Characterization of postures to analyze people's emotions using Kinect technology

Julián Alberto Monsalve-Pulido [a] & Carlos Alberto Parra-Rodríguez [b]

[a] *Universidad Santo Tomás Tunja, Colombia. julian.monsalve@usantoto.edu.co*
[b] *Pontificia Universidad Javeriana, Bogotá, Colombia. carlos.parra@javeriana.edu.co*

Abstract
This article synthesizes the research undertaken into the use of classification techniques that characterize people's positions, the objective being to identify emotions (astonishment, anger, happiness and sadness). We used a three-phase exploratory research methodology, which resulted in technological appropriation and a model that classified people's emotions (in standing position) using the Kinect Skeletal Tracking algorithm, which is a free software. We proposed a feature vector for pattern recognition using classification techniques such as SVM, KNN, and Bayesian Networks for 17,882 pieces of data that were obtained in a 14-person training sample. As a result, we found that that the KNN algorithm has a maximum effectiveness of 89.0466%, which surpasses the other selected algorithms.

Keywords: analysis of emotions; recognition of postures; free software; Kinect, KNN.

Caracterización de posturas para el análisis de emociones de personas, por medio de la tecnología Kinect.

Resumen

El presente artículo sintetiza la investigación realizada en el uso de técnicas de clasificación para un proceso de caracterización de posturas de personas que tiene como objetivo la identificación de emociones (Asombro, Enfado, Felicidad y Tristeza). En este proyecto de investigación fue necesario utilizar una metodología de investigación exploratoria en tres fases donde el resultado es una apropiación tecnológica y un modelo de clasificación de emociones en personas en posición de pie, usando el algoritmo de Skeletal Tracking de Kinect basado en software libre. Se propuso un vector de características para el reconocimiento de patrones usando técnicas de clasificación como SVM, KNN y Redes Bayesianas en 17.882 datos obtenidos en una muestra de entrenamiento de 14 personas. Como resultado se evidenció que el algoritmo KNN tiene una efectividad máxima del 89.0466% superando a los demás algoritmos seleccionados.

Palabras clave: análisis de emociones; reconocimiento de posturas; software libre; Kinect; KNN.

1. Introduction

Human-machine interaction has been evolving over recent years, in particular Natural User Interface (NUI), which aims to integrate the user interaction with a computer system using natural perception. NUI can be manipulated depending on user-needs through direct or intermediate devices that create a transparent and discreet perception [1,2]. This research appropriates the Kinect technology as a Natural User Interface, and the objective is to characterize people's positions to identify emotions.

The application of this research focuses on the development of a regional problem in Boyacá, Colombia. Problems have been identified in the tourism sector due to the absence of effective mechanisms to promote and market tourist destinations resulting from the poor coordination and execution of good strategies to boost the sector. To help the development of tourism in the region it is necessary to carry out an analysis of experiential tourism as this is a new form of tourism that is based on the emotions and experiences tourists experience through interacting with the
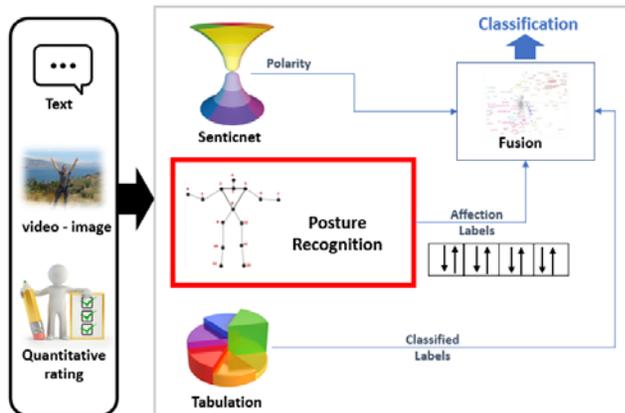
Figure 1. Integration of posture recognition to the multimodal model
Source: Authors

destination; it can be defined as an extraordinary personal experience that combines both tangible aspects that are represented in tourism products, and intangible aspects such as freedom, security, tranquility, and relaxation [3,4]. As digital media is used in all aspects of a tourist's experience, it is necessary to create automatic interpretation mechanisms that qualify emotions or feelings about a tourist product or service. These mechanisms are part of the Natural Language Processing area that can be defined as: *"Discipline focused on the design and implementation of computer applications that communicate with people through the use of natural language"* [5]. Also, when applying sentiment analysis opinions, the most complete definition is the following: It is a *"Set of computational techniques for the extraction, classification, understanding and evaluation of opinions expressed in sources published on the Internet, comments on web portals and other content generated by users"* [6].

