

Human Body Language Analysis: A Preliminary Study Based on Kinect Skeleton Tracking

Danilo Avola¹, Luigi Cinque², Stefano Levialdi², and Giuseppe Placidi¹

¹ Department of Life, Health and Environmental Sciences, University of L'Aquila
Via Vetoio Coppito 2, 67100, L'Aquila, Italy
{danilo.avola,giuseppe.placidi}@univaq.it
<http://www.univaq.it/en/section.php?id=262>

² Department of Computer Science, Sapienza University
Via Salaria 113, 00198, Rome, Italy
{cinque,levialdi}@di.uniroma1.it
<http://w3.uniroma1.it/dipinfo/english/index.asp>

Abstract. The nonverbal communication can be informally defined as the communicative process between two or more entities (e.g., persons) which achieving an informative exchange without using the semantic meaning of the words. This process can be accomplished by using one or more language forms, including the body language (i.e., movements, gestures, and postures) which in turn can be composed by voluntary and involuntary behaviours. The analysis and interpretation of these behaviours can infer different internal states of persons (e.g., feelings, attitudes, emotions) which in turn can support the development of a wide range of automatic applications in different fields, such as: rehabilitation, security, people identification, human behaviour analysis, biometric.

In recent years, we have focused our efforts in developing a first implementation of Kinematic, a novel multimodal framework designed to support advanced human-machine interfaces. The purpose of the framework is to provide a tool to analyze and interpret verbal and nonverbal human-to-human communication in order to transfer this ability to the human-machine interaction. In this paper we face a specific aspect of the framework regarding the first calibration phase of the numerical measures related to the Kinect skeleton used to analyze and interpret the body language. The numerical measures was obtained analyzing the movements of the skeleton during individual and social contexts. A preliminary qualitative and quantitative study has been reported and discussed.

Keywords: nonverbal communication, body language, human-machine interfaces, skeleton, numerical measures.

1 Introduction

The term communication derives from Latin *communicare* which means “to share” or “to make common”. Over the centuries, this term has taken on increasingly heterogeneous meanings, representing material/immaterial as well as

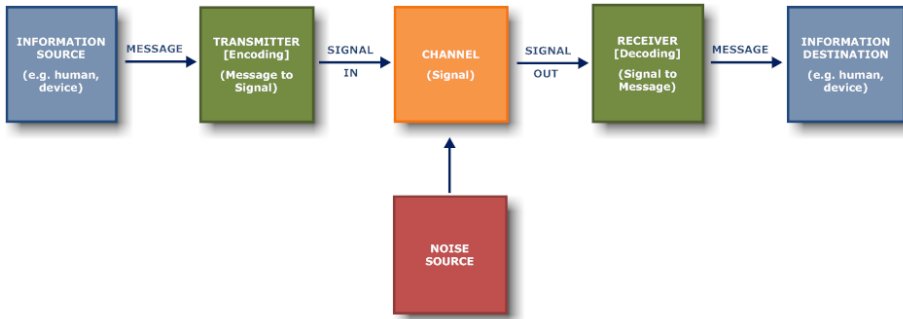


Fig. 1. Schematic Diagram of a General Communication System

concrete/abstract aspects of the interpersonal and social interaction processes. Subsequently, the term has been progressively focused on defining the incorporeal and conceptual information exchange between people and/or devices. In this context, a main contribute was provided from [16], where the authors formalized the basic theoretical aspects of communication. Their work, as shown in Fig. 1, consisted in providing an abstraction of the general steps involved in a communication process which can be summarized as follows:

Information Source: which produces a message or sequence of messages to be communicated to the receiving terminal.

Transmitter: which operates on the message to produce a signal suitable for transmission over the channel.

Noise Source: which represents the possible noise within the communication channel.

Receiver: which ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal.

Information Destination: which represents the “entity” (e.g., person, device) to whom the message or sequence of messages are intended.

