leventon, W.T. Freeman, Bayesian Reconstruction of 3D Human Motion from Single-Camera Video, in: NIPS, 1999, pp. 820-826.
[13] Y. Katsuki, Y. Yamakawa, M. Ishikawa, High-speed Human/Robot Hand Interaction System, in: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, ACM, 2015, pp. 117-118.
[14] M. Elmezain, A. Al-Hamadi, J. Appenrodt, B. Michaelis, A hidden markov model-based continuous gesture recognition system for hand motion trajectory, in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE, 2008, pp. 1-4.
[15] Y. Matsumoto, A. Zelinsky, An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement, in: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, IEEE, 2000, pp. 499-504.
[16] J.P. Wachs, M. Kölsch, H. Stern, Y. Edan, Vision-based hand-gesture applications, Communications of the ACM, 54 (2011) 60-71.
[17] J. Suarez, R.R. Murphy, Hand gesture recognition with depth images: A review, in: RO-MAN, 2012 IEEE, IEEE, 2012, pp. 411-417.
[18] P. Doliotis, A. Stefan, C. McMurrough, D. Eckhard, V. Athitsos, Comparing gesture recognition accuracy using color and depth information, in: Proceedings of the 4th international conference on PErvasive technologies related to assistive environments, ACM, 2011, pp. 20.
[19] M. Hansard, S. Lee, O. Choi, R.P. Horaud, Time-of-flight cameras: principles, methods and applications, Springer Science & Business Media, 2012.
[20] T. Kapuściński, M. Oszust, M. Wysocki, Hand gesture recognition using time-of-flight camera and viewpoint feature histogram, in: Intelligent Systems in Technical and Medical Diagnostics, Springer, 2014, pp. 403-414.
[21] S.B. Gokturk, H. Yalcin, C. Bamji, A time-of-flight depth sensor-system description, issues and solutions, in: Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on, IEEE, 2004, pp. 35-35.
[22] J.D. Arango Paredes, B. Munoz, W. Agredo, Y. Ariza-Araujo, J.L. Orozco, A. Navarro, A reliability assessment software using Kinect to complement the clinical evaluation of Parkinson's disease, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 6860-6863.
[23] F. Anderson, T. Grossman, J. Matejka, G. Fitzmaurice, YouMove: enhancing movement training with an augmented reality mirror, in: Proceedings of the 26th annual ACM symposium on User interface software and technology, ACM, 2013, pp. 311-320.
[24] S. Obdrzalek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, M. Pavel, Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population, in: Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE, IEEE, 2012, pp. 1188-1193.
[25] C. Wang, Z. Liu, S.-C. Chan, Superpixel-Based Hand Gesture Recognition With Kinect Depth Camera, Multimedia, IEEE Transactions on, 17 (2015) 29-39.
[26] A. Kurakin, Z. Zhang, Z. Liu, A real time system for dynamic hand gesture recognition with a depth sensor, in: Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, IEEE, 2012, pp. 1975-1979.
[27] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore, Real-time human pose recognition in parts from single depth images, Communications of the ACM, 56 (2013) 116-124.
[28] T. Sharp, C. Keskin, D. Robertson, J. Taylor, J. Shotton, D.K.C.R.I. Leichter, A.V.Y. Wei, D.F.P.K.E. Krupka, A. Fitzgibbon, S. Izadi, Accurate, Robust, and Flexible Real-time Hand Tracking, in: Proc. CHI, 2015.
[29] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, X. Twombly, Vision-based hand pose estimation: A review, Computer Vision and Image Understanding, 108 (2007) 52-73.
[30] T. Labs, Myo, in, 2015, pp. https://www.myo.com/.
[31] Y. Zhang, C. Harrison, Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition, in: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, ACM, 2015, pp. 167-173.
[32] N. Haroon, A.N. Malik, Multiple Hand Gesture Recognition using Surface EMG Signals, Journal of Biomedical Engineering and Medical Imaging, 3 (2016) 1.

[33] S. Roy, S. Ghosh, A. Barat, M. Chattopadhyay, D. Chowdhury, Real-time Implementation of Electromyography for Hand Gesture Detection Using Micro Accelerometer, in: Artificial Intelligence and Evolutionary Computations in Engineering Systems, Springer, 2016, pp. 357-364.