The data to be analyzed come from various sources (social networks, travel planners, blogs, etc.) and from different types of data (text, images, sounds, videos, and numerical values) for which it is necessary to use multimodal methodologies to perform a classification and thus identify a good polarity. As a solution to the problem, we propose creating a multimodal model to generally analyze feelings or by using a fusion process that would integrate the results of text classification, postures, and quantitative qualifications of a tourist product or service. This article only documents the results of recognizing positions that, in the future will be integrated into the multimodal model by interpreting the resulting vector through a merger at the decision or identity level. In Fig. 1, the multimodal model is described.

The diversity of information and the volume of data that must be analyzed is due to the mass storage of digital media information that is generated by people and electronic devices. Digital's January 2016 report,[1] for example, mentions that 46% of the global population has Internet access (3,419 billion people), and 31% (2,307 billion people) are active users of social networks. Facebook is top on the list with 1,550 million users, Qzone has 653 million users, Tumblr has 555 million users, Instagram has 400 million,

Twitter has 320 million, baidu has 300 million, sina weibo has 222 million, and YY has 122 million users. Facebook users generate more than 500 Terabytes of content each day; there are more than 2,700 million "likes" and around 300 million photographs [7]. This data source is desired by the large marketing industries whose objective is to undertake a large-scale analysis of structured and unstructured information using BigData or Data Mining techniques.

Identifying emotions is a complex process that is the result of physical and psychological reactions that develop in behavior such as thinking and creating ambiguous natural language variables that are difficult to interpret such as surprise, anxiety, fear, and irony [8,9]. Emotions influence different ways of acting depending on our thoughts. Unexpected events that affect our normal behavior can lead to changes in behavior and decision making. To develop applications, human emotions are important in terms of usability, especially in intelligent environments, where emotions influence cognition, and, therefore, intelligence. This is particularly true when social decisions [8,10] are made. Therefore, the research focuses on identifying emotions (happiness, amazement, anger, and sadness) based on supervised postures.

The article begins with an explanation of the methodology that was applied to the investigation followed by a brief description of the state of art, and then some pattern recognition techniques are considered. Finally, we present the conclusions of the results obtained.

2. Related work

Non-verbal communication is the communication process in which messages are sent and received without words: through signs, and gestures [11]. This has no syntactic structure, so sequences of hierarchical constituents cannot be analyzed. The first impression a person makes occurs within seven seconds, and 93% of the information we communicate depends on our body language. A conversation is constituted by two parts: the verbal or conscious and the non-verbal or unconscious and emotional. This research only analyzes unconscious non-verbal conversation and focuses on body positions where gestures communicate feelings, emotions, intentions in a fraction of a second using Kinect technology [12].

When identifying emotions, several authors have investigated models that combine different areas such as psychology, biology, and neuroscience; their results include how emotions and intelligence are combined. One example is the Sentic Computing [8] research for which investigators have developed a 3D sandglass model that represents affective states through labels. Four independent but related affective dimensions are used that can potentially describe the full range of emotional experiences rooted in any of us.

Sentiment analysis in digital environments has changed the way of analyzing user-opinions through quantitative or qualitative indicators that qualify a product or service from which information generated by consumers is extracted in large volumes freely and spontaneously during the process of
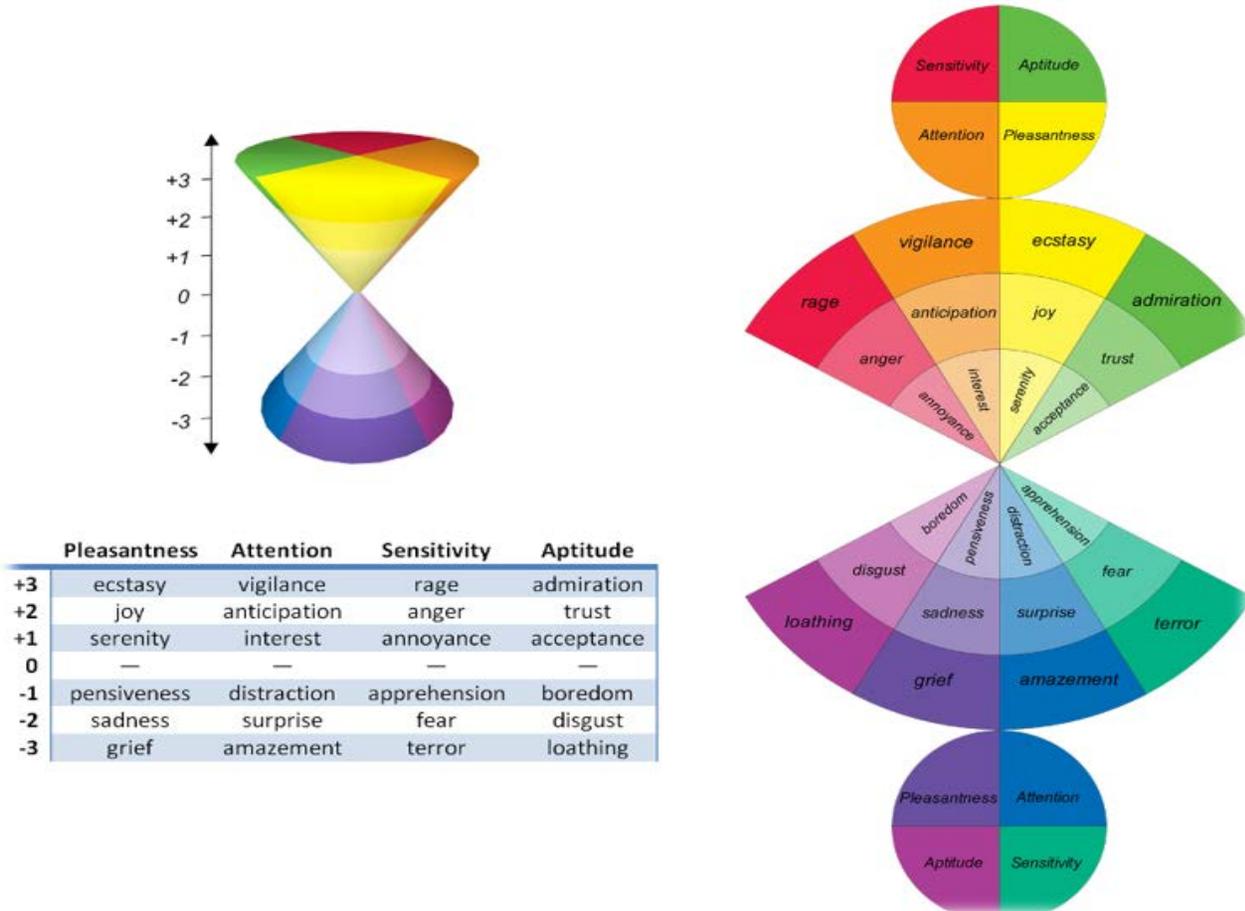
---

