Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

A Review Paper on Involuntary Human Motion Acknowledgement to Assist Feature Robotic Skills

[1]Dr. Nidhi Mishra, [2]Dr. F Rahman, [3]Mr. Kapil Kelkar

[1, 2]Assistant Professor, Faculty of Information & Technology, Kalinga University, near Mantralaya, Atal Nagar-Nava Raipur, Chhattisgarh 492101

[3]Assistant Professor, Faculty of Science, Kalinga University, near Mantralaya, Atal Nagar-Nava Raipur, Chhattisgarh 492101

[1]nidhi.mishra@kalingauniversitya.ac.in, [2]f.rahman@kalingauniversitya.ac.in, [3]kapil.kelkar@kalingauniversitya.ac.in

Article Info
Page Number: 2621-2630
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract

It is a difficult effort due to issues such as backdrop clutter, partial occlusion, changes in scale, viewpoint, lighting, and appearance to recognize human activities from video sequences or still photos. A multiple activity recognition system is required for a wide variety of applications, such as video surveillance systems, human-computer interaction, and robots for the characterization of human behavior. In this article, we present a comprehensive evaluation of recent and cutting-edge research developments in the field of human activity classification. We propose a classification of human activity research strategies and then evaluate the benefits and drawbacks of each of these approaches. In particular, we divide human activity classification algorithms into two broad categories based on whether or not they use data from multiple modalities. The first category is called "without using data from multiple modalities," and the second is called "using data from multiple After that, each of these categories is broken down even further into sub-categories, which indicate the manner in which they simulate human behaviors and the kinds of activities in which they have an interest. In addition, we present a complete study of the human activity classification datasets that now exist and are accessible to the public, and we investigate the characteristics that should be met by an ideal human activity identification dataset. Finally, we address several unanswered questions about human activity recognition and reflect on the characteristics of potential future research directions. It is common practise for humanoid robots to employ a template-based dialogue system. This type of system is able to reply successfully inside a specific discourse domain, but it is unable to react effectively to information that falls outside of that discourse domain. Because the interactive elements don't have an emotional detection system, the rules for the dialogue system are hand-drawn instead of being automatically generated. In order to achieve this goal, a humanoid robot open-domain chat system and a deep neural network emotion analysis model were both developed. The former is intended to assess the feelings that interacting objects may have. Emotional state analysis, in addition to research on Word2vec and language coding, are all components of the process. Following this, the emotional state of a humanoid robot is taught with the help of a Training and emotional state analysis paradigm, which is described. The conventional dialogue system for humanoid robots is

based on template construction. This system can provide acceptable answers within the designated discussion region, but it cannot go beyond this area. The rules of the dialogue system are based on manual creation, and it does not include any emotional recognition. This research built an open-domain dialogue system for a humanoid robot in addition to an emotion analysis model that was based on a deep neural network. The model was used to assess the emotions of interacting objects. Language processing, coding, feature analysis, and Word2vec are all essential components of emotional state analysis. A humanoid robot's emotional state analysis training findings are detailed here along with those results' implications. As science and technology continue to grow, robots have gradually made their way into every facet of human existence. Robots find use in manufacturing, the armed forces, home healthcare, education, and laboratories [1]. The three guiding principles of robotics [2] state that the ultimate goal of robot development is to have robots behave like humans, assist people in performing activities in a more effective manner, and accomplish goals [3]. To accomplish goals in human-robot cooperation, people need to interact more effectively with the robot [4, 5]. The traditional method of human-computer interaction involves a person inputting data through the use of a keyboard, mouse, and various other manual input devices, while a computer would output data to a person through the use of a display and various other peripherals. This contact requires a number of supplementary items. In the actual world, a computer is not accessible to everyone [6]. The natural routes of communication between people and machines include speech, vision, touch, hearing, proximity, and other human interactions [7]. This manner of connection is not only common but also productive. [8] So that human beings and robots can work together more efficiently. The emotion analysis model of the humanoid robot is able to assess and detect the emotional information of the interacting object while the object is interacting [9]. During contact, the language of the object carries a wealth of emotional information, and the textual content represents human cognition at a very advanced level.

