International Journal of Industrial Ergonomics xxx (2017) 1–13

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Gesture recognition for human-robot collaboration: A review

Hongyi Liu, Lihui Wang*

KTH Royal Institute of Technology, Department of Production Engineering, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 20 July 2016
Received in revised form
15 November 2016
Accepted 15 February 2017
Available online xxx

Keywords:
Human-robot collaboration
Gesture
Gesture recognition

ABSTRACT

Recently, the concept of human-robot collaboration has raised many research interests. Instead of robots replacing human workers in workplaces, human-robot collaboration allows human workers and robots working together in a shared manufacturing environment. Human-robot collaboration can release human workers from heavy tasks with assistive robots if effective communication channels between humans 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 collaborative manufacturing. 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 also presented after technical analysis. Towards the end of this paper, future research trends are outlined.

1. Introduction

1.1. Human-robot collaboration

Robotic systems have already become essential components in various industrial sectors. Recently, the concept of Human-Robot Collaboration (HRC) has generated more interests. Literature suggested that human workers have unique problem-solving skills and sensory-motor capabilities, but are restricted in force and precision (Krüger et al., 2009; Green et al., 2008). Robotic systems, on the other hand, provide better fatigue, higher speed, higher repeatability and better productivity, but are restricted in flexibility. Jointly, HRC can release human workers from heavy tasks by establishing communication channels between humans and robots for better overall performance.

Ideally, an HRC team should work similarly as a human-human collaborative team in a manufacturing environment. However, time-separation or space-separation is dominant in HRC systems, which reduced productivity for both human workers and robots (Krüger et al., 2009). To build an efficient HRC team, human-human collaboration can be analysed as an example. In human teamwork and collaboration, there are two theories: joint intention theory and situated learning theory (Cohen and Levesque, 1990, 1991; Vygotsky, 1980; Breazeal et al., 2004). To apply the theories in an HRC team, three experiences will benefit the HRC manufacturing 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 forms of communication.

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

1.2. Gesture recognition

Gesture is one type of communication methods. Head nodding, hand gestures and body postures are effective communication channels in human-human collaboration (Green et al., 2008; Bauer et al., 2008). Gestures can be categorised into three types (Mitra and Acharya, 2007):

* Corresponding author.
E-mail addresses: hongyil@kth.se (H. Liu), lihuiw@kth.se (L. Wang).

http://dx.doi.org/10.1016/j.ergon.2017.02.004
0169-8141/

- 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. To collaborate with human workers, robots need to understand human gestures correctly and act based on the gestures efficiently. In HRC manufacturing environment, a natural form of gesture communication between humans and robots should be made available.

1.3. Gesture recognition for human-robot collaboration

To recognise gestures in the HRC manufacturing context, it is beneficial to investigate into a generic and simplified human information processing model. As shown in Fig. 1, Parasuraman et al. (2000) generalised human information processing into a four-stage model. Based on this generic model, 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, explained as follows.

- Sensor data collection: the raw data of a gesture 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.

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

2.1. Image based approaches

Technologies are often inspired by nature. As a 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.

Fig. 1. A four-stage model of human information processing (Parasuraman et al., 2000).

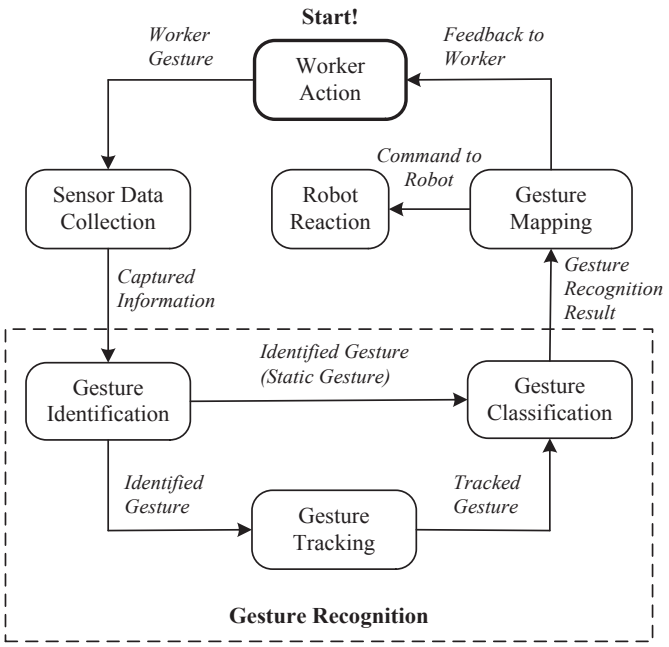

Fig. 2. A process model of gesture recognition for human-robot collaboration.

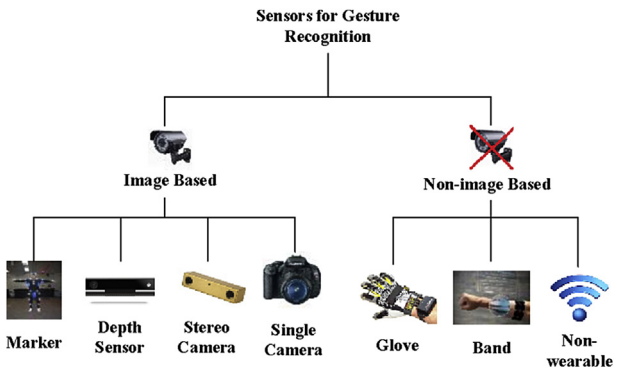

Fig. 3. Different types of gesture recognition sensors.

2.1.1. Marker

In marker-based approaches, a sensor is a conventional optical camera. In most marker-based solutions, users need to wear obvious markers (Mitra and Acharya, 2007). Today, we enjoy much faster graphical processing speed compared with twenty years ago. As a result, more gesture recognition sensors are available on the market.

2.1.2. Single camera

In the early 90th, researchers started to analyse gestures using a single camera (Starner, 1995; Starner et al., 1998). A drawback of single-camera-based approaches is the restriction of view angles, which affects a system's robustness (Howe et al., 1999). Recent research, however, applied a single camera in high-speed gesture recognition (Katsuki et al., 2015). The system utilises the speed image sensor and specially designed visual computing processor to achieve high-speed gesture recognition.

2.1.3. Stereo camera

To achieve robust gesture recognition, researchers suggested stereo camera based approaches to construct 3D environment. They have been applied in applications that use two stereo cameras

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

to construct 3D depth information. Many stereo camera based approaches follow a similar workflow (Elmezain et al., 2008; Matsumoto and Zelinsky, 2000). Although stereo camera systems have improved robustness in outdoor environment, they still suffered from problems such as computational complexity and calibration difficulties (Wachs et al., 2011).

2.1.4. Depth sensor

Recently, depth sensing technologies have emerged rapidly. We define a depth sensor as a non-stereo depth sensing device. Non-stereo depth sensor enjoys several advantages compared to the traditional stereo cameras. For example, the problems of setup calibration and illumination conditions can be prevented (Suarez and Murphy, 2012). 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 (Mitra and Acharya, 2007). A comparison of gesture identification accuracy by using colour and depth information can be found in (Doliotis et al., 2011). Time-of-Flight (ToF) technology is one of the popular depth sensing techniques. The fundamental principle of the ToF technology is to identify light travel time (Hansard et al., 2012). Recently, Microsoft Kinect 2 has applied the ToF technology. The advantage of the ToF technology is the higher frame rate. The limitation of the ToF technology is that the camera resolution highly depends on its light power and reflection (Gokturk et al., 2004).

Depth sensor provides a cheap and easy solution for gesture recognition. It is widely used in entertainment, education, and research, which has introduced a large developer community (Arango Paredes et al., 2015; Anderson et al., 2013; Obdrzalek et al., 2012; Kapuściński et al., 2014). With a large developer community, many open source tools and projects are available. Due to resolution restriction, currently, depth sensors are especially popular in body gesture recognition and close-distance hand and arm gesture recognition (Kapuściński et al., 2014; Wang et al., 2015; Kurakin et al., 2012; Shotton et al., 2013).