[1] https://www.slideshare.net/wearesocialsg/digital-in-2016

	Pleasantness	Attention	Sensitivity	Aptitude
+3	ecstasy	vigilance	rage	admiration
+2	joy	anticipation	anger	trust
+1	serenity	interest	annoyance	acceptance
0	—	—	—	—
-1	pensiveness	distraction	apprehension	boredom
-2	sadness	surprise	fear	disgust
-3	grief	amazement	terror	loathing

Figure 2. 3D model and emotions hourglass
Source: [8]

buying a product or service (the before, during, or at the end). In most cases, opinion-mining is applied to large volumes of information, specifically in text analysis, where approaches are used such as subjective lexicon, the use of the N-Gram model, and machine learning [13]. In sentiment analysis research, several methodologies that function as a basis for an effective process have been proposed. An example is the one proposed by [14] which includes five steps to develop an effective analysis. These steps are: lexicon generation, subjectivity detection, polarity detection, sentimental structure, and sentiment visualization.

Information that is currently stored in diverse sources is multimodal, and the combination of text, image, video, or sound generates a broader problem when analyzing feelings, and it is necessary to create and identify models or recognition techniques for a multimodal classification. The ability to perform multimodal fusion is an important prerequisite to successfully implement agent-user interaction. One of the main obstacles to multimodal fusion is the development and specification of a methodology that integrates cognitive and affective information from variant sources at diverse time scales and with different measurement values [15]. There are fusion techniques that will help with more effective interpretation. One example is
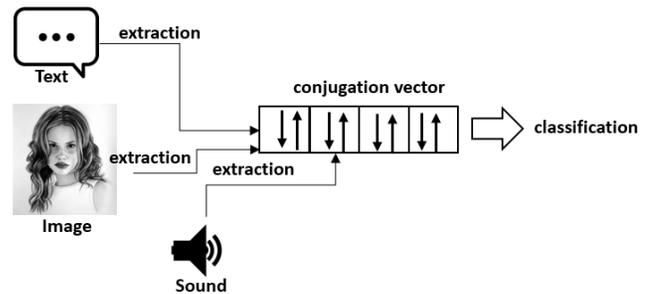

Figure 3. Fusion at the identity level
Source: Adapted from [15].

fusion on an entity level, which combines the characteristics extracted from each input channel in a vector of the conjunction before any classification operation is performed in any fusion [16]. This presents a problem that integrates highly disparate input characteristics in the synchronization of multiple inputs by unnecessary repetitions and increasing the computational resource [15].