Two elements remain implicitly defined in the schema:

Channel: representing the medium used to transmit the signal from transmitter to receiver.

Code: representing the set of symbols (e.g. words, signs, gestures) and related rules (i.e. to combine them) adopted to define a specific language.

The proposed definition has been subject of many variants and interpretations [12,11]. In our context, we adopt the steps of general communication system both to study the human-to-human communication processes and to develop the related human-machine interactions. In particular, we consider both “information source”/“transmitter”, and “receiver”/“information destination” belonging to two different combined entities: human and device, respectively. As shown in [7], the communication concept can be further characterized according to the following classification:

Nonverbal Communication: which represents the information exchange with the use of spoken language.

Verbal Communication: which represents the information exchange without the use of spoken language.

In particular, the **Nonverbal Communication** can be further classified as follows:

Paralinguistic System: which represents the set of sounds of the vocal (non-verbal) communication system (independently from the meaning of the words). It takes into account several aspects, such as: *tone* (which is influenced from different aspects: age, gender, and so on), *frequency* (which is tied to the general contexts and events), and *rhythm* (which is driven by specific contexts and social status).

Kinesic System: which represents two different classes of “movements”. The first class considers the *body movements*, including: gestures (such as: hands, legs or arms movements), and posture (i.e., pose of the whole body or parts of it). The second class considers *facial expressions*, including: eyes movements, emotions, and so on.

Proxemics: which represents the study of the space use during a communication process. In particular, it explains that each person divides the surrounding space in four classes: *intimate zone* (from 0 to 50 centimeters), *personal zone* (from 50 cm to 1 meter), *social zone* (from 1 m to 3/4 m), *public zone* (over 4 m).

Haptic: which represents the relationship between the communication process and the physical contact. Also in this case, according to the specific “action” (e.g., handshake, hug) it is possible to derive peculiar information about the nature of the informative exchange.

The previous items have voluntary and involuntary aspects that highlight intrinsic behaviours of the human according to events, situations and personal internal states. Lately, we have intensified our efforts in developing a first implementation of Kinematic [3,4], a novel multimodal framework, based on Kinect [15], designed to support advanced human-machine interfaces. The purpose of the framework is to provide a tool to analyze and interpret verbal and nonverbal human-to-human communication in order to transfer this ability to the human-machine interaction. In this paper we face a specific aspect of the framework regarding the calibration phase and the first numerical measures related to the Kinect skeleton used to analyze and interpret the gesture class belonging to the kinesic system (i.e., body language).

This paper is structured as follows. Section 2 introduces some basic works supporting this preliminary measuring phase. Section 3 shortly presents an overview of the framework. Section 4 summarizes the calibration phase of the numerical measures related to the Kinect skeleton. Finally, Section 5 concludes the paper and introduces the current developments.