2.2. Non-image based approaches

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

2.2.1. 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 a load of cables can potentially cause problems in HRC manufacturing environment (Mitra and Acharya, 2007; Sharp et al., 2015). Glove-based approaches also introduced complex calibration and setup procedures (Erol et al., 2007).

2.2.2. Band

Another contactless technology uses band-based sensors. Band-based sensors rely on a wristband or similar wearable devices. Band-based sensors adopt wireless technology and electromyogram sensors, which avoid connecting cables. The sensors only need to contact with wrist; user's hand and fingers are released. One example is Myo gesture control armband (Labs, 2015). Recently, several band-based sensor gesture control systems have been reported (Zhang and Harrison, 2015; Haroon and Malik, 2016; Roy et al., 2016).

2.2.3. 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 (RF) signal based hand gesture tracking and recognition system (Google, 2015). As shown in Fig. 4(a), the device has an RF signal sender and a receiver. It 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 (Smith et al., 1998). A recent discovery from MIT introduced WiTrack and RF-Capture system that captures user motion by radio frequency signals reflected from human body (Adib et al., 2014, 2015; Adib and Katabi, 2013). As shown in Fig. 4(b), the RF-Capture system selects particular RF signals that can traverse through walls and reflect off the human body. The system can capture human motion even from another room with a precision of 20 cm. Although the precision is not acceptable in HRC manufacturing, 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, summarising the advantages and disadvantages of different technologies. It is clear that there is no sensor that fits all HRC applications. Two observations of the sensor technologies are provided based on the above analyses:

- In indoor HRC manufacturing environment, depth sensors are the most promising image-based techniques. Depth sensors possess advantages of easy calibration and accurate data processing. A large application developer community exists, which provides immediate solutions.
- Non-wearable approaches are the most promising technology among non-image based approaches. They can avoid direct contact with users, which provide advantages in an HRC manufacturing environment. Non-wearable sensing is also a fast-growing field.

3. Gesture identification

Gesture identification is the first step in gesture recognition 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 skeleton models.

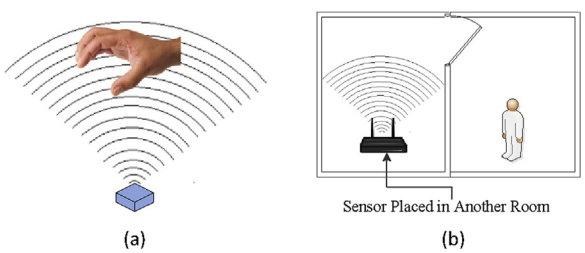

Fig. 4. Project Soli and RF-Capture system: (a) concept of Project Soli; (b) concept of RF-Capture gesture capturing system.

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

Table 1
Advantages and disadvantages of different sensor technologies.

	Advantages	Disadvantages
Marker	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 m
Glove	Fast response, precise tracking	Cumbersome device with a 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

3.1. 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.

3.1.1. 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 (Letessier and Bérard, 2004). Another common problem in skin colour detection is that human skin colour 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.

3.1.2. Local features

In image-based gesture recognition, illumination conditions significantly influence gesture identification quality. Therefore, many researchers have utilised the local features method that is not sensitive to lighting conditions. The local features approach is a detailed texture-based approach. It decomposes an image into smaller regions that are not corresponding to body parts (Weinland et al., 2011). As shown in Fig. 5, one of the most important local features is Scale Invariant Feature Transform (SIFT) (Lowe, 1999). Several similar local features approaches, for example, SURF and ORB were proposed in later years (Bay et al., 2006; Rublee et al., 2011). Normally, local features approaches are 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.

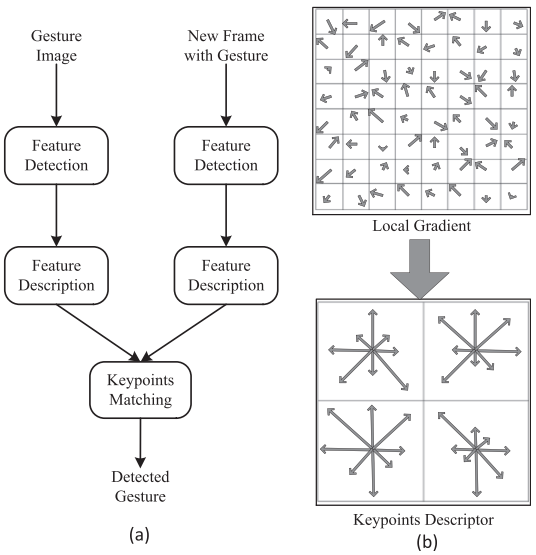

Fig. 5. SIFT algorithm: (a) SIFT algorithm for gesture identification; (b) SIFT feature description example (Lowe, 1999).

3.1.3. Shape and contour

Another intuitive and simple way to identify gestures is to utilise the unique shape and contour of a human body in HRC environment. Shape model based approach matches a pre-constructed shape model and shape features from observation. A milestone for shape detection and matching was reported by Belongie et al. (2002). 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 measure surface shapes. The 3D models generated from the technologies enable highly detailed representation of human body shape (Allen et al., 2002; Oikonomidis et al., 2011).

3.1.4. Motion

In certain HRC manufacturing environment, a human worker is the only moving object in the raw data. Therefore, the human 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 were implemented based on optical flow method (Cutler and Turk, 1998; Barron et al., 1994). Dalal and Thurau (Thurau and Hlaváč, 2008) introduced the well-known 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 (Adib and Katabi, 2013; Pu et al., 2013). 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 use 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 (Ronfard et al., 2002; Lee et al., 2003; Tang et al., 2014).

3.3. Skeleton model

To identify body gestures, a detailed model of the human body is

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

often useless. Different from the aforementioned approaches, skeleton model approach uses a human skeleton to discover human body poses (Taylor et al., 2012). As shown in Fig. 6, a skeleton model is a simplified human body model that preserves only the most valuable information from a human body. Skeleton model approach also provides advantages for simplifying gesture classification (Han et al., 2013). With the benefits mentioned above, the skeleton model approach has become an attractive solution for depth sensors (Han et al., 2013; Li, 2012).

3.4. Summary of gesture identification approaches

A gesture identification quality comparison case study was presented in the paper by Han (Han et al., 2013). The comparison result is shown in Table 2. It is easy to summarise that depth-based approach outperforms RGB-based approach. Skeleton model belongs to depth-based approach. Most of the visual features approaches belong to RGB-based approach. In Table 3, the advantages and disadvantages of different gesture identification methods are summarised. Moreover, according to different sensors, different gesture identification methods should be applied. In Table 4, the most suitable gesture identification method for each popular sensor is summarised. Due to the nature of HRC in manufacturing environment, human workers are the most important members of an HRC team. Despite understanding human body gestures, the skeleton model approach can also monitor human movements, which provides a secure environment for the HRC team. As mentioned earlier, skeleton model simplifies human body, while valuable information is well preserved. Therefore, subsequent gesture classification can be simplified by skeleton model approaches. Currently, skeleton model approach is an appropriate solution for gesture recognition in HRC manufacturing systems.

4. Gesture tracking

In gesture recognition, the notion of tracking is used differently in different literature. We define the notion of tracking as the process of finding temporal correspondences between frames. Specifically, we focus on the continuous gesture tracking problem

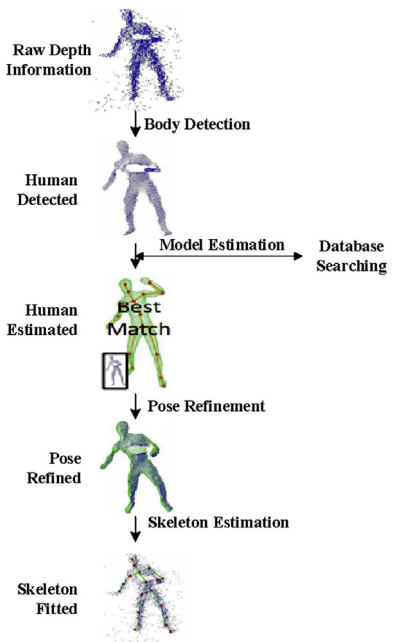

Fig. 6. Example of skeleton model identification (Ye et al., 2011).