1. Introduction

Recognition of human activity has a tremendous impact on human-to-human interactions and interpersonal bonds. A person's identity, personality, and psychological condition are all contained in their fingerprints. Computer vision and machine learning researchers are particularly interested in how people perceive the actions of others. As a result of this investigation, a system for multimodal activity recognition is now required in various applications, such as video surveillance systems, human-computer interaction, and robotics for characterizing human behaviour.

"What action?" and "What action?" are two of the most often asked questions in the field of classification. ("Where in the video?") and "Where in the video are you?" (i.e., the challenge of localization) A person's kinetic states must be determined in order for the computer to accurately identify the activity. Typical human actions, such as "walking" and "running," can be easily identified in daily life. On the other hand, it is more difficult to recognize more complex tasks, such as "peeling an apple." Complex actions can be broken down into smaller,

easier-to-understand tasks. We can learn more about human activities by analysing what things are there in a scene and how they relate to what is going on around them (Gupta and Davis, 2007).

A figure-centric setting with an uncluttered background, where the actor is free to conduct an activity, is assumed in most human activity recognition work. Human activity recognition is a difficult topic to solve because of issues such as backdrop clutter and partial occlusion. Other difficulties include variations in scale, viewpoint and lighting, as well as frame resolution. Additionally, annotating behavioral roles is time-consuming and requires an understanding of the specific event in order to do so effectively. As a result, the problem becomes even more difficult to solve because of the many similarities between classes. It's possible, then, that acts performed by members of different classes may be difficult to differentiate from those performed by members of the same class using the same body movements. There are many ways in which people conduct an activity, and this makes it difficult to understand what the underlying action is. It's also difficult to create a real-time visual model for learning and evaluating human movement with insufficient benchmark datasets.

Three components are needed to overcome these issues: I background subtraction (Elgammal et al., 2002; Mumtaz et al., 2014), in which the system tries to separate the parts of the image that are invariant over time (background) from those that are moving or changing (foreground); (ii) human tracking, in which the system locates human motion over time; and (iii) acquiescence detection.

Video or still photographs of people's movements can be used to study their behaviour using human activity recognition. Human activity recognition systems are motivated by this reality and seek to appropriately classify input data into its underlying activity category. Gestures, atomic actions, human-to-object or human-to-human interactions, group actions, behaviours, and events are all examples of human activities. According to the degree of complexity, the breakdown of human activities can be shown in Figure 1.

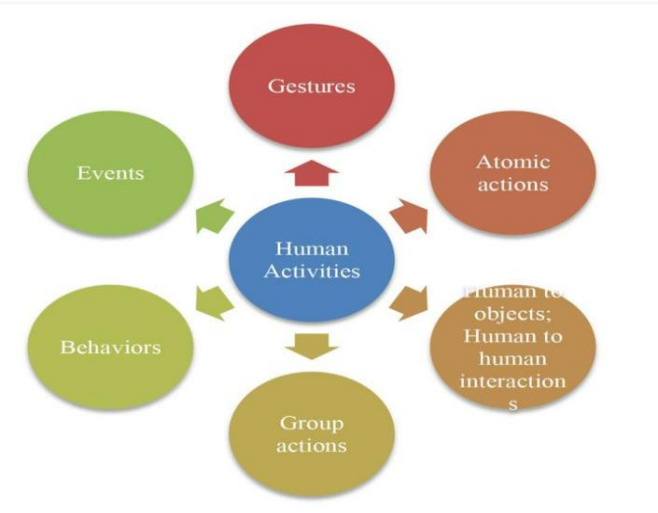

Figure 1. Decomposition of human activities

When a person makes a gesture, they are making simple movements of their body to express an idea or an activity (Yang et al., 2013). Movements of a person in which they describe a specific motion, which may be part of more complicated tasks, are called "atomic actions" (Ni et al., 2015). Encounters between people and objects are known as "human to human" or "human to object" interactions (Patron-Perez et al., 2012). Activities carried out by a group or individuals are referred to as "group actions" (Tran et al., 2014b). The term "human behaviour" refers to the physical activities that are linked to a person's feelings, personality, and mental state (Martinez et al., 2014). To sum things up, events are high-level activities that describe the behaviours of persons and reveal the intentions or social roles of individuals (Lan et al., 2012a).