Moreover, when considering merger at the decision level, each modality is modeled and classified independently. Unimodal results are combined at the end of the process by
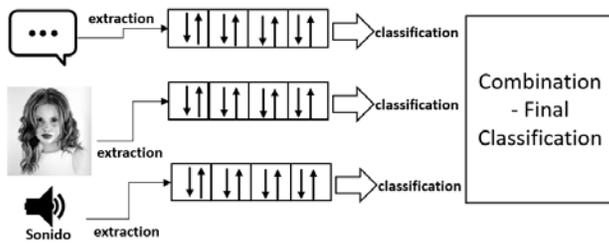
Figure 4. Fusion at the decision level
Source: Adapted from [17]

Figure 5. Physical structure of the camera
Source: [20, 21].

choosing suitable metrics such as expert rules and simple operators, including majority votes, sums, products, and statistical weighting. This merger has a decision-level benefit as the preferred method of data fusion since the different classifier errors tend not to be correlated, and the methodology is independent of the characteristics [17].

The investigation in [18] is an example of the above; the authors developed a new approach to recognizing bimodal emotions based on facial expression and speech and used the fusion method of range regression (SKRRR). Furthermore, in the research carried out by [19], a methodology was developed to analyze multimodal feelings, which aims to collect feelings in web videos and then demonstrate a model that integrates audio, visuals, and text by fusing characteristics and extracting affective information in multiple modalities. The results obtained are accurate by almost 80%; this figure surpasses all the vanguard systems by more than 20%.

An evolution of NUI was the creation of the Kinect sensor, for which the initial functionality evolved and improved the user experience in video games by using a natural interaction such as movement and voice. This has been used in research in various areas, and important results have been obtained in image and sound recognition [20]. The Kinect[2] sensor incorporates several detection components; and it contains a depth sensor, a color camera, and a matrix of four microphones that provide the entire body with 3D motion capture, facial recognition, and speech recognition capabilities.

Kinect technology has been used in the identification of emotions during multimodal analysis. [22], for example, recorded video and depth images of some students in a classroom with Kinect technology; the data were processed with techniques that tracked posture and face-to-face gestures. The results obtained related to the tutor´s identification of perception that was implicitly found in students´ physical demand and frustration. In addition, posture and gesture were correlated with student´s cognitive-affective states that tutors perceived through the implicit affective channel. [23] used Kinect technology to identify consumer reactions in a food testing kiosk with a multimodal system programmed to recognize affection to classify if a consumer likes or dislikes a tested product. The consumer´s facial expression, body posture, hand gestures, and voice were analyzed after testing the product. The result was that a classifier was created through an algorithm that assigned emotion templates using vector support machines.
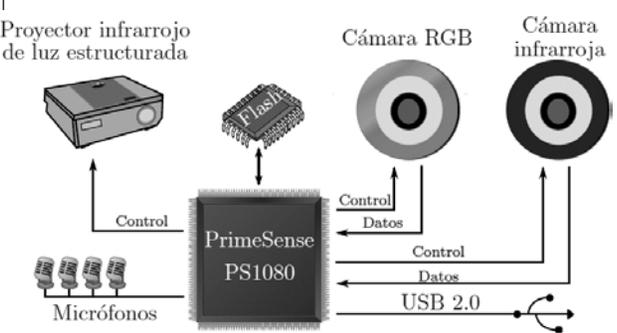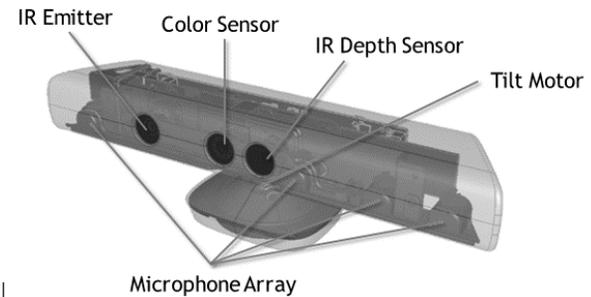
3. Materials and methods

To undertake this research, we proposed a three-phase methodology, and the main objective is to characterize emotions through postures. The first phase included detailing the state of the art regarding previous research in the following areas: feelings analysis using Kinect technology, feelings analysis in psychology, technical analysis of the Kinect device, and social analysis when identifying emotions and feelings. The second phase included the creation of a model to collect and interpret information postures generated by Kinect technology, and the characterization of four feelings using pattern recognition algorithms. Phase three concludes with the results obtained and with a proposal for future work.

One of the main functions of the Kinect sensor "Skeletal tracking" was used during the research. This is based on a skeleton tracking algorithm that manages to identify people´s body parts who are in the sensor´s field of vision. Using this algorithm, we can obtain points that refer to a person´s body parts and then identify gestures and / or postures. The sensor identifies twenty reference points (head, center of the shoulder, shoulder right, shoulder left, elbow right, elbow left, right wrist, left wrist, right hand, left hand, spine, center of hip, left hip, right hip, right knee, left knee, right ankle, left ankle, right foot, and left foot).

To track the Kinect skeleton, the depth images must be processed, human forms must be detected, and the body parts of the user in the image must be identified. Each body part is abstracted as a 3D coordinate called an articulation; a set of articulations forms a virtual skeleton for each of Kinect´s depth images, that is, 30 skeletons are obtained per second.