[34] Google, Project Soli, in, Google, 2015, pp. https://www.google.com/atap/project-soli/.
[35] J. Smith, T. White, C. Dodge, J. Paradiso, N. Gershenfeld, D. Allport, Electric field sensing for graphical interfaces, Computer Graphics and Applications, IEEE, 18 (1998) 54-60.
[36] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, F. Durand, Capturing the human figure through a wall, ACM Transactions on Graphics (TOG), 34 (2015) 219.
[37] F. Adib, D. Katabi, See through walls with WiFi!, ACM, 2013.
[38] F. Adib, Z. Kabelac, D. Katabi, R.C. Miller, 3d tracking via body radio reflections, in: Usenix NSDI, 2014.
[39] J. Letessier, F. Bérard, Visual tracking of bare fingers for interactive surfaces, in: Proceedings of the 17th annual ACM symposium on User interface software and technology, ACM, 2004, pp. 119-122.
[40] D. Weinland, R. Ronfard, E. Boyer, A survey of vision-based methods for action representation, segmentation and recognition, Computer Vision and Image Understanding, 115 (2011) 224-241.
[41] D.G. Lowe, Object recognition from local scale-invariant features, in: Computer vision, 1999. The proceedings of the seventh IEEE international conference on, Ieee, 1999, pp. 1150-1157.
[42] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, in: Computer vision–ECCV 2006, Springer, 2006, pp. 404-417.
[43] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to SIFT or SURF, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 2564-2571.
[44] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24 (2002) 509-522.
[45] B. Allen, B. Curless, Z. Popović, Articulated body deformation from range scan data, in: ACM Transactions on Graphics (TOG), ACM, 2002, pp. 612-619.
[46] R. Cutler, M. Turk, View-based interpretation of real-time optical flow for gesture recognition, in: fg, IEEE, 1998, pp. 416.
[47] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of optical flow techniques, International journal of computer vision, 12 (1994) 43-77.
[48] C. Thurau, V. Hlaváč, Pose primitive based human action recognition in videos or still images, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1-8.
[49] Q. Pu, S. Gupta, S. Gollakota, S. Patel, Whole-home gesture recognition using wireless signals, in: Proceedings of the 19th annual international conference on Mobile computing & networking, ACM, 2013, pp. 27-38.
[50] R. Ronfard, C. Schmid, B. Triggs, Learning to parse pictures of people, in: Computer Vision—ECCV 2002, Springer, 2002, pp. 700-714.
[51] S.-J. Lee, C.-S. Ouyang, S.-H. Du, A neuro-fuzzy approach for segmentation of human objects in image sequences, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 33 (2003) 420-437.
[52] D. Tang, H.J. Chang, A. Tejani, T.-K. Kim, Latent regression forest: Structured estimation of 3d articulated hand posture, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, 2014, pp. 3786-3793.
[53] J. Han, L. Shao, D. Xu, J. Shotton, Enhanced computer vision with microsoft kinect sensor: A review, Cybernetics, IEEE Transactions on, 43 (2013) 1318-1334.
[54] Y. Yao, Y. Fu, Real-time hand pose estimation from RGB-D sensor, in: Multimedia and Expo (ICME), 2012 IEEE International Conference on, IEEE, 2012, pp. 705-710.
[55] I. Oikonomidis, N. Kyriazis, A.A. Argyros, Efficient model-based 3D tracking of hand articulations using Kinect, in: BMVC, 2011, pp. 3.
[56] J. Taylor, J. Shotton, T. Sharp, A. Fitzgibbon, The vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 103-110.
[57] Y. Li, Hand gesture recognition using Kinect, in: Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on, IEEE, 2012, pp. 196-199.
[58] M. Ye, X. Wang, R. Yang, L. Ren, M. Pollefeys, Accurate 3d pose estimation from a single depth image, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 731-738.
[59] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz, K.-q. Wang, Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors, in: Proceedings of the 14th international conference on Intelligent user interfaces, ACM, 2009, pp. 401-406.
[60] D. Comaniciu, V. Ramesh, P. Meer, Real-time tracking of non-rigid objects using mean shift, in: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, IEEE, 2000, pp. 142-149.
[61] S. Thrun, W. Burgard, D. Fox, Probabilistic robotics, MIT press, 2005.
[62] R.E. Kalman, A new approach to linear filtering and prediction problems, Journal of Fluids Engineering, 82 (1960) 35-45.