2. Previous Surveys and Taxonomies

There are a number of studies in the field of human activity detection. Gavrila (1999) distinguished between 2D (using explicit shape models or not) and 3D approaches to research. Human motion analysis, tracking from single and multiview cameras, and recognition of human activities were the emphasis of Aggarwal and Cai (1999). An action categorization hierarchy similar to that provided by Wang et al. (2003) has been developed. The study by Moeslund et al. (2006) focused primarily on posture-based action recognition methods and suggested a four-fold taxonomy, which included initialization of human motion, tracking, pose estimation, and recognition methods. '

As Turaga et al. (2008) found when classifying activity recognition algorithms, there is a distinct difference between the meaning of "action" and "activity." Poppe (2010) distinguished between "top-down" and "bottom-up" techniques of human activity recognition. However, the taxonomy of human activity identification methods proposed by Aggarwal and Ryoo (2011) is based on a tree-structured taxonomy, which divides the methods into two main categories: "single layer" approaches and "hierarchical" approaches, both of which include several layers of categorization.

In the literature, the terms "activity" and "behaviour" are often used interchangeably (Castellano et al., 2007; Song et al., 2012a). It is important to distinguish between these two concepts in this survey since the term "activity" is used to define a series of actions that correlate to a certain body movement. Both activities and events that are linked to a single person's gestures, emotions, facial expressions and audible signals are referred to as "behaviour." Figure 2 depicts some of the most common human actions in a simplified form.

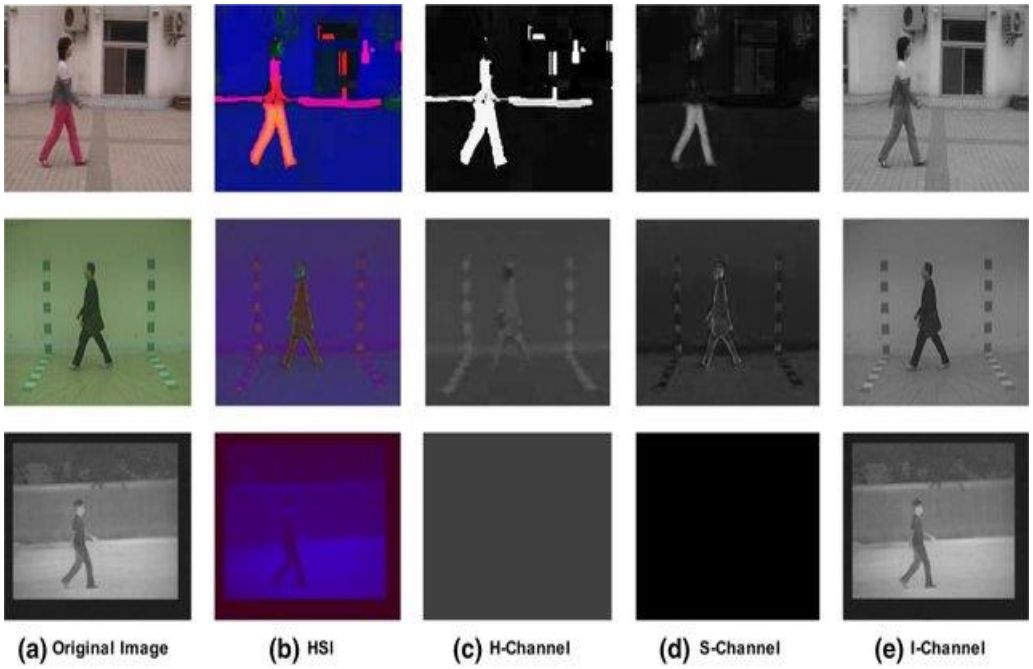

Figure 2. Representative frames of the main human action classes for various datasets.