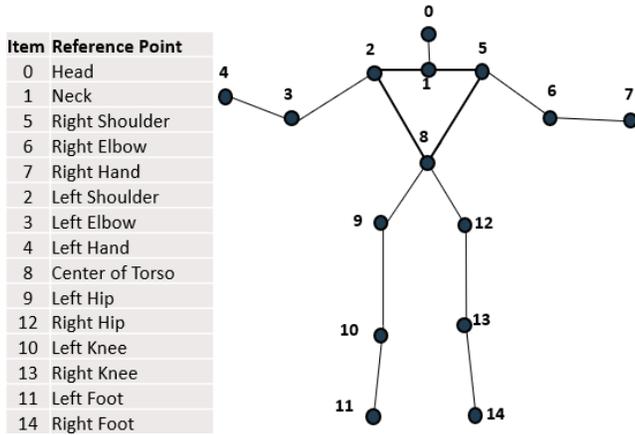| Item	Reference Point
0	Head
1	Neck
5	Right Shoulder
6	Right Elbow
7	Right Hand
2	Left Shoulder
3	Left Elbow
4	Left Hand
8	Center of Torso
9	Left Hip
12	Right Hip
10	Left Knee
13	Right Knee
11	Left Foot
14	Right Foot

Figure 6. Kinect benchmarks in OpenNI
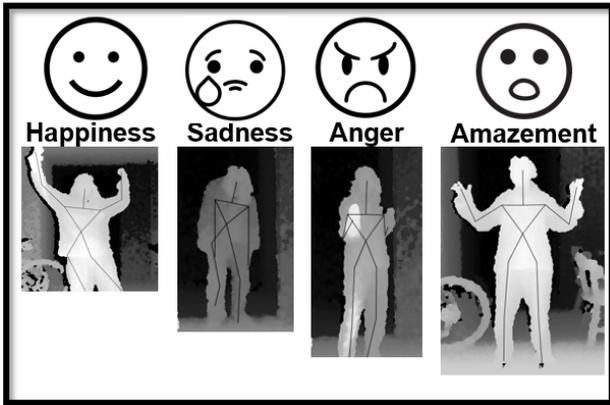Source: Adapted from [20]

Figure 7. Emotions for classifier training.
Source: Authors

The articulations generated vary according to the Kinect library used [20]. For this research, the free distribution framework OpenNI[3] (Open Natural Interaction) was used with an open source license and multiplatform development. This supports a middleware that implements characteristics of complete analysis and skeletal follow-up and an analysis of the position and follow-up of the hands or gesture recognition. The framework incorporates the NITE module. This module integrates a library that identifies each skeleton with its 15 articulations $a_i = \{x_i, y_i, z_i\}$ with $z_i > 0$ (see Fig. 6). The coordinates of this are expressed in millimeters with respect to the position of Kinect in the scene. In the Microsoft SDK and the Xbox console, five joints are added (the ankles, the wrists, and the center of the hip). The configuration and programming of the framework used was made on a GNU / Linux platform with java programming language using a three-layer architecture to integrate the storage and query of information in a posture characteristics model. Weka libraries are integrated so we could use the recognition algorithms and a Postgres database.

This research identified four feelings (happiness, sadness, anger, and amazement), which are described in Fig. 7.

Happiness is a feeling of fullness, joy, fulfillment, and enjoyment; sadness a feeling of emptiness, restlessness, decay, and demotivation; anger a feeling of annoyance and offense; and amazement a feeling of discovering something unforeseen or unexpected.

To construct of the classifier, we tested unsupervised algorithms for pattern recognition: one of the main ones evaluated was the Support Vector Machines (SVM). The main objective of this method is to find an optimal hyperplane margin using support vectors capable of forming a decision border around the learning data domain. The hyperplane is defined by:

$$F(x) = w^t . x + b = 0 \qquad (1)$$

$$\text{Margin: } min \ \frac{1}{2} \ \|w\|^2 \qquad (2)$$

where w is the normal weight vector of the separation hyperplane, b is the partial term, and x is the vector of n characteristics. The classification of a new individual x (i) is given by his or her position relative to that of the Hyperplane. SVM is based on the use of kernel functions that allow optimal separation data to be obtained. These are some of the kernel examples used in SVM [24]:

$$\text{Polynomial: } K(x,z) = \langle x, z \rangle^a \qquad (3)$$

$$\text{Radial base functions: } K(x,z) = e^{-\|x-z\|^2/2\sigma} \qquad (4)$$

$$\text{Sigmoid: } K(x,z) = \tanh(\tanh(k\langle x, z \rangle - \delta \qquad (5)$$

$$\text{Reverse quadratic multi: } K(x,z) = (\|x - z\|^{\frac{1}{2}} 2\sigma + C^2)^{-1} \qquad (6)$$

$$\text{Insertion kernel: } K(x,z) = \sum_{i=1}^{n} \min(x(i), z(i)) \qquad (7)$$

Other classification methods explored are the nearest k-neighbor, the KNN target given a vector x (i), and a set N of neighbors marked with mN. The task of the classifier is to predict the class tag of x (i) based on the class labels of set N by majority vote. In KNN, the most important parameter is the number of neighbors k. The choice of k is essential to build the nearest k-neighbors model. Thus, k can strongly influence the performance of generalization. The value of k must be large enough to minimize the probability of error, but also reasonably low compared to mN or the size of the set N [25].

