IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 October 2022, accepted 17 November 2022, date of publication 5 December 2022, date of current version 9 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3226696

RESEARCH ARTICLE

A Prototype for Mexican Sign Language Recognition and Synthesis in Support of a Primary Care Physician

CANDY OBDULIA SOSA-JIMÉNEZ[1], HOMERO VLADIMIR RÍOS-FIGUEROA[2], AND ANA LUISA SOLÍS-GONZÁLEZ-COSÍO[3]

[1]School of Statistics and Informatics, University of Veracruz (UV), Xalapa, Veracruz 91020, Mexico
[2]Research Institute in Artificial Intelligence, University of Veracruz (UV), Xalapa, Veracruz 91097, Mexico
[3]School of Sciences, National Autonomous University of Mexico (UNAM), Mexico City 04510, Mexico

Corresponding author: Candy Obdulia Sosa-Jiménez (cansosa@uv.mx)

This work was supported in part by the National Council of Science and Technology of Mexico (CONACYT) under Grant 388930.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ph.D. Research Program Committee of the Research Center in Artificial Intelligence of the University of Veracruz.

ABSTRACT Few hearing people know and use Mexican Sign Language (MSL). Consequently, this is the main barrier between people who having total or partial hearing loss and hearing people. This study proposes a system that recognizes and animates in real time a set of signs belonging to the semantic field of general medicine consultation services. Therefore, a linkage between a hearing doctor and a deaf patient can be established in a non-intrusive way and with easy dynamic interaction. Our main contribution is a bidirectional translator system for Mexican Sign Language in the context of primary care health services, in addition to basic signs to fingerspell alphabet and numbers as a complement to provide personal information such as name, age, etc. The recognition module uses a Microsoft Kinect sensor to obtain sign trajectories and images to feed hidden Markov Models (HMMs) for processing sign samples in real time. The experiments showed the recognition of 82 different signs by 22 participants. As a result, accuracy and F1 scores average rates of 99% and 88%, respectively, were obtained.

INDEX TERMS Mexican sign language, depth sensor, dynamic sign language recognition, medicine consultation services, sign language synthesis.

I. INTRODUCTION

Sign languages use observable gestures in space for communication. The International Sign System (ISS) is a pidgin, a simplified language created and used by people from communities that do not have a common language (e.g., Deaflympics); however, it is limited.

There are more than 137 different sign languages worldwide that communicate the expressions of each language [21]. Every sign language evolves by integrating new signs, eliminating old ones, and generating different ones that refer to the same words or concepts. Moreover, there are variants of sign languages by language, country, and geographical region. For instance, Spanish Sign Language (SSL) and Mexican Sign Language (MSL) are very different, although both refer to Spanish language.

Hearing loss refers to a loss greater than 40 dB in adults and greater than 30 dB in children. Most people with hearing loss live in low- income and middle-income countries [47].

According to data from the Mexican National Survey of Demographic Dynamics, among the population aged ≥3 years with limited listening, only 4.5% use hearing aids. Of these people, nine out of ten have problems accessing this type of aid. Consequently, they have fewer opportunities to communicate and interact with others under equal conditions. This situation affects the integration of the population in social, educational, and employment areas [26].

Mexico has 120 million inhabitants, and the prevalence of disability in Mexico is 6% (7.2 million inhabitants), of which 33.5% have hearing disabilities. Only 83.3% of

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

the population with disabilities in Mexico are affiliated with health services. The main causes of this disability are birth, illness, accidents, and advanced age [26].

In the Mexican educational system, none of its educational levels include the teaching of MSL, so the majority of the Mexican population does not have knowledge about it, unless it is for their own interest and they carry out studies afforded by themselves, which hinders deaf inclusion and communication in daily life; in addition, there is not enough standardized material developed by government institutions in MSL for access to public services such as education, health, justice, etc.

The main problem is related to the lack of knowledge of medical staff to communicate with deaf patients in Mexico, given the great needs of people with hearing loss, this study focuses on providing a technological tool that supports communication between hearing and deaf, facilitating communication during a consultation with a Primary Care Physician, and promoting inclusion in society, so deaf people can have better access to health services.