Table 2
Gesture identification quality comparison from depth information and RGB information (Han et al., 2013).

	True positive	False positive
Depth information	96.7%	0%
RGB information	45.3%	2.3%

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. 7.

4.1. Single hypothesis tracking

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

4.1.1. Mean shift

Mean shift tracker is a basic tracking method. Mean shift tracker performs matching with RGB-colour histograms (Comaniciu et al., 2000). 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 (Comaniciu et al., 2000).

4.1.2. 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. 8. A complete KF mathematical derivation can be found in (Thrun et al., 2005; Kalman, 1960). Nowadays, KF has evolved and applied in different fields such as aerospace, robotics, and economics.

4.1.3. 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 (Haykin, 2004). Another algorithm that solves the same problem from a different angle is Unscented Kalman Filter (UKF) (Wan and Van Der Merwe, 2000). 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 HRC manufacturing scenarios, many human workers are working at the same station at the same time (Krüger et al., 2009). To track multiple workers' gestures simultaneously, multiple hypotheses tracking technologies should be applied.

4.2.1. Particle filter

Particle filter (PF) is a popular technology in robotics. Different from KF, PF does not make assumption on posterior model. The PF representation is a nonparametric approximation that can represent a broader space of distribution. Therefore, PF satisfies multiple hypotheses tracking requirements (Okuma et al., 2004). An example of PF is shown in Fig. 9. Several advanced tracking algorithms also apply PF to scan probability density function (Oron et al., 2012; Kwon and Lee, 2011; Kwon et al., 2009).

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

Table 3
Advantages and disadvantages of different gesture identification methods.

	Advantages	Disadvantages
Visual features	Low computational workload	Low quality
Learning algorithms	Background removal can be avoided	Higher computational cost
Skeleton model	Only the most important information is abstracted from a human body	Only possible to be used in depth sensor based systems

Table 4
Gesture identification methods for different sensors.

Sensor	Gesture identification method
Single camera	In single camera based systems, visual features method and learning algorithms can be implemented. To achieve robust performance, learning algorithms should be applied. To achieve faster image processing, visual features method should be applied (Katsuki et al., 2015).
Depth sensor	Since skeleton model method utilises and simplifies point cloud information, skeleton model method is a better option for depth sensor based systems (Han et al., 2013).
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 can be implemented in later gesture classification (Zhang and Harrison, 2015; Zhang et al., 2009).
Non-wearable	The non-wearable sensors also receive signal data instead of images. Due to the fact that RF signals contain noises (Adib and Katabi, 2013; Adib et al., 2014), advanced filtering and processing solutions need to be implemented in non-wearable sensor based systems.

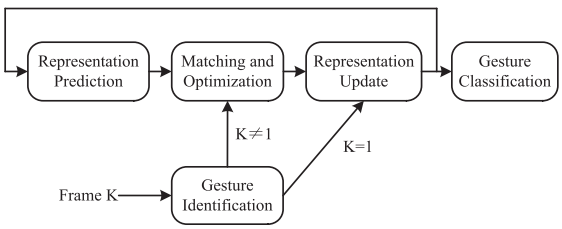

Fig. 7. Gesture tracking example.

4.2.2. Particle filter extensions

Many researchers have 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 (Li et al., 2014). Some other researchers also improved PF resampling strategy (Li et al., 2012; Rincón et al., 2011).

4.3. Advanced tracking methods

Recently, there have been 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.

4.3.1. 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 details for tracking (Ross et al., 2008). Kwon and Lee, (2011) presented a 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.

4.3.2. 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 (Kwon et al., 2009). One representative approach is Tracking, Learning and Detection Tracker (Kalal et al., 2010). The approach integrates the result of an object detector with an optical

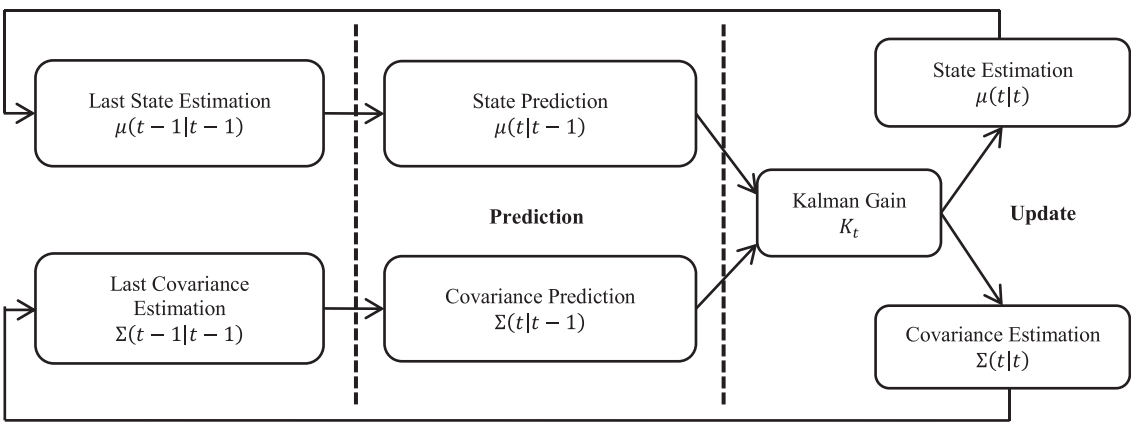

Fig. 8. Kalman filter process flow (Gao et al., 2015).

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

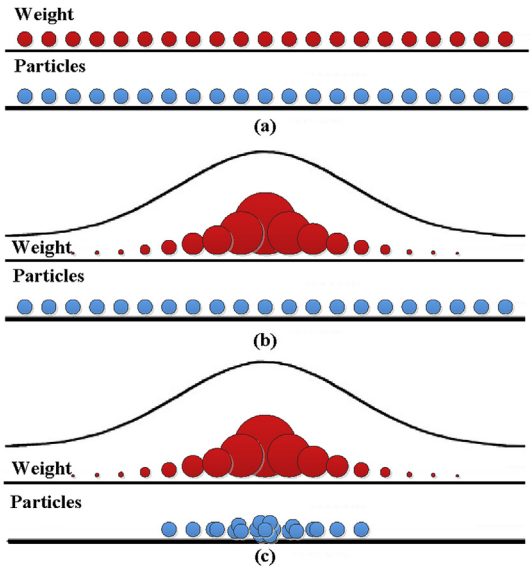

Fig. 9. Particles and weight factors: (a) after particles initialisation; (b) after weight factor calculation; (c) after resampling (Gao et al., 2015).

flow tracker. Another typical tracking-by-detection technology is to apply Multiple Instance Learning (Babenko et al., 2009). The learning algorithm can increase a tracker's robustness and decrease parameter tweaks.

4.4. Comparison of different gesture tracking approaches

Smeulders et al. (2014) presented a test result of different gesture tracking algorithms. The resulting score is normalised F-score. F-score provides an insight of the average coverage of the tracked object bounding box and the ground truth bounding box. Therefore, the tracking algorithms with higher F-score have better tracking quality. In Fig. 10, 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. However, the simple gesture tracking algorithms generate less computational load. Depending on computation power and tracking quality requirement, a gesture tracking algorithm can be selected for HRC manufacturing. A summary of different gesture tracking approaches is presented in Table 5.

5. Gesture classification

Gesture classification is the last but the most important step in gesture recognition. Being a typical machine learning problem, gesture classification can be solved by many popular machine learning algorithms.

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 (Peterson, 2009). Application of KNN in gesture classification can be found in (Peterson, 2009).