3. Recognition and Detection of Emotional States in Images and Videos

3.1.1. Speech from the interactive object must be recorded using a microphone and converted into an audio file before the voice recognition module can extract text information from the audio file during use. Using pre-processed text data, the emotional state of the interactive object is produced from the model's emotion analysis. A model for analysing emotions based on collected data was developed using machine learning in this article [21]. The model is reserved once offline training with data sets is completed using a specific way. The previously saved model is used to make predictions. Text emotions can be detected using a machine learning technique, as seen in Figure 2.

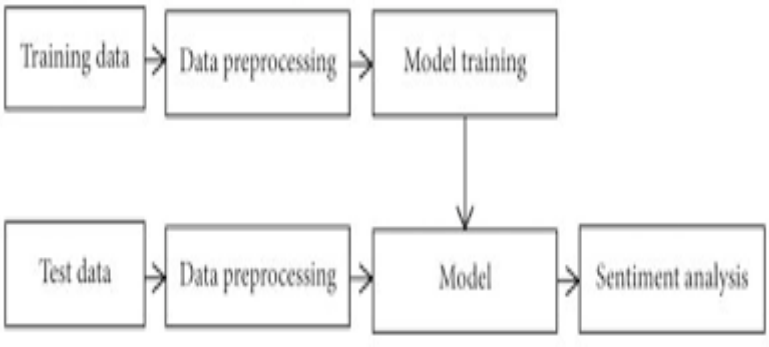

Figure 3 Text emotion analysis process based on machine learning.

Data Collection and Preprocessing 3.2

Data capture and pre-processing are included in the data. The following is a breakdown of the contents.

3.2.1. Data Collection

We used the "Microblog Cross-Language Emotion Recognition Dataset" to build an emotion analysis model for a humanoid robot. Corpus has negative and positive categories. The data collection includes 12,153 negative and 12,178 positive corpus groupings. The microblog corpus' casual language are appropriate for training an emotion identification system.

Data prep Down sampling excluded 25 negative label items from the negative label corpus, and positive and negative samples were blended into 12,153 items. Since most of the corpus originates from Weibo, it's full of emoticons and repetitive punctuation. Voice recognition eliminates repetitive punctuation [22]. Preprocessing eliminated repetitive punctuation and emoticons as characteristics. This study's word segmentation phase employed Jieba's toolkit. Textual processing table 1 compares punctuation and facial expressions. Table 1: Comparison table of punctuation and facial expressions in textual processing There are several ways to represent a phrase in a vector space, but the word bag model is one of the most often used methods.

4. A Study of Characteristics

The chi-square test, information gain, mutual information technique, and TF-IDF are all typical feature selection approaches in the text vector space model. Absolute term frequency (TF) and inverse document frequency (IDF) are used to emphasise keywords in documents using the word bag model's TF-IDF combination (IDF). The training text's feature item's absolute word frequency (TF) indicates its spelling. Absolute word frequency may be used to readily identify the most important terms in a document. Following this formula, the inverted document frequency (IDF) may be calculated.

For example, ni denotes how many times each feature item occurs in the training set in the total number of documents. Fewer but better classified terms are highlighted by the IDF in this list. IDF will ensure that unusual words are not omitted from the corpus during the actual calculating phase.

Google came up with the concept of Word2vec in 2013. An example of dense feature representation, or distributed representation, in which words are represented by their features. Models for training Word2vec include CBOW (continuous bag of words), and Skip-Gram. Hierarchical SoftMax and Negative Sampling are two kinds of enhanced algorithms for Word2vec, both of which are aimed at speeding up computation and training. For the Hierarchical SoftMax model, the projection layer's output is the mean word vector sum under the CBOW model, and the projection layer's output is the same for the Skip-Gram model. The Hierarchical SoftMax technique employs Huffman trees instead of SoftMax mapping of the projection layer to the output layer in order to avoid computing the probability of every words. Sampling in the Negative For neural networks, word2vec is a pretraining approach that may improve the neural network's training starting point and make optimization simpler [23]. Word vector is also a pretraining method.