[63] S. Haykin, Kalman filtering and neural networks, John Wiley & Sons, 2004.
[64] E. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in: Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000, IEEE, 2000, pp. 153-158.
[65] K. Okuma, A. Taleghani, N. De Freitas, J.J. Little, D.G. Lowe, A boosted particle filter: Multitarget detection and tracking, in: Computer Vision-ECCV 2004, Springer, 2004, pp. 28-39.
[66] S. Oron, A. Bar-Hillel, D. Levi, S. Avidan, Locally orderless tracking, in: Computer Vision and Pattern

Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 1940-1947.
[67] J. Kwon, K.M. Lee, Tracking by sampling trackers, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 1195-1202.
[68] J. Kwon, K.M. Lee, F.C. Park, Visual tracking via geometric particle filtering on the affine group with optimal importance functions, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 991-998.
[69] R. Gao, L. Wang, R. Teti, D. Dornfeld, S. Kumara, M. Mori, M. Helu, Cloud-enabled prognosis for manufacturing, CIRP Annals-Manufacturing Technology, 64 (2015) 749-772.
[70] T. Li, S. Sun, T.P. Sattar, J.M. Corchado, Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches, Expert Systems with applications, 41 (2014) 3944-3954.
[71] T. Li, T.P. Sattar, S. Sun, Deterministic resampling: unbiased sampling to avoid sample impoverishment in particle filters, Signal Processing, 92 (2012) 1637-1645.
[72] J.M.D. Rincón, D. Makris, C.O. Uruňuela, J.-C. Nebel, Tracking human position and lower body parts using Kalman and particle filters constrained by human biomechanics, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 41 (2011) 26-37.
[73] D.A. Ross, J. Lim, R.-S. Lin, M.-H. Yang, Incremental learning for robust visual tracking, International Journal of Computer Vision, 77 (2008) 125-141.
[74] Z. Kalal, J. Matas, K. Mikolajczyk, Pn learning: Bootstrapping binary classifiers by structural constraints, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 49-56.
[75] B. Babenko, M.-H. Yang, S. Belongie, Visual tracking with online multiple instance learning, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 983-990.
[76] A.W. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah, Visual tracking: an experimental survey, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 36 (2014) 1442-1468.
[77] L.E. Peterson, K-nearest neighbor, Scholarpedia, 4 (2009) 1883.
[78] A.D. Wilson, A.F. Bobick, Hidden Markov models for modeling and recognizing gesture under variation, International Journal of Pattern Recognition and Artificial Intelligence, 15 (2001) 123-160.
[79] A.D. Wilson, A.F. Bobick, Parametric hidden markov models for gesture recognition, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21 (1999) 884-900.
[80] S. Lu, J. Picone, S. Kong, Fingerspelling Alphabet Recognition Using A Two-level Hidden Markov Model, in: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013, pp. 1.
[81] J. McCormick, K. Vincs, S. Nahavandi, D. Creighton, S. Hutchison, Teaching a Digital Performing Agent: Artificial Neural Network and Hidden Markov Model for recognising and performing dance movement, in: Proceedings of the 2014 International Workshop on Movement and Computing, ACM, 2014, pp. 70.
[82] S.-Z. Yu, Hidden semi-Markov models, Artificial Intelligence, 174 (2010) 215-243.
[83] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77 (1989) 257-286.
[84] M.A. Hearst, S.T. Dumais, E. Osman, J. Platt, B. Scholkopf, Support vector machines, Intelligent Systems and their Applications, IEEE, 13 (1998) 18-28.
[85] B. Schölkopf, A. Smola, Support Vector Machines, Encyclopedia of Biostatistics, (1998).
[86] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, The journal of machine learning research, 1 (2001) 211-244.
[87] B. Schiilkopf, The kernel trick for distances, in: Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference, MIT Press, 2001, pp. 301.
[88] A. Cenedese, G.A. Susto, G. Belgioioso, G.I. Cirillo, F. Fraccaroli, Home Automation Oriented Gesture Classification From Inertial Measurements, Automation Science and Engineering, IEEE Transactions on, 12 (2015) 1200-1210.
[89] K.-p. Feng, F. Yuan, Static hand gesture recognition based on HOG characters and support vector machines, in: Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd International Symposium on, IEEE, 2013, pp. 936-938.