5.2. Hidden Markov Model

Hidden Markov Model (HMM) is a popular gesture classification algorithm. The HMM is a combination of an unobservable Markov chain and a stochastic process. An example of HMM is shown in Fig. 11, the unobservable 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* (Wilson and Bobick, 1999, 2001). To solve the problem, Expectation-Maximisation (EM) algorithm is applied (Wilson and Bobick, 1999). Many papers discussed HMM gesture recognition applications (Lu et al., 2013; McCormick et al., 2014; Yu, 2010). Some articles combined HMM with other classification approaches (McCormick et al., 2014). Others extended HMM algorithm into wider range of applications (Yu, 2010).

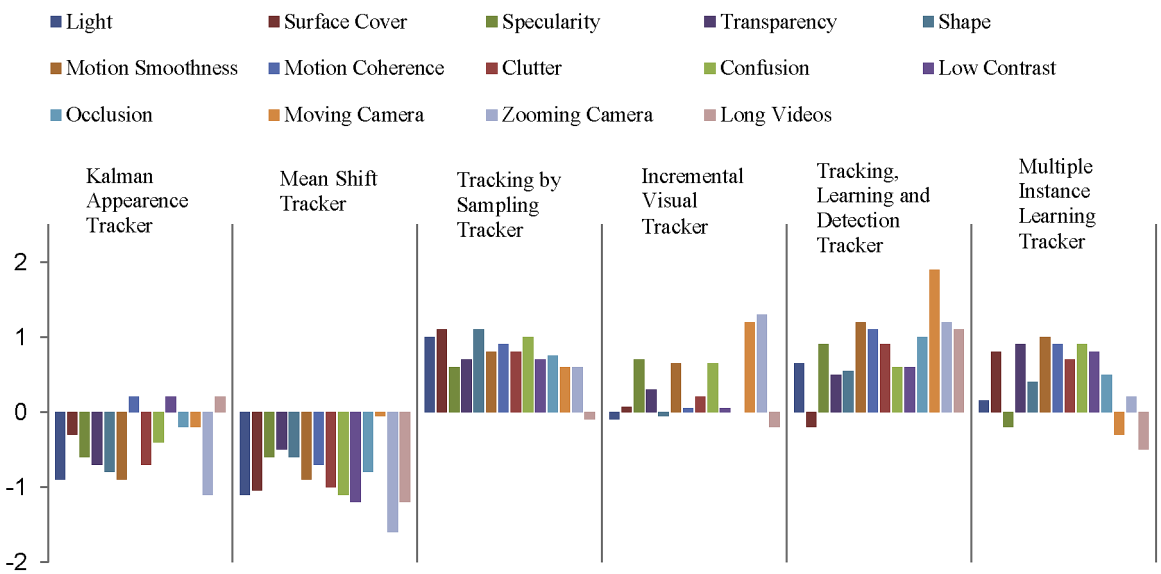

Fig. 10. Tracking algorithms test results in different video conditions (Smeulders et al., 2014).

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

Table 5
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 tasks.
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.

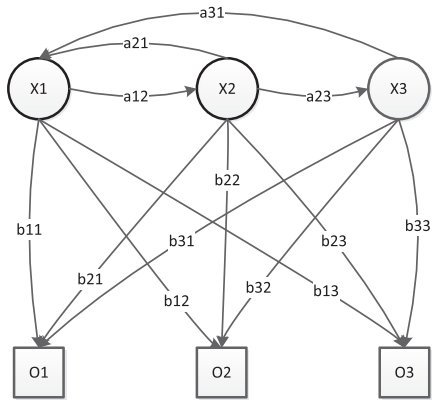

Fig. 11. Example of hidden Markov model (Rabiner, 1989).

5.3. Support Vector Machine

As shown in Fig. 12, Support Vector Machine (SVM) is a discriminative classifier defined by a separating hyperplane (Hearst et al., 1998; Schölkopf and Smola, 1998). 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 the training sets. Some researchers proposed Relevance Vector Machine (RVM) to solve the problem (Tipping, 2001). SVM kernel trick was introduced by Scholkopf (Schiilkopf, 2001). SVM kernel trick enables linear SVM in nonlinear problems. SVM kernel transforms low-dimensional training data into high-dimensional feature space with nonlinear method (Cenedese et al., 2015). There are also many papers that combined SVM with other classification methods to improve gesture classification performance (Feng and Yuan, 2013; Ghimire and Lee, 2013; Patsadu et al., 2012).

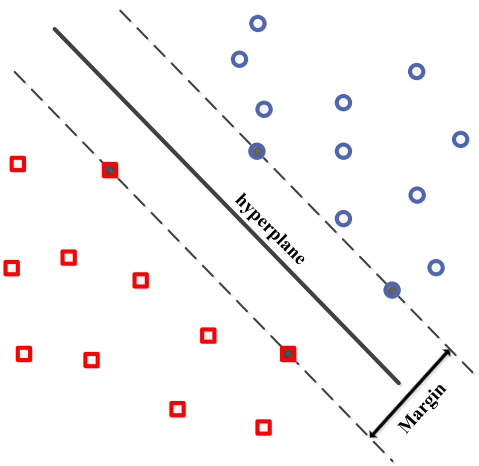

Fig. 12. Example of linear Support Vector Machine (Hearst et al., 1998; Schölkopf and Smola, 1998).

5.4. Ensemble method

Ensemble method is another type of widely-used gesture classification algorithm. The primary assumption of ensemble method is that ensembles are more accurate than individual weak classifiers. One of the famous ensemble methods is Boosting by Schapire et al. (Schapire, 2003; Freund and Schapire, 1997). The boosting algorithm starts with several weak classifiers. The weak classifiers are repeatedly applied. In a training iteration, a 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. 13, 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 (Viola and Jones, 2001; Micilotta et al., 2005). The classification performs better when AdaBoost algorithm is used.

5.5. Dynamic time warping

Dynamic time warping (DTW) is an optimal alignment algorithm for two sequences (Müller, 2007; Keogh and Pazzani, 2001). 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 have been many DTW applications in gesture recognition (Celebi et al., 2013; Arici et al., 2014). Some papers also introduced Derivative Dynamic Time Warping (DDTW) as an extension to DTW (Keogh and Pazzani, 2001; Rautaray and Agrawal, 2015).

5.6. Artificial Neural Network

Artificial Neural Network (ANN) is a family of information processing models inspired by biological neural networks (Haykin et al., 2009). ANN consists of many interconnected processing units (neurons) that work in parallel. Each 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, many researchers have utilised ANN for gesture recognition (Maung, 2009; Hasan and Abdul-Kareem, 2014; D'Orazio et al., 2014). Several papers also presented gesture recognition systems that combined ANN with other classification methods (El-Baz and Tolba, 2013; Subramanian and Suresh, 2012; Zhou et al., 2002).

5.7. Deep learning

Deep learning is an emerging and fast-growing branch of

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

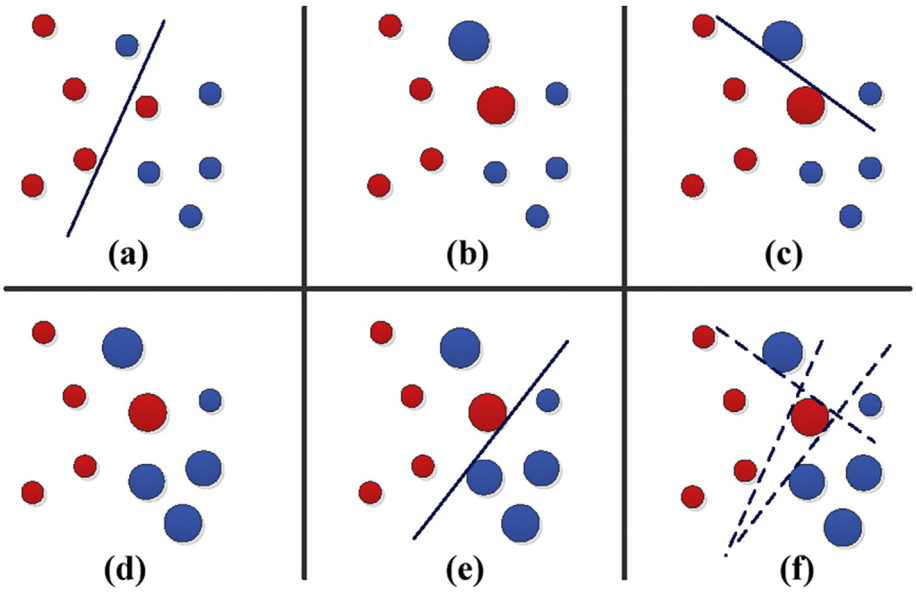

Fig. 13. Example of AdaBoost: (a) weak classifier 1, (b) weights increase, (c) weak classifier 2, (d) weights increase, (e) weak classifier 3, (f) final decision boundary (Schapire, 2003).