Furthermore, the Bayesian classifier is a supervised learning algorithm and a statistical method based on Bayes theorem. Given a sample x (i) and a set of training samples S, each with its class label Cl with l ∈ [1, L] and L being the number of classes, the classifier predicts that x (i) belongs to the class with the highest posterior probability:

---

[3] http://openni.ru/reference-guide/index.html?t=index.html

$$P(C_l \mid x^{(i)}) = \frac{P(x^{(i)} \mid C_l)P(C_l)}{P(x^{(i)})} \qquad (8)$$

4. Results

To develop the system, we used free software tools, and the objective was to integrate a scalable solution to continue with the research without any limitations regarding proprietary licenses. The following tools were used in the model: was Processing[4]; which is the core of flexible and adapted solutions to learn about visual arts in digital environments; OpenNI, as previously mentioned, stands for Open Natural Interaction, and is a tool that focuses on the certification and improvement of the interoperability of the natural user interface and the organic user interface for natural interaction devices for device applications such as the Kinect [20]. To recognize the skeleton of the 15 reference points, the Simple-Openni library was used and libraries extracted from WEKA[5] were used to apply the pattern recognition algorithms. The model´s integral solution was developed with the Eclipse development interface using the Java language and libraries.

When constructing the model, we identified the existing protocols to collect information on people´s movements using a video device; the Protocol of positioning marked Davis [26] is taken as a basis for this research since it is one of the most commonly used in biomechanics. It consists of using the anatomical points of bony eminences depending on the movement analysis that must be analyzed. The fifteen points obtained from the skeleton that are captured in the camera are stored in a "Capture" table with the structure (node, PosX, PosY, PosZ); they are then consulted during the training and classification process.

The classifier starts with a capture process using the Kinect camera for an initial calibration; it then starts the training process so that the classifier interprets the data input using the Kinect camera (see Fig. 8).

The general model begins with capturing the information through the Skeletal Tracking´s 15 points that are generated by Kinect; the information is then stored on a database created in Postgres where the classifier can consult it. For the training process, we created a file.arff with a specific structure to be verified by the Weka library (see Fig. 9).

Weka captures information and generates a classification model according to the selected algorithm (SVM, KNN, NB) and creates the final classification according to the Skeletal Tracking input information based on the proposed model (see Fig. 10).

For the model´s tests, 17,882 training data were used from fourteen randomly selected people who simulated basic emotions (astonishment, anger, happiness, and sadness). The coordinates xi, yi, and zi (0 <i <14), depend on the size and position of the person in the scene, and the following feature vector was identified:

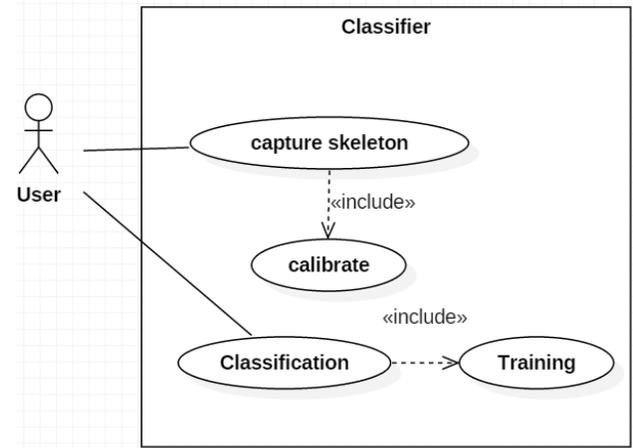[x0,y0,z0,x1,y1,z1,x2,y2,z2,x2,y2,z3,x3,y3,z3,x4,y4,z4,x5,y5,z5,x6,y6,z6,x7,y7,z7]