[90] D. Ghimire, J. Lee, Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines, Sensors, 13 (2013) 7714-7734.
[91] O. Patsadu, C. Nukoolkit, B. Watanapa, Human gesture recognition using Kinect camera, in: Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on, IEEE, 2012, pp. 28-32.
[92] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of computer and system sciences, 55 (1997) 119-139.
[93] R.E. Schapire, The boosting approach to machine learning: An overview, in: Nonlinear estimation and classification, Springer, 2003, pp. 149-171.
[94] P. Viola, M. Jones, Robust real-time object detection, International Journal of Computer Vision, 4 (2001) 51-52.
[95] A.S. Micilotta, E.-J. Ong, R. Bowden, Detection and Tracking of Humans by Probabilistic Body Part Assembly, in: BMVC, 2005.
[96] S. Celebi, A.S. Aydin, T.T. Temiz, T. Arici, Gesture Recognition using Skeleton Data with Weighted Dynamic Time Warping, in: VISAPP (1), 2013, pp. 620-625.
[97] T. Arici, S. Celebi, A.S. Aydin, T.T. Temiz, Robust gesture recognition using feature pre-processing and weighted dynamic time warping, Multimedia Tools and Applications, 72 (2014) 3045-3062.
[98] E.J. Keogh, M.J. Pazzani, Derivative Dynamic Time Warping, in: SDM, SIAM, 2001, pp. 5-7.

[99] S.S. Rautaray, A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, Artificial Intelligence Review, 43 (2015) 1-54.
[100] S.S. Haykin, S.S. Haykin, S.S. Haykin, S.S. Haykin, Neural networks and learning machines, Pearson Education Upper Saddle River, 2009.
[101] T.H.H. Maung, Real-time hand tracking and gesture recognition system using neural networks, World Academy of Science, Engineering and Technology, 50 (2009) 466-470.
[102] H. Hasan, S. Abdul-Kareem, Static hand gesture recognition using neural networks, Artificial Intelligence Review, 41 (2014) 147-181.
[103] T. D'Orazio, G. Attolico, G. Cicirelli, C. Guaragnella, A Neural Network Approach for Human Gesture Recognition with a Kinect Sensor, in: ICPRAM, 2014, pp. 741-746.
[104] A. El-Baz, A. Tolba, An efficient algorithm for 3D hand gesture recognition using combined neural classifiers, Neural Computing and Applications, 22 (2013) 1477-1484.
[105] K. Subramanian, S. Suresh, Human action recognition using meta-cognitive neuro-fuzzy inference system, International journal of neural systems, 22 (2012) 1250028.
[106] Z.-H. Zhou, J. Wu, W. Tang, Ensembling neural networks: many could be better than all, Artificial intelligence, 137 (2002) 239-263.
[107] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature, 521 (2015) 436-444.
[108] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks, 61 (2015) 85-117.
[109] J. Tompson, M. Stein, Y. Lecun, K. Perlin, Real-time continuous pose recovery of human hands using convolutional networks, ACM Transactions on Graphics (TOG), 33 (2014) 169.
[110] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, in: Advances in Neural Information Processing Systems, 2014, pp. 568-576.
[111] J. Nagi, F. Ducatelle, G. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, L.M. Gambardella, Max-pooling convolutional neural networks for vision-based hand gesture recognition, in: Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on, IEEE, 2011, pp. 342-347.
[112] A. Jain, J. Tompson, Y. LeCun, C. Bregler, Modeep: A deep learning framework using motion features for human pose estimation, in: Computer Vision--ACCV 2014, Springer, 2014, pp. 302-315.
[113] S. Bilal, R. Akmeliawati, A.A. Shafie, M.J.E. Salami, Hidden Markov model for human to computer interaction: a study on human hand gesture recognition, Artificial Intelligence Review, 40 (2013) 495-516.
[114] M.K. Brown, L.R. Rabiner, Dynamic time warping for isolated word recognition based on ordered graph searching techniques, in: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82., IEEE, 1982, pp. 1255-1258.
[115] C.A. Ratanamahatana, E. Keogh, Everything you know about dynamic time warping is wrong, in: Third Workshop on Mining Temporal and Sequential Data, Citeseer, 2004.
[116] J.V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, Journal of clinical epidemiology, 49 (1996) 1225-1231.