machine learning. Deep learning enables data modelling with high-level abstractions by using multiple processing layer neural networks. 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 (LeCun et al., 2015). Nowadays, deep learning is applied in image recognition, speech recognition, particle accelerator data analysis, etc. (Schmidhuber, 2015). 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 (Tompson et al., 2014; Simonyan and Zisserman, 2014). Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are two popular deep learning algorithms (LeCun et al., 2015). Several gesture recognition systems have applied above deep learning algorithms recently (Nagi et al., 2011; Jain et al., 2014).

5.8. Comparison of gesture classification approaches

Table 6 lists the advantages and disadvantages of the gesture classification approaches. One of the trends for HRC systems is to adopt deep learning. The primary constraint of deep learning is the limited computational power. However, the exponentially increasing computation power can solve the problem quickly. The number of deep learning based gesture classification applications is growing rapidly. Another trend is to combine different classification algorithms. Every classification algorithm has own advantages and disadvantages. To utilise advantages, different classifiers can be combined to achieve better performance in manufacturing environment. 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. 14, we selected 9 different journals and conferences that each published more than 5 papers within the 285 reviewed papers related to gesture recognition. Regarding the 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. 15 shows the yearly distributions of the reviewed gesture recognition papers. It is clear that the number of gesture recognition papers has increased rapidly since 1994, indicating the growing interests in this field.

Table 6
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 (Bilal et al., 2013)	Many free parameters need to be adjusted (Bilal et al., 2013)
Support Vector Machine	Different kernel function can be applied (Schiilkopf, 2001)	Number of support vectors grows linearly with the size of training set (Tipping, 2001)
Ensemble Method	Do not need large number of training data	Over fit easily, sensitive to noise and outliers
Dynamic Time Warping	Reliable nonlinear alignment between patterns (Brown and Rabiner, 1982)	Time and space complexity (Ratanamahatana and Keogh, 2004)
Artificial Neural Network	Can detect complex nonlinear relationships between variables (Tu, 1996)	"Black box" nature and cannot be used for small training data set (Tu, 1996)
Deep Learning	Do not need good design of features, outperform other machine learning methods (LeCun et al., 2015)	Need large number of training data and computationally expensive

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

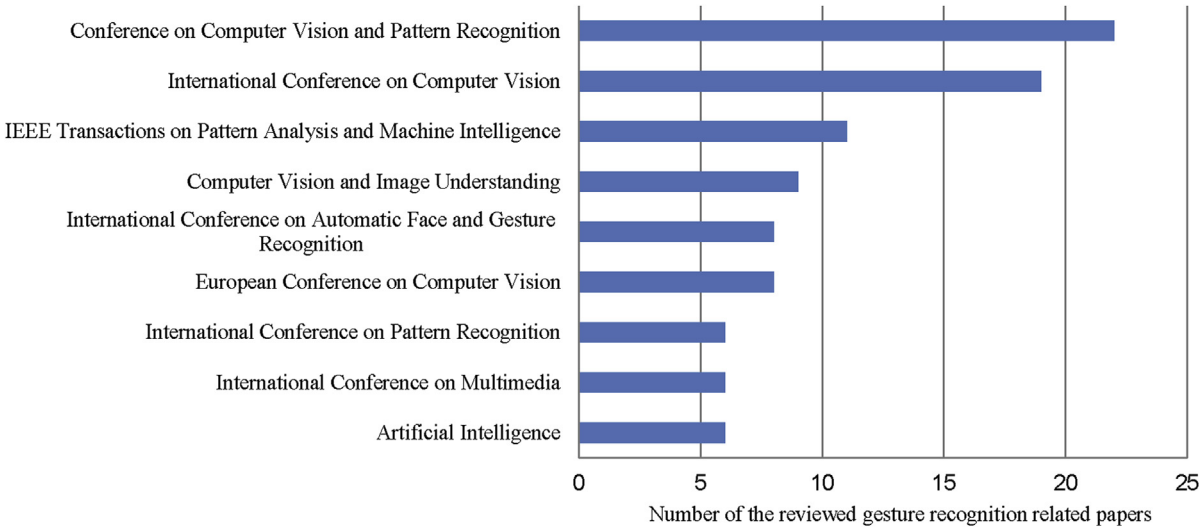

Fig. 14. Journals and conferences that publishing most gesture recognition related papers.

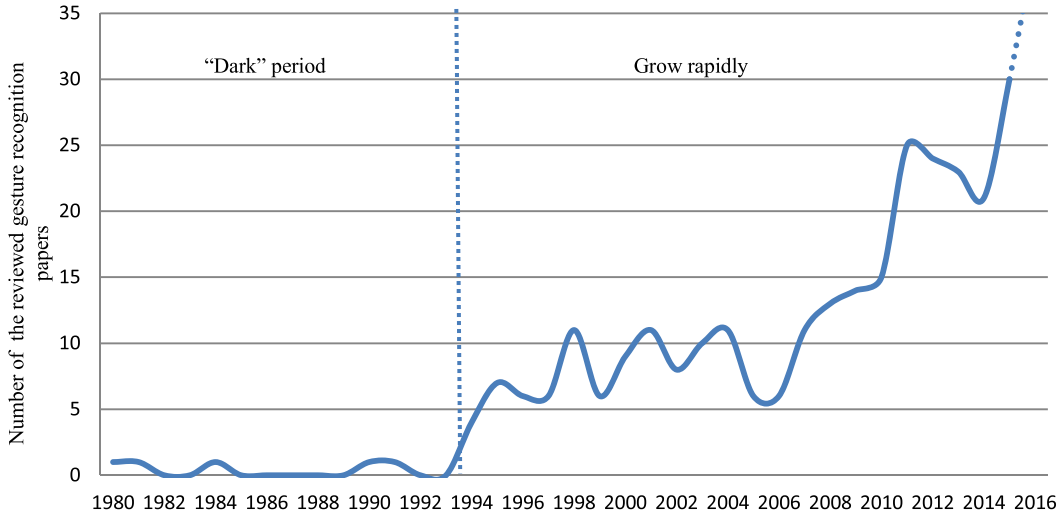

Fig. 15. Yearly distribution of the reviewed gesture recognition related papers.

Fig. 16 shows the yearly development trend of the four technical components in the field of gesture recognition. The horizontal axis represents the percentage of the papers cited in a particular period of time as compared with all the papers reviewed on this technology. It is clear to observe:

- Depth sensing is the rapidly developing technology among sensor technologies. Band sensors and non-wearable sensors are also growing quickly.
- Skeleton 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 methods such as HMM and SVM 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 usually start from hardware. Software technologies and algorithms are then designed to utilise the performance of hardware. Therefore, we would introduce some of the predicted future trends starting with sensor technologies.

- Depth sensor and skeleton model based gesture recognition: due to the nature of HRC in manufacturing environment, human workers are the most important members of any HRC team. Despite the understanding of human body gestures, depth sensors together with skeleton models will monitor human movements, which provide a safer environment for an HRC manufacturing system. Moreover, skeleton models will simplify gesture classification, making much simpler gesture tracking and classification methods applicable.