Figure 8. Use case diagram
Source: Authors


```
@RELATION gestos

@ATTRIBUTE nodo REAL
@ATTRIBUTE x    REAL
@ATTRIBUTE y    REAL
@ATTRIBUTE z    REAL
@ATTRIBUTE class    {Asombro,Enfado,Felicidad,Tristeza}

@DATA
0,325.2185,71.8983,2090.5146,Asombro
```
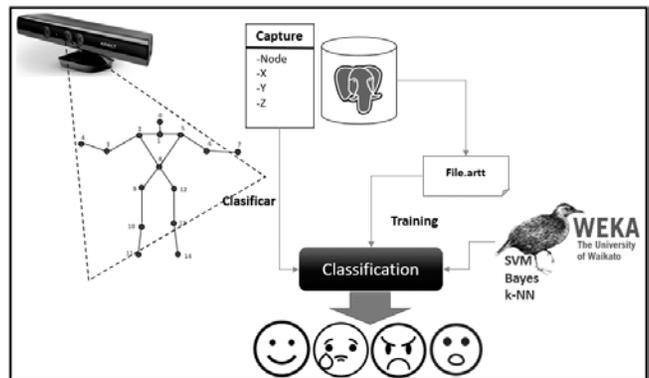Figure 9. ARFF file Structure
Source: Authors

Figure 10. Posture characterization proposed model.
Source: Authors

Table 1.
Distribution of data on emotions for training purposes.

Emotion	No data
Amazement	4076
Anger	4680
Happiness	4851
Sadness	4275

Source: Authors

---

[4] https://processing.org

[5] https://www.cs.waikato.ac.nz/ml/weka/

Table 2.
Performance of the classifiers.

Training data	SVM RBF	KNN -	NB -
10	80.5894 %	85.807 %	81.6575 %
20%	77.5269 %	89.0466 %	78.9808 %
40%	79.6812 %	88.5357 %	80.9582 %
60%	80.7773 %	87.7254 %	81.8817 %
80%	80.8166 %	86.5213 %	82.019 %

Source: Authors

For the classifiers analysis tests, five different types of training data were used (see Table 2), The first included ten random samples and had important results with a KNN that was 86.1928% effective. The second had 20% training data, and it was shown that the KNN algorithm is 89.0466% effective, which significantly surpasses the other classifiers. In the other training tests (using 40%, 60%, and 80% of total data for training) the KNN algorithm remains above the rest of the classification algorithms.

5. Conclusions

This study showed that the KNN algorithm performs more completely than SVM and NB and has a maximum effectiveness of 89.0466% for the set of selected data. It can also be used as a foundation to develop applications that recognize basic emotions (astonishment, anger, happiness, and sadness) using Kinect technology. The algorithms based on SVM and NB have a lower percentage than the KNN, but future studies could consider their effectiveness since this can be improved with unsupervised learning.

The applied vector of characteristics can vary for classifications of more complex negative emotions, for example ((+) rage, anger, annoyance, (-) apprehension, fear, terror). This means that it would be necessary to identify additional points of the skeleton, which would increase the complexity when recognizing emotions.

The data results of this investigation will be integrated into a general multimodal sentiment analysis model that considers the tourist area in the department of Boyacá Colombia. It will be merged using a conjunction vector for a final sentiment classification that analyzes data types such as text and images.

Using free distribution tools for research processes creates a channel of collaborative help communication. This takes advantage of solutions from communities around the world to make contributions to science without the need to reinvent the wheel or have a large budget to carry out a research project. For this research, we used tools including OpenNI, NITE, Weka, Java that helped classify emotions, and the result was a technically functional and economically viable product.

Human-computer interaction (HCI) has improved over recent years; the Natural User Interfaces (NUI) have created usability solutions where valuable information is stored and can be analyzed to identify users´ emotions. Kinect is not only for use in video games; it can also be applied in various areas of knowledge due to its great potential for innovative hardware and its increasing usefulness in research.

Referencias

[1] Mann, S., Intelligent image processing. IEEE, John Wiley & Sons, Inc., 2002. DOI: /10.1002/0471221635

[2] Valli, A., Natural interaction white paper, 2007.

[3] Rivera-Mateos, M., El turismo experiencial como forma de turismo responsable e intercultural, en: García-Rodríguez, L., Roldán-Tapía, A.R., Eds., Relac. Intercult. en la Divers., 2013, pp. 199-217.

[4] Smith, W.L., Experiential tourism around the world and at home: definitions and standards, Int. J. Serv. Stand., 2(1), 1 P, 2006. DOI: 10.1504/IJSS.2006.008156

[5] Dale, R., Moisl, H. and Somers, H.L., Handbook of natural language processing, Marcel Dekker, 2000.

[6] Cambria, E. and Hussain, A., Sentic album: content-, concept-, and context-based online personal photo management system, Cognit. Comput., 4(4), pp. 477-496, 2012. DOI: 10.1007/s12559-012-9145-4ch

[7] Simon-Kemp, W.A.S., Digital in 2016, 2016.