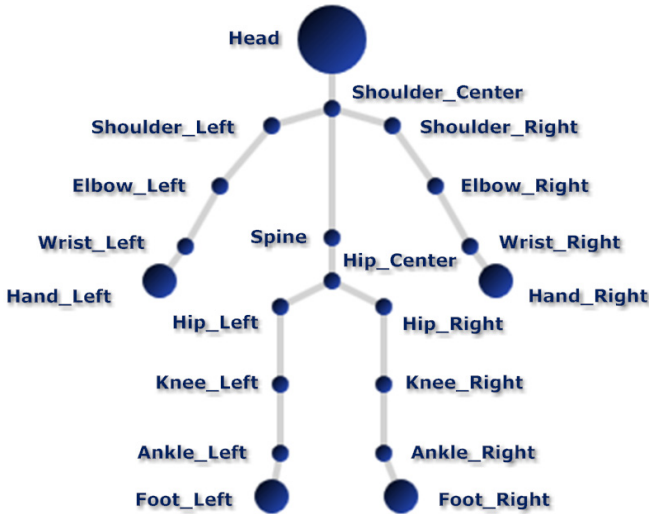


Fig. 2. Kinect Skeleton Model

2 Background

This section introduces some basic works supporting the set of numerical measures adopted to perform the first calibration phase related to the Kinect skeleton. As shown in Fig. 2 the kinect skeleton model of a human can be represented with a stickman having up to 20 joints (according to the Microsoft joint classification [17]), which can be used to analyze pose and tracking of the subject. Our intuition was to study the spatial movements over the time of the joints using well known numerical measures inherited from handwriting and freehand drawing analysis. The concept is that the spatial-temporal evolution of each joint, during a common activity (e.g., interpersonal or social interactions), tends to form “graphical shapes” which can be analyzed with usual mathematical and/or geometrical measures (e.g., curvature measure) to infer human behavioural patterns. In our context, we just wanted to classify the “disorder level” of the joints movement. The “disorder level” can be seen as the accuracy through which each “graphical shape” related to each joint (within a time interval) is drawn. In other words, the “disorder level” expressed by a set of joints during a specific interaction of the human subject can characterize a class of gestures and/or behaviours (e.g., social behaviour classification, emotion detection).

From an abstract point of view different works confirm our intuition. For instance, advanced works in graphology [8,6] show as the handwriting analysis can be used to evaluate and predict some basic aspects of the human personality. On the same direction, we analyze the graphic over the time composed by the tracking of the joints during a user interaction activity. To reach this task we were inspired by specific handwriting/freehand drawing measures related to the sketch recognition activity. A first interesting work is shown in [2], where the authors

describe an intelligent framework able to automatically distinguish, in on-line way, freehand drawing from handwriting. Their approach takes into account only the mathematical features belonging to the sketch performed by users during interaction activity. In particular, their framework adopts a set of numerical measures (i.e., a feature vector) to analyse an object composed of a set of strokes in order to classify it as text or drawing. A similar work is presented in [1], where a gesture recognition system is introduced. Their system acquires freehand gestures performed by users which manage a led pen. A color tracking algorithm analyzes in real time the sequential frames and extracts the center of gravity of the blob on each of them. These centers, combined to form a sketch, are interpreted by means of mathematical features which identify the related gesture. Other basic works are presented in [10,9], where a set of geometrical features and a fuzzy logic approach to classify elementary geometric shapes, combined with an extensible set of heuristics is presented. Both works introduce a recognizer able to identify elementary geometrical shapes (e.g., triangles, diamonds, circles) which are recognized independently of changes in rotation, size or number of individual strokes. These last aspects play a key role in our context since the tracking of the joints can be influenced by this kind of issues. Another remarkable method is shown in [5], where a system to separate textual from graphical domain is presented. In their approach the authors set out first to find the single logically coherent feature which distinguishes shape from text. The authors observed that, when using any general set of coordinate equations, handwritten text symbols are more difficult to describe than common shapes (which are geometrically simpler). In that sense, text strokes are more randomly structured. For this reason, the measure of the degree of randomness of text strokes (high) compared with shape strokes (low) can be considered an accurate criterion of classification. We have applied a similar measure to support the characterization (concurrently with other measures) of the “graphical shapes” derived from the tracking of the joints. Other works were considered in relation to the used approach. A first example is described in [14], where the authors combine two algorithms for application to the recognition of unconstrained isolated handwritten numerals. The first algorithm employs a modified quadratic discriminant function utilizing direction sensitive spatial features of the numeral image. The second algorithm utilizes features derived from the profile of the character in a structural configuration to recognize the numerals. While both algorithms yield very low error rates, the authors combine the two algorithms in different ways to study the best polling strategy and realize very low error and rejections rates. In a similar way, we have analyzed the correlation among our selected features. A last remarkable work is presented in [13], where the authors adopt a Genetic Algorithm (GA) for solving the off-line handwriting recognition issue. Their basic idea is to use the GA both to combine various styles of writing a character and to generate new styles. In other words, the authors tried to transfer the ability of the human mind in understanding new character style (seen for the first time) within an application. They obtained excellent results. In similar way, we have tried to classify a set of general behavioural patterns by observing a sub-set of them. During our research