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

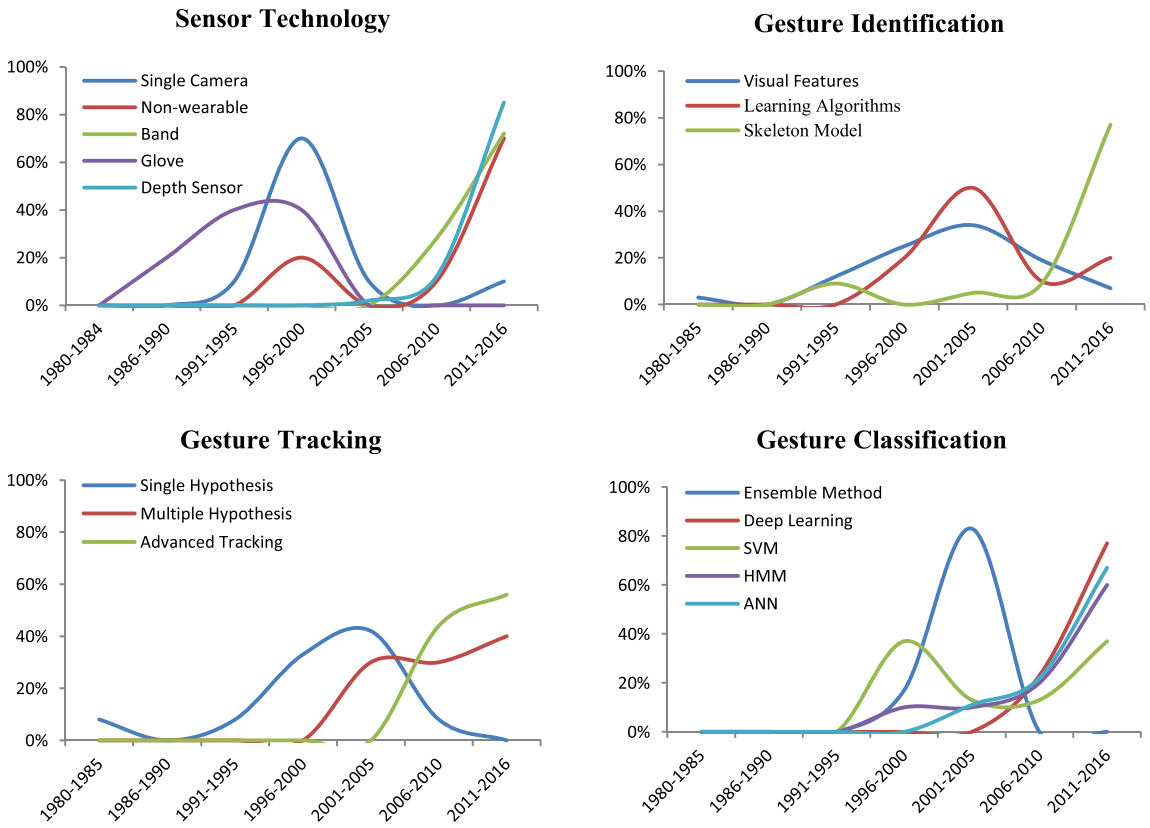

Fig. 16. Statistical analysis of papers in four gesture recognition technical components.

- Non-wearable sensor and deep learning based gesture recognition: although non-wearable sensor technologies are not ready, they are the most promising ones. In HRC manufacturing systems, human workers should be able to communicate with robots naturally. For this very purpose, nothing should be attached to workers' body. Non-wearable sensors today still suffer from low gesture identification and classification quality. This problem can potentially be solved by using the deep learning methods.
- Hard real-time gesture recognition system: one of the most important requirements of a manufacturing system is the real-time requirement. Especially, in an HRC manufacturing system, the safety of human workers is of paramount importance. Therefore, real-time gesture recognition is another future direction. Currently, band and glove sensors provide the fastest response. Moreover, high-speed single-camera gesture recognition is also emerging recently (Katsuki et al., 2015). In gesture identification, tracking and classification, quick and effective methods can be applied.
- 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 achieve the best system performance, different gesture recognition sensors can be used in the same system.
- Algorithms combination approach: similar to sensors, different gesture classification algorithms also have their advantages and disadvantages. As mentioned in the gesture classification section, appropriate combination of algorithms can improve efficiency.

References

Adib, F., Katabi, D., 2013. See through Walls with WiFi! ACM.
Adib, F., Kabelac, Z., Katabi, D., Miller, R.C., 2014. 3D tracking via body radio reflections. In: Usenix NSDI.
Adib, F., Hsu, C.-Y., Mao, H., Katabi, D., Durand, F., 2015. Capturing the Human Figure through a Wall, vol. 34. ACM Transactions on Graphics (TOG), p. 219.
Allen, B., Curless, B., Popović, Z., 2002. Articulated Body Deformation from Range Scan Data. In: ACM Transactions on Graphics (TOG). ACM, pp. 612–619.
Anderson, F., Grossman, T., Matejka, J., Fitzmaurice, G., 2013. YouMove: enhancing movement training with an augmented reality mirror. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 311–320.
Arango Paredes, J.D., Munoz, B., Agredo, W., Ariza-Araujo, Y., Orozco, J.L., Navarro, A., 2015. 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, pp. 6860–6863.
Arici, T., Celebi, S., Aydin, A.S., Temiz, T.T., 2014. Robust gesture recognition using feature pre-processing and weighted dynamic time warping. Multimed. Tools Appl. 72, 3045–3062.
Babenko, B., Yang, M.-H., Belongie, S., 2009. Visual tracking with online multiple instance learning. In: Computer Vision and Pattern Recognition, pp. 983–990. CVPR 2009. IEEE Conference on, IEEE, 2009.
Barron, J.L., Fleet, D.J., Beauchemin, S.S., 1994. Performance of optical flow techniques. Int. J. Comput. Vis. 12, 43–77.
Bauer, A., Wollherr, D., Buss, M., 2008. Human–robot collaboration: a survey. Int. J. Humanoid Robot. 5, 47–66.
Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: speeded up robust features. In: Computer Vision–ECCV 2006. Springer, pp. 404–417.
Belongie, S., Malik, J., Puzicha, J., 2002. Shape Matching and Object Recognition Using Shape Contexts, Pattern Analysis and Machine Intelligence. IEEE Transactions on, 24, pp. 509–522.
Bilal, S., Akmeliawati, R., Shafie, A.A., Salami, M.J.E., 2013. Hidden Markov model for human to computer interaction: a study on human hand gesture recognition. Artif. Intell. Rev. 40, 495–516.
Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A., Mulanda, D., 2004. Humanoid robots as cooperative partners for people. Int. J. Humanoid Robot. 1, 1–34.
Brown, M.K., Rabiner, L.R., 1982. Dynamic time warping for isolated word