The dense feature is easier to compute and does not suffer from the issue of dimension explosion, which has a great generalisation ability. Compared to independent thermal coding. It is possible to compare features using dense feature representation. In natural language processing, such as Chinese word segmentation, sentiment analysis, and reading comprehension, this distributed form of the word vector is extensively utilised.

5. Developing a Model for Emotional State Analysis

Text categorization may also be done using the support vector machine (SVM). A support vector machine's primary goal is to identify the feature space hyperplane with the biggest interval. A major benefit of this method is that it works well even when the number of dimensions exceeds the number of samples, even in a high-dimensional space. Support vector machines may be configured to use a variety of different kernel functions. However, when the number of features exceeds the number of samples, SVM's performance suffers.

As a sample or phrase for each sample in the training data set T, N stands for the number of samples that are included in the training data set. The formula for solving optimization issues using support vector machines is as follows:

Calculate the normal vector that separates the hyperplane in formula (3) by substituting. The issue may be stated in two ways:

A semidefinite matrix of shape is used in formula (4) to represent the upper limit of the Lagrange daily number.

Scaling was utilised to increase the support vector machine output category probability in this work. Here, we utilise the Sigmoid function, which is a parametric technique for fitting logistic regression model output values, to map the values between, and then we use this model's original output values to map probability values, as indicated in the following formula:

A and B are trainable parameters in formula (3), which represents the support vector machine's decision function, which may output labels corresponding to any input X.

using support vector machines is as follows:

Calculate the normal vector that separates the hyperplane in formula (3) by substituting. The issue may be stated in two ways:

A semidefinite matrix of shape is used in formula (4) to represent the upper limit of the Lagrange daily number.

Scaling was utilised to increase the support vector machine output category probability in this work. Here, we utilise the Sigmoid function, which is a parametric technique for fitting logistic regression model output values, to map the values between, and then we use this

model's original output values to map probability values, as indicated in the following formula:

A and B are trainable parameters in formula (3), which represents the support vector machine's decision function, which may output labels corresponding to any input X.

For example, the cross-entropy loss may be seen in the following formulas:

Platt Scaling allows support vector machines to output the likelihood of a category. When it comes to matching points, it's all about how near they are to the interface, and how distant they are from the interface, and how close they are to each other.

Conclusion

Emotion recognition, data gathering, pre-processing, emotional state analysis language processing, model construction, and integration are the main activities in emotion analysis. In the experiment, TF-IDF characteristics and the word bag model impact training results. In a single modelling framework, TF-IDF and SVM get the best results. The best model integration approach uses an SVM learner and a stacking strategy. This paper's main contribution is a fresh viewpoint. Each component's procedures are just briefly detailed. This paper examines an SVM-based sentiment analysis model. Attention processes affect the neural network model, and a typical machine learning model's performance is compared with different inputs. Support vector machine with TF-IDF obtained the best classification result in the single model trial (F1 = 0.795, 78.94% accuracy, AUC = 0.863).

Reference:

1. L. Gabriella, G. Márta, K. Veronika et al., “Emotion attribution to a non-humanoid robot in different social situations,” PLoS One, vol. 9, no. 12, Article ID e114207, 2014.View at: Publisher Site | Google Scholar
2. A. Rozanska and M. Podpora, “Multimodal sentiment analysis applied to interaction between patients and a humanoid robot pepper,” IFAC-Papers Online, vol. 52, no. 27, pp. 411–414, 2019.View at: Publisher Site | Google Scholar
3. M. Viríkova and S. Peter, “Teach your robot how you want it to express emotions,” Advances in Intelligent Systems and Computing, vol. 316, pp. 81–92, 2015.View at: Google Scholar
4. X. Ke, Y. Shang, and K. Lu, “Based on hyper works humanoid robot facial expression simulation,” Manufacturing Automation, vol. 137, no. 1, pp. 118–121, 2015.View at: Google Scholar
5. F. Azni Jafar, N. Abdullah, N. Blar, M. N. Muhammad, and A. M. Kassim, “Analysis of human emotion state in collaboration with robot,” Applied Mechanics and Materials, vol. 465-466, pp. 682–687, 2013.View at: Publisher Site | Google Scholar
6. Z. Shao, R. Chandramouli, K. P. Subbalakshmi, and C. T. Boyadjiev, “An analytical system for user emotion extraction, mental state modeling, and rating,” Expert Systems

with Applications, vol. 124, no. 7, pp. 82–96, 2019.View at: Publisher Site | Google Scholar