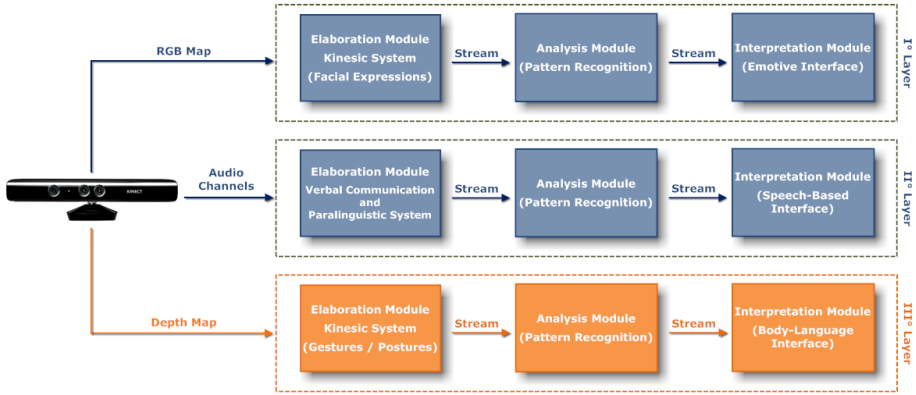


Fig. 3. The Framework Architecture

we have considered several measures, in this work we will show a restricted set of them since here our purposes is to define the basic calibration phase related to the Kinect skeleton.

3 The Framework Architecture

In order to provide a complete overview of the Kinematic framework, in Fig. 3 the whole simplified architecture is provided. As shown, it is composed by three parallel layers which elaborate the different input coming from the Kinect device: RGB Map, Audio Channels, and Depth Map. As previously discussed, in this work we are only interested in studying and providing preliminary results related to the Kinect skeleton model (i.e., third layer) according to a set of specific interactions. To achieve this result we have implemented (within the analysis module) three main mathematical features: *curvature*, *entropy*, and *homogeneity*.

Each feature treats the abstract “graphical shapes” drawn by each joint during the interpersonal or social activity as a set of strokes. The *curvature feature* measures the speed with which a stroke changes its trajectory. This task is achieved by dividing each stroke (when possible) in different sub-strokes according to the direction changes depending on the tracing direction, and by analyzing each of them by means of its spatial derivative. The computation of the distances among maximum and minimum points of all derivatives can be considered a characterizing measure of the main stroke. The *entropy feature* measures the stochastic distribution of the pixels composing a stroke. This task is achieved by considering the pixel disposition above and below the horizontal line of symmetry of the bounding box containing the stroke and by comparing their degree of randomness. Also in this case the computation of the absolute difference between the two sub-distributions can be considered a distinguishing measure of the stroke. The *homogeneity feature* measures both interception and closeness of the strokes

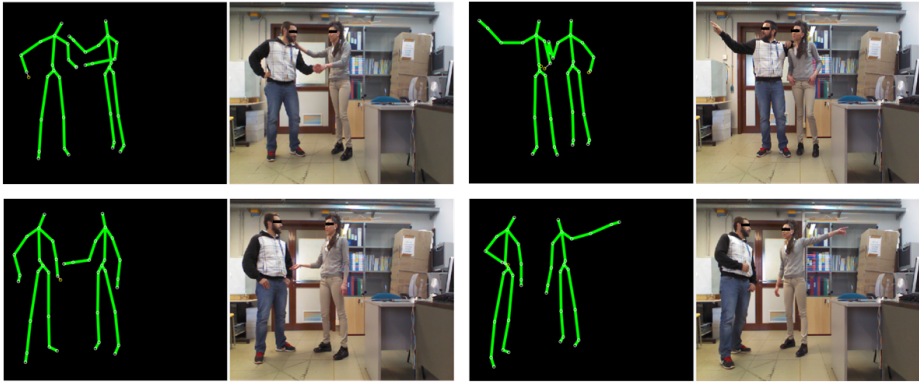


Fig. 4. Examples of Social Interaction

coming from the different joints. In particular, this last measure tend to identify spatial relationship among six groups of joints:

- $group_1$ (head, shoulder center),
- $group_2$ (spine, hip center),
- $group_3$ (shoulder left, elbow left, wrist left, hand left),
- $group_4$ (hip left, knee left, ankle left, foot left),
- $group_5$ (shoulder right, elbow right, wrist right, hand right),
- $group_6$ (hip right, knee right, ankle right, foot right).

In our context, the interpretation of the three measures according to specific interpersonal or social interaction can support the classification of different human behavioural patterns.

4 Experimental Results

With the aim to evaluate the degree of randomness related to the movements of the joints we have planned three activities: in the first, 5 subjects had to draw on the air graphical and alphabet symbols using one or both hands; in the second, the same subjects had to explain a research paper using gestures (with hands and/or body); in the last, they were encouraged to perform a free exchange of opinions (e.g., social interaction) on free arguments (e.g., politics, sport). The screenshots in Fig. 4 show some examples related to the last activity.

Table 1 summarizes the basic intervals of values related to the three features in consideration of each specific group of joints. Within each group we have computed an average value (i.e., maximum and minimum) to provide a single interval. Although the obtained results are preliminary and rough, a careful analysis of the Table 1 has allowed us to extract some useful information. First of all, we have observed that all the joints can be numerically characterized within

Table 1. Case Study - Planned Activities and Features

Features/Groups	<i>Group</i> ₁	<i>Group</i> ₂	<i>Group</i> ₃	<i>Group</i> ₄	<i>Group</i> ₅	<i>Group</i> ₆
Curvature	[10,5..25,9]	[10,1..13,2]	[12,4..60,5]	[10,8..64,8]	[23,6..40,3]	[25,2..42,4]
Entropy	[22,1..33,2]	[20,3..25,5]	[30,4..65,2]	[28,4..68,5]	[28,6..45,8]	[26,6..44,7]
Homogeneity	[75,0..88,0]	[70,0..82,0]	[77,0..89,0]	[78,0..90,0]	[79,0..90,0]	[78,0..91,0]

a well-defined numerical interval. This means that we can hope to be able to classify human behavioural patterns according to the different activities. Moreover, we have observed consistent differences between the numerical values in relation to the specific interaction. Studying the distribution of the numerical values, we have also noted a direct relationship between *curvature* and *entropy*, and an inverse relationship between *curvature* and *homogeneity*. These last aspects highlight that the measures have a satisfactory degree of correlation which will be useful for detailing the intervals of classification according to the different activities. Finally, we have also tested that some gestures/interactions not belonging to the three introduced activities do not fall within the fixed intervals.

5 Conclusions

The verbal and nonverbal human-to-human communication can be analyzed and interpreted in order to transfer this ability to the human-machine interaction. In this way, a wide range of automatic applications in different fields (e.g., rehabilitation, security, people identification, human behaviour analysis) can be supported. In recent years, we have focused our efforts in developing a first implementation of Kinematic, a novel multimodal framework, based on Kinect sensor, designed to support advanced human-machine interfaces. A basic step of the framework regards the discrimination of the degree of randomness related to the movements of the kinect skeleton model with respect the ordered one. This target was achieved by implementing and evaluating three mathematical feature (*curvature*, *entropy*, and *homogeneity*) working on the joint movements.