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

recognition based on ordered graph searching techniques. In: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82. IEEE, pp. 1255–1258.
Celebi, S., Aydin, A.S., Temiz, T.T., Arici, T., 2013. Gesture recognition using skeleton data with weighted dynamic time warping. In: VISAPP, vol. 1, pp. 620–625.
Cenedese, A., Susto, G.A., Belgioioso, G., Cirillo, G.I., Fraccaroli, F., 2015. Home Automation Oriented Gesture Classification from Inertial Measurements, Automation Science and Engineering. IEEE Transactions on, 12, pp. 1200–1210.
Cohen, P.R., Levesque, H.J., 1990. Persistence, Intention, and Commitment, Reasoning about Actions and Plans, pp. 297–340.
Cohen, P.R., Levesque, H.J., 1991. Teamwork, Nous, pp. 487–512.
Comaniciu, D., Ramesh, V., Meer, P., 2000. Real-time tracking of non-rigid objects using mean shift. In: Computer Vision and Pattern Recognition, pp. 142–149. Proceedings. IEEE Conference on, IEEE, 2000.
Cutler, R., Turk, M., 1998. View-based Interpretation of Real-time Optical Flow for Gesture Recognition. in: fg. IEEE, p. 416.
D'Orazio, T., Attolico, G., Cicirelli, G., Guaragnella, C., 2014. A neural network approach for human gesture recognition with a kinect sensor. In: ICPRAM, pp. 741–746.
Doliotis, P., Stefan, A., McMurrough, C., Eckhard, D., Athitsos, V., 2011. Comparing gesture recognition accuracy using color and depth information. In: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, p. 20.
El-Baz, A., Tolba, A., 2013. An efficient algorithm for 3D hand gesture recognition using combined neural classifiers. Neural Comput. Appl. 22, 1477–1484.
Elmezain, M., Al-Hamadi, A., Appenrodt, J., Michaelis, B., 2008. A Hidden Markov Model-based Continuous Gesture Recognition System for Hand Motion Trajectory. In: Pattern Recognition, pp. 1–4. ICPR 2008. 19th International Conference on, IEEE, 2008.
Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X., 2007. Vision-based hand pose estimation: a review. Comput. Vis. Image Underst. 108, 52–73.
Feng, K.-p., Yuan, F., 2013. Static hand gesture recognition based on HOG characters and support vector machines. In: Instrumentation and Measurement, Sensor Network and Automation (IMSNA), pp. 936–938, 2nd International Symposium on, IEEE, 2013.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 119–139.
Gao, R., Wang, L., Teti, R., Dornfeld, D., Kumara, S., Mori, M., Helu, M., 2015. Cloud-enabled prognosis for manufacturing. CIRP Annals-Manufacturing Technol. 64, 749–772.
Ghimire, D., Lee, J., 2013. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors 13, 7714–7734.
Gokturk, S.B., Yalcin, H., Bamji, C., 2004. A time-of-flight depth sensor-system description, issues and solutions. In: Computer Vision and Pattern Recognition Workshop, p. 35. CVPRW'04. Conference on, IEEE, 2004.
Google, Project Soli, in, Google, 2015, pp. https://www.google.com/atap/project-soli/.
Green, S.A., Billinghurst, M., Chen, X., Chase, G., 2008. Human-robot collaboration: a literature review and augmented reality approach in design. Int. J. Adv. Robot. Syst. 1–18.
Han, J., Shao, L., Xu, D., Shotton, J., 2013. Enhanced Computer Vision with Microsoft Kinect Sensor: a Review, Cybernetics. IEEE Transactions on, 43, pp. 1318–1334.
Hansard, M., Lee, S., Choi, O., Horaud, R.P., 2012. Time-of-flight Cameras: Principles, Methods and Applications. Springer Science & Business Media.
Haroon, N., Malik, A.N., 2016. Multiple hand gesture recognition using surface EMG signals. J. Biomed. Eng. Med. Imaging 3, 1.
Hasan, H., Abdul-Kareem, S., 2014. Static hand gesture recognition using neural networks. Artif. Intell. Rev. 41, 147–181.
Haykin, S., 2004. Kalman Filtering and Neural Networks. John Wiley & Sons.
Haykin, S.S., Haykin, S.S., Haykin, S.S., Haykin, S.S., 2009. Neural Networks and Learning Machines. Pearson Education Upper Saddle River.
Hearst, M.A., Dumais, S.T., Osman, E., Platt, J., Scholkopf, B., 1998. Support Vector Machines, Intelligent Systems and Their Applications, vol. 13. IEEE, pp. 18–28.
Howe, N.R., Leventon, M.E., Freeman, W.T., 1999. Bayesian reconstruction of 3D human motion from single-camera video. In: NIPS, pp. 820–826.
Jain, A., Tompson, J., LeCun, Y., Bregler, C., 2014. Modeep: a deep learning framework using motion features for human pose estimation. In: Computer Vision–ACCV 2014. Springer, pp. 302–315.
Kalal, Z., Matas, J., Mikolajczyk, K., 2010. Pn learning: bootstrapping binary classifiers by structural constraints. In: Computer Vision and Pattern Recognition (CVPR), pp. 49–56. IEEE Conference on, IEEE, 2010.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Fluids Eng. 82, 35–45.
Kapuściński, T., Oszust, M., Wysocki, M., 2014. Hand gesture recognition using time-of-flight camera and viewpoint feature histogram. In: Intelligent Systems in Technical and Medical Diagnostics. Springer, pp. 403–414.
Katsuki, Y., Yamakawa, Y., Ishikawa, M., 2015. High-speed human/robot hand interaction system. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-robot Interaction Extended Abstracts. ACM, pp. 117–118.
Keogh, E.J., Pazzani, M.J., 2001. Derivative dynamic time warping. In: SDM. SIAM, pp. 5–7.
Krüger, J., Lien, T., Verl, A., 2009. Cooperation of human and machines in assembly lines. CIRP Annals-Manufacturing Technol. 58, 628–646.
Kurakin, A., Zhang, Z., Liu, Z., 2012. 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, pp. 1975–1979.
Kwon, J., Lee, K.M., 2011. Tracking by Sampling Trackers. In: Computer Vision (ICCV), pp. 1195–1202. IEEE International Conference on, IEEE, 2011.
Kwon, J., Lee, K.M., Park, F.C., 2009. Visual tracking via geometric particle filtering on the affine group with optimal importance functions. In: Computer Vision and Pattern Recognition, pp. 991–998. CVPR 2009. IEEE Conference on, IEEE, 2009.
T. Labs, Myo, in, 2015, pp. https://www.myo.com/.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lee, S.-J., Ouyang, C.-S., Du, S.-H., 2003. A Neuro-fuzzy Approach for Segmentation of Human Objects in Image Sequences, Systems, Man, and Cybernetics, Part B: Cybernetics. IEEE Transactions on, 33, pp. 420–437.
Letessier, J., Bérard, F., 2004. Visual tracking of bare fingers for interactive surfaces. In: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 119–122.
Li, Y., 2012. Hand gesture recognition using kinect. IEEE 3rd International Conference on, IEEE, 2012. In: Software Engineering and Service Science (ICSESS), pp. 196–199.
Li, T., Sattar, T.P., Sun, S., 2012. Deterministic resampling: unbiased sampling to avoid sample impoverishment in particle filters. Signal Process. 92, 1637–1645.
Li, T., Sun, S., Sattar, T.P., Corchado, J.M., 2014. Fight sample degeneracy and impoverishment in particle filters: a review of intelligent approaches. Expert Syst. Appl. 41, 3944–3954.
Lowe, D.G., 1999. Object recognition from local scale-invariant features. In: Computer Vision, pp. 1150–1157. The proceedings of the seventh IEEE international conference on, Ieee, 1999.
Lu, S., Picone, J., Kong, S., 2013. 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), p. 1.
Matsumoto, Y., Zelinsky, A., 2000. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In: Automatic Face and Gesture Recognition, pp. 499–504. Proceedings. Fourth IEEE International Conference on, IEEE, 2000.
Maung, T.H.H., 2009. Real-time hand tracking and gesture recognition system using neural networks, world academy of science. Eng. Technol. 50, 466–470.
McCormick, J., Vincs, K., Nahavandi, S., Creighton, D., Hutchison, S., 2014. 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, p. 70.
Micilotta, A.S., Ong, E.-J., Bowden, R., 2005. Detection and tracking of humans by probabilistic body part assembly. In: BMVC.
Mitra, S., Acharya, T., 2007. Gesture Recognition: a Survey, Systems, Man, and Cybernetics, Part C: Applications and Reviews. IEEE Transactions on, 37, pp. 311–324.
Müller, M., 2007. Dynamic Time Warping. Information Retrieval for Music and Motion, pp. 69–84.
Nagi, J., Ducatelle, F., Di Caro, G., Cireşan, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J., Gambardella, L.M., 2011. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In: Signal and Image Processing Applications (ICSIPA), pp. 342–347. IEEE International Conference on, IEEE, 2011.
Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., Pavel, M., 2012. 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, pp. 1188–1193.
Oikonomidis, I., Kyriazis, N., Argyros, A.A., 2011. Efficient model-based 3D tracking of hand articulations using kinect. In: BMVC, p. 3.
Okuma, K., Taleghani, A., De Freitas, N., Little, J.J., Lowe, D.G., 2004. A boosted particle filter: multitarget detection and tracking. In: Computer Vision-ECCV 2004. Springer, pp. 28–39.
Oron, S., Bar-Hillel, A., Levi, D., Avidan, S., 2012. Locally orderless tracking. In: Computer Vision and Pattern Recognition (CVPR), pp. 1940–1947. IEEE Conference on, IEEE, 2012.
Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2000. A Model for Types and Levels of Human Interaction with Automation, Systems, Man and Cybernetics, Part a: Systems and Humans. IEEE Transactions on, 30, pp. 286–297.
Patsadu, O., Nukoolkit, C., Watanapa, B., 2012. Human gesture recognition using Kinect camera. In: Computer Science and Software Engineering (JCSSE), pp. 28–32. International Joint Conference on, IEEE, 2012.
Peterson, L.E., 2009. K-nearest neighbor. Scholarpedia 4, 1883.
Pu, Q., Gupta, S., Gollakota, S., Patel, S., 2013. Whole-home gesture recognition using wireless signals. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking. ACM, pp. 27–38.
Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE, vol. 77, pp. 257–286.
Ratanamahatana, C.A., Keogh, E., 2004. Everything you know about dynamic time warping is wrong. In: Third Workshop on Mining Temporal and Sequential Data. Citeseer.
Rautaray, S.S., Agrawal, A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. Artif. Intell. Rev. 43, 1–54.
Rincón, J.M.D., Makris, D., Uruñuela, C.O., Nebel, J.-C., 2011. Tracking Human Position and Lower Body Parts Using Kalman and Particle Filters Constrained by Human