7. J. Hernandez-Vicen, S. Martinez, J. Garcia-Haro, and C. Balaguer, "Correction of visual perception based on neuro-fuzzy learning for the humanoid robot TEO," Sensors, vol. 18, no. 4, pp. 972-973, 2018.View at: Publisher Site | Google Scholar
8. A. Zaraki, D. Mazzei, M. Giuliani, and D. De Rossi, "Designing and evaluating a social gaze-control system for a humanoid robot," IEEE Transactions on Human-Machine Systems, vol. 44, no. 2, pp. 157–168, 2014.View at: Publisher Site | Google Scholar
9. J. Wainer, B. Robins, F. Amirabdollahian, and K. Dautenhahn, "Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism," IEEE Transactions on Autonomous Mental Development, vol. 6, no. 3, pp. 183–199, 2014.View at: Publisher Site | Google Scholar
10. L. Tang, Z. Li, X. Yuan, W. Li, and A. Liu, "Analysis of operation behavior of inspection robot in human-machine interaction," Modern Manufacturing Engineering, vol. 3, no. 3, pp. 7-8, 2021.View at: Google Scholar
11. Z. Li and H. Wang, "Design and implementation of mobile robot remote human-computer interaction software platform," Computer Measurement & Control, vol. 25, no. 4, pp. 5-6, 2017.View at: Google Scholar
12. H. Huang, N. Liu, M. Hu, Y. Tao, and L. Kou, "Robot cognitive and affective interaction model based on game," Journal of Electronics and Information Technology, vol. 43, no. 6, pp. 8-9, 2021.View at: Google Scholar
13. Lufei, Y. Jiang, and G. Tian, "Autonomous cognition and personalized selection of robot service based on emotion-spatiotemporal information," Robot, vol. 40, no. 4, pp. 9-10, 2018.View at: Google Scholar
14. J. Law, P. Shaw, and M. Lee, "A biologically constrained architecture for developmental learning of eye-head gaze control on a humanoid robot," Autonomous Robots, vol. 35, no. 1, pp. 77–92, 2013.View at: Publisher Site | Google Scholar
15. A. Cela, J. Yebes, R. Arroyo, L. R. Bergasa, and E. López, "Complete low-cost implementation of a teleoperated control system for a humanoid robot," Sensors, vol. 13, no. 2, pp. 1385–1401, 2013.View at: Publisher Site | Google Scholar
16. E. Tidoni, P. Gergondet, A. Kheddar, and S. M. Aglioto, "Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot," Frontiers in Neurorobotics, vol. 8, 2014.View at: Google Scholar
17. A. M. BaTula, Y. E. Kim, and H. Ayaz, "Virtual and actual humanoid robot control with four-class motor-imagery-based optical brain-computer interface," BioMed Research International, vol. 2017, Article ID 1463512, 13 pages, 2017.View at: Publisher Site | Google Scholar
18. T. Sato, Y. Nishida, J. Ichikawa, and Y. Hatamura, "Active understanding of human intention by a robot through monitoring of human behavior," in Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems, IROS'94, pp. 405–414, IEEE, Munich, Germany, September 1994.View at: Google Scholar
19. H. Clint, Modelling Intention Recognition for Intelligent Agent Systems, DTIC, Mexico City, Mexico, 2004.

20. K. A. Tahboub, “Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition,” Journal of Intelligent and Robotic Systems, vol. 45, no. 1, pp. 31–52, 2006.View at: Google Scholar
21. T. Koolen, S. Bertrand, G. Thomas et al., “Design of a momentum-based control framework and application to the humanoid robot atlas,” International Journal of Humanoid Robotics, vol. 13, no. 1, 2016.View at: Publisher Site | Google Scholar