References

1. Avola, D., Bottoni, P., Dafinei, A., Labella, A.: Color-based recognition of gesture-traced 2d symbols. In: Proceedings of the 17th International Conference on Distributed Multimedia Systems, DMS 2011, pp. 5–6. Knowledge Systems Institute, Convitto della Calza (2011), <http://dblp.uni-trier.de>
2. Avola, D., Del Buono, A., Del Nostro, P., Wang, R.: A novel online textual/Graphical domain separation approach for sketch-based interfaces. In: Damiani, E., Jeong, J., Howlett, R.J., Jain, L.C. (eds.) New Directions in Intelligent Interactive Multimedia Systems and Services - 2. SCI, vol. 226, pp. 167–176. Springer, Heidelberg (2009), http://dx.doi.org/10.1007/978-3-642-02937-0_15

3. Avola, D., Cinque, L., Levaldi, S., Placidi, G.: Kinematic: A kinect based framework to support advanced human-machine interfaces - ver. i. In: Internal Technical Report in Human-Computer Interfaces, ITR-HCI 2012, pp. 1–85. Sapienza University and Univeristy of L'Aquila (2012), <http://dblp.uni-trier.de>
4. Avola, D., Cinque, L., Levaldi, S., Placidi, G.: Kinematic: A kinect based framework to support advanced human-machine interfaces - ver. ii. In: Internal Technical Report in Human-Computer Interfaces, ITR-HCI 2013, pp. 1–158. Sapienza University and Univeristy of L'Aquila (2013), <http://dblp.uni-trier.de>
5. Bhat, A., Hammond, T.: Using entropy to distinguish shape versus text in hand-drawn diagrams. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI 2009, pp. 1395–1400. Morgan K. Publishers Inc., USA (2009), <http://dl.acm.org/citation.cfm?id=1661445.1661669>
6. Champa, H., AnandaKumar, K.: Automated human behavior prediction through handwriting analysis. *Integrated Intelligent Computing* 1(1), 160–165 (2010)
7. Charbonnier, M.: The understanding of nonverbal communication in bilingual children. Ph.D. Thesis in Developmental and Social Psychology, XIX Cycle, pp. 1–126. University of Padova, Padova (2008), <http://paduaresearch.cab.unipd.it/290/>
8. Coll, R., Fornes, A., Lladós, J.: Graphological analysis of handwritten text documents for human resources recruitment. In: International Conference on Document Analysis and Recognition, vol. 2(3), pp. 1081–1085 (2009)
9. Fonseca, M., Pimentel, C., Jorge, J.: Cali: An online scribble recognizer for calligraphic interfaces. In: AAAI 2002 Spring Symposium, AAAI 2002, pp. 51–58 (2002), http://dx.doi.org/10.1007/978-1-84882-812-4_7
10. Fonseca, M.J., Jorge, J.A.: Experimental evaluation of an on-line scribble recognizer. *Pattern Recognition Letters* 22(12), 1311–1319 (2001), <http://www.sciencedirect.com/science/article/pii/S0167865501000769>
11. Grice, H.: Meaning. *Philosophical Review* 66(1), 377–388 (1957), <http://www.sciencedirect.com/science/article/pii/S0167865501000769>
12. Jakobson, R.: *Language in Literature*, 2nd edn. Belknap Press Series. Harvard University Press (1987)
13. Kala, R., Vazirani, H., Shukla, A., Tiwari, R.: Offline handwriting recognition using genetic algorithm. *International Journal of Computer Science Issues* 7(1), 16–25 (2010)
14. Kimura, F., Shridhar, M.: Handwritten numerical recognition based on multiple algorithms. *Pattern Recogn.* 24(10), 969–983 (1991), <http://dx.doi.org/10.1016/0031-32039190094-L>
15. Kinect (2013), <http://www.microsoft.com/en-us/kinectforwindows/>
16. Shannon, C., Weaver, W.: *The Mathematical Theory of Communication*, Illini Books Edn., vol. 1. University of Illinois Press (1949)
17. Skeleton (2013), <http://microsoft.com/library/microsoft.kinect.jointtype.aspx>