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004

Biomechanics, Systems, Man, and Cybernetics, Part B: Cybernetics. IEEE Transactions on, 41, pp. 26–37.
Ronfard, R., Schmid, C., Triggs, B., 2002. Learning to parse pictures of people. In: Computer Vision—ECCV 2002. Springer, pp. 700–714.
Ross, D.A., Lim, J., Lin, R.-S., Yang, M.-H., 2008. Incremental learning for robust visual tracking. Int. J. Comput. Vis. 77, 125–141.
Roy, S., Ghosh, S., Barat, A., Chattopadhyay, M., Chowdhury, D., 2016. Real-time implementation of electromyography for hand gesture detection using micro accelerometer. In: Artificial Intelligence and Evolutionary Computations in Engineering Systems. Springer, pp. 357–364.
Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: an efficient alternative to SIFT or SURF. In: Computer Vision (ICCV), pp. 2564–2571. IEEE International Conference on, IEEE, 2011.
Schapire, R.E., 2003. The boosting approach to machine learning: an overview. In: Nonlinear Estimation and Classification. Springer, pp. 149–171.
Schiilkopf, B., 2001. The kernel trick for distances. In: Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference. MIT Press, p. 301.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.
Schölkopf, B., Smola, A., 1998. Support Vector Machines, Encyclopedia of Biostatistics.
Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J., Leichter, D.K.C.R.I., Wei, A.V.Y., Krupka, D.F.P.K.E., Fitzgibbon, A., Izadi, S., 2015. Accurate, robust, and flexible real-time hand tracking. In: Proc. CHI.
Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R., 2013. Real-time human pose recognition in parts from single depth images. Commun. ACM 56, 116–124.
Simonyan, K., Zisserman, A., 2014. Two-stream convolutional networks for action recognition in videos. In: Advances in Neural Information Processing Systems, pp. 568–576.
Smeulders, A.W., Chu, D.M., Cucchiara, R., Calderara, S., Dehghan, A., Shah, M., 2014. Visual Tracking: an Experimental Survey, Pattern Analysis and Machine Intelligence. IEEE Transactions on, 36, pp. 1442–1468.
Smith, J., White, T., Dodge, C., Paradiso, J., Gershenfeld, N., Allport, D., 1998. Electric Field Sensing for Graphical Interfaces, Computer Graphics and Applications, vol. 18. IEEE, pp. 54–60.
Starner, T.E., 1995. Visual Recognition of American Sign Language Using Hidden Markov Models.
Starner, T., Weaver, J., Pentland, A., 1998. Real-time american sign language recognition using desk and wearable computer based video. In: Pattern Analysis and Machine Intelligence, pp. 1371–1375. IEEE Transactions on.
Suarez, J., Murphy, R.R., 2012. Hand gesture recognition with depth images: a review. In: Ro-man, 2012 IEEE. IEEE, pp. 411–417.
Subramanian, K., Suresh, S., 2012. Human action recognition using meta-cognitive neuro-fuzzy inference system. Int. J. neural Syst. 22, 1250028.
Tang, D., Chang, H.J., Tejani, A., Kim, T.-K., 2014. Latent regression forest: structured estimation of 3d articulated hand posture. In: Computer Vision and Pattern Recognition (CVPR), pp. 3786–3793. IEEE Conference on, IEEE, 2014.
Taylor, J., Shotton, J., Sharp, T., Fitzgibbon, A., 2012. The Vitruvian Manifold: Inferring Dense Correspondences for One-shot Human Pose Estimation. In: Computer Vision and Pattern Recognition (CVPR), pp. 103–110. IEEE Conference on, IEEE, 2012.
Thrun, S., Burgard, W., Fox, D., 2005. Probabilistic Robotics. MIT press.
Thurau, C., Hlaváč, V., 2008. Pose primitive based human action recognition in videos or still images. In: Computer Vision and Pattern Recognition, pp. 1–8. CVPR 2008. IEEE Conference on, IEEE, 2008.
Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.
Tompson, J., Stein, M., Lecun, Y., Perlin, K., 2014. Real-time Continuous Pose Recovery of Human Hands Using Convolutional Networks, vol. 33. ACM Transactions on Graphics (TOG), p. 169.
Tu, J.V., 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J. Clin. Epidemiol. 49, 1225–1231.
Viola, P., Jones, M., 2001. Robust real-time object detection. Int. J. Comput. Vis. 4, 51–52.
Vygotsky, L.S., 1980. Mind in Society: the Development of Higher Psychological Processes. Harvard university press.
Wachs, J.P., Kölsch, M., Stern, H., Edan, Y., 2011. Vision-based hand-gesture applications. Commun. ACM 54, 60–71.
Wan, E., Van Der Merwe, R., 2000. The unscented Kalman filter for nonlinear estimation. In: Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000. IEEE, pp. 153–158.
Wang, C., Liu, Z., Chan, S.-C., 2015. Superpixel-based Hand Gesture Recognition with Kinect Depth Camera, Multimedia. IEEE Transactions on, 17, pp. 29–39.
Weinland, D., Ronfard, R., Boyer, E., 2011. A survey of vision-based methods for action representation, segmentation and recognition. Comput. Vis. Image Underst. 115, 224–241.
Wilson, A.D., Bobick, A.F., 1999. IEEE Transactions on. Parametric Hidden Markov Models for Gesture Recognition, Pattern Analysis and Machine Intelligence, 21, pp. 884–900.
Wilson, A.D., Bobick, A.F., 2001. Hidden Markov models for modeling and recognizing gesture under variation. Int. J. Pattern Recognit. Artif. Intell. 15, 123–160.
Ye, M., Wang, X., Yang, R., Ren, L., Pollefeys, M., 2011. Accurate 3d pose estimation from a single depth image. IEEE International Conference on, IEEE, 2011. In: Computer Vision (ICCV), pp. 731–738.
Yu, S.-Z., 2010. Hidden semi-Markov models. Artif. Intell. 174, 215–243.
Zhang, Y., Harrison, C., 2015. 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, pp. 167–173.
Zhang, X., Chen, X., Wang, W.-h., Yang, J.-h., Lantz, V., Wang, K.-q., 2009. 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, pp. 401–406.
Zhou, Z.-H., Wu, J., Tang, W., 2002. Ensembling neural networks: many could be better than all. Artif. Intell. 137, 239–263.

Please cite this article in press as: Liu, H., Wang, L., Gesture recognition for human-robot collaboration: A review, International Journal of Industrial Ergonomics (2017), http://dx.doi.org/10.1016/j.ergon.2017.02.004