[8] Cambria, E., Livingstone, A. and Hussain, A. The Hourglass of Emotions. In: Esposito, A., Esposito, A.M., Vinciarelli, A., Hoffmann, R. and, Müller, V.C., (eds.), Cognitive Behavioural Systems. Lecture Notes in Computer Science, vol 7403. Springer, Berlin, Heidelberg. 2012. DOI: 10.1007/978-3-642-34584-5_11

[9] Minsky, M., The emotion machine: commonsense thinking, artificial intelligence, and the future of the human mind, 2007.

[10] Vesterinen, E., Affective computing. Pattern Analysis and Applications, 1(1), pp. 71-73, 1998.

[11] Siegman, A.W. and Feldstein, S., Nonverbal behavior and communication. L. Erlbaum, 1987.

[12] Pons, C., Comunicación no verbal. Barcelona: Editorial Kairós, 2015.

[13] Kaur, A. and Gupta, V., A survey on sentiment analysis and opinion mining techniques, J. Emerg. Technol. Web Intell., 5(4), pp. 367-371, 2013. DOI: 10.4304/jetwi.5.4.367-3

[14] Gamon, M., Aue, A., Corston-Oliver, S. and Ringger, E., Pulse: mining customer opinions from free text, Springer, Berlin, Heidelberg, 2005, pp. 121-132. DOI: 10.1007/11552253_12

[15] Poria, S., Cambria, E., Hussain, A. and Bin Huang, G., Towards an intelligent framework for multimodal affective data analysis, Neural Networks, 63, pp. 104-116, 2015. DOI: 10.1016/j.neunet.2014.10.005

[16] Kapoor, A., Burleson, W. and R Picard, W., Automatic prediction of frustration, Int. J. Hum. Comput. Stud., 65(8), pp. 724-736, 2007. DOI: 10.1016/J.IJHCS.2007.02.003

[17] Lisetti, C.L., Pattern Analysis & Applic, 1, J. Wiley, 1998, 71 P. DOI: 10.1007/BF01238028

[18] Yan, J., Zheng, W., Xu, Q., Lu, G., Li, H. and Wang, B., Sparse Kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech, IEEE Trans. Multimed., 18(7), pp. 1319-1329, 2016. DOI: 10.1109/TMM.2016.2557721

[19] Poria, S., Cambria, E., Howard, N., Bin Huang, G. and Hussain, A., Fusing audio, visual and textual clues for sentiment analysis from multimodal content, Neurocomputing, 174, pp. 50-59, 2016. DOI: 10.1016/j.neucom.2015.01.095

[20] Benhumea, H.S., Interfaz de lenguaje natural usando Kinect. Unidad Zacatenco, 2012.

[21] Zeev-Zalevsky, J.G., Shpunt, A. and Maizels, A., Method and system for object reconstruction [online]. [date of reference: Sept. 04th of 2016]. Available at: https://patents.google.com/patent/US20170004623

[22] Grafsgaard, J.F., Fulton, R.M., Boyer, K.E., Wiebe, E.N. and Lester, J.C., Multimodal analysis of the implicit affective channel in computer-mediated textual communication, Proc. 14th ACM Int. Conf. Multimodal Interact., pp. 145-152, 2012. DOI: 10.1145/2388676.2388708

[23] Patwardhan, A.S. and Knapp, G.M., Multimodal affect analysis for product feedback assessment, 2013, pp. 178-187.

[24] Choubik, Y. and Mahmoudi, A., Machine learning for real time poses classification using kinect skeleton data, in: 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), 2016, pp. 307-311. DOI: 10.1109/CGiV.2016.66

[25] Shum, H.P. H., Ho, E.S.L., Jiang, Y. and Takagi, S., Real-time posture reconstruction for Microsoft Kinect, IEEE Trans. Cybern., 43(5), pp. 1357-1369, 2013. DOI: 10.1109/TCYB.2013.2275945

[26] Davis, R.B., Ounpuu, S., Tyburski, D. and Gage, J.R., A gait analysis data collection and reduction technique, Hum. Mov. Sci., 10(5), pp. 575-587, 1991. DOI: 10.1016/0167-9457(91)90046-Z

J.A. Monsalve-Pulido, obtained his BSc. in Systems and Computing degree in 2005 and his MSc. degree in Free Software in 2008 from the Universidad Autónoma de Bucaramanga, Colombia and the Universitat Oberta de Catalunya, Spain respectively. Heb is currently undertaking his PhD in engineering at the Pontificia Universidad Javeriana, Colombia. Also, he works as a research professor in the Systems Engineering Faculty at St. Thomas University in Tunja, Colombia, and he is a leader of the GIDINT research group. His research has included big data, information systems, and learning networks, and his research interests include software engineering and data mining.
ORCID: 0000-0002-0706-5881

C.A. Parra-Rodríguez, is a full professor in the Department of Electronics at the Pontificia Universidad Javeriana, Colombia and the director of the School of Engineering. prof. Parra has experience in the application of data science in robotics and perception for robotics and has published over fifty publications.
ORCID: 0000-0001-7209-6150