The remainder of this study is organized as follows. Section II presents related works on static or dynamic sign language recognition and studies of MSL recognition and synthesis. Section III summarizes the contributions of this study. Section IV presents the apparatus, sensors, hardware, and software used. Section V explains our approach, which includes a synthesis module, an MSL recognition module, and a medical graphical user interface. Section VI presents the components of evaluation, vocabulary, participants, experiments, and metrics. Section VII explains the results of the recognition of alphabetic vocabulary, numbers, and medical consultation service words. Section VIII presents the discussion of this study. Finally, the conclusions and future work are presented.

II. RELATED WORKS

The development of low-cost depth sensors has opened new avenues for Human-Computer Interaction. With new depth cameras and sensors, gesture recognition has gradually shifted from a 2D technique to 3D analysis. It is difficult for computers to extract gestures from a target image because of the unpredictable environment and complex background. The hand is a key body part in gesture recognition, particularly in people with hearing disabilities. A hand is a flexible body with more than 20 degrees of freedom; therefore, people can perform the same gesture by starting the hand movement at different positions.

For instance, Kinect has been widely used to recognize several sign languages, such as American Sign Language (ASL) [28], [30], which studies x, y, and z coordinates, and German Sign Language (GSL) [20], which analyzes the speed, position, and distance between hands. The Albanian Sign Language [13] studied hand contours and hand centers. The Turkish Sign Language (TID) [42] used the trajectories of sign responses. Chinese Sign Language [5] analyzes hand posture frames from video and 3D trajectories, and India Sign Language [18] analyzes angles, speed, and curvature fingertips.

Moreover, these studies have employed three main techniques to recognize signs: the hidden Markov model [20], [30], dynamic time warping [13], [42], neural networks [28], [34], [49], and support vector machines [49]. Consequently, these studies achieved accuracy rates between 83% [42] and 97% [20], [28].

Additionally, Kinect has been used jointly with the Leap Motion controller to detect Indian Sign Language [19]. Specifically, they used depth and image data as inputs to a hidden Markov model and neural network classifiers, achieving an accuracy of 94.55%.

In the design of gesture-based user interfaces, continuously recognizing complex dynamic gestures is a challenging task because of the high dimensionality, ambiguous semantic meaning, and presence of unpredictable non-gesture body motions.

Sign languages have not been the only object of study, but also other types of hand gestures, such as movements to direct an orchestra [44], commands to direct a Smart TV [48], or activities of daily life [29].

Other approaches do not use depth sensors, such as video cameras in conjunction with special gloves [25], time-of-flight cameras [24], web cams [37], or inertial sensors that measure acceleration and angular velocity [46].

Regarding the progress in the automated recognition and synthesis of MSL, there are few works that are described below. However, none of them has considered vocabulary from a primary care medical consultation service or the recognition and synthesis of MSL.

Reference [23] presented the use of MSL for the control of a service robot using eight alphabetic letters. In addition, they recognized 23 static alphabet letters in the color images. Their method uses active contours to segment and shape the signature for description, and a neural network for classification. The dataset was acquired under controlled conditions with a background in contrast to the hand. One person generated all the sign samples. A recognition rate of 95.8 % was achieved.

Reference [32] presented two methods for recognizing the MSL alphabet: One method uses a controlled background and illumination to aid Red, Green, Blue (RGB) image segmentation. Then, Hu moments and other descriptors were used to obtain 2D invariance in translation, rotation, and scale. The dataset was generated for two subjects. The author reported a 100% classification for 20 letters of the MSL alphabet. The second method uses a Kinect camera to segment the hand depth. A 2D template for each letter was then transformed using evolutionary matrices. In this case, 25 letters were recognized with 90% accuracy.

Reference [43] presented a method to estimate the 3D posture and recognize 27 letters of the MSL alphabet with 90.27% precision using a dataset generated by one subject.

Reference [2] presented a Mexican speech-to-sign language system. After the speech recognizer, each character, word, or sentence was performed using an animated avatar.

In this study, 70 words were selected from the MSL dictionary without a particular semantic domain.

Reference [38] recognized 24 static MSL letters from RGB images and achieved a 98.53% recognition rate. The dataset was generated by one person performing each sign five times. Then, the background was controlled. Their method used 2D normalized geometric moments and a neural network.

Reference [12] recognized five vowels and two consonants from MSL alphabet with 76.19% precision. Their method uses images, depth, and skeleton from a Kinect sensor, with random forest and a neural network.

Reference [39] recognized 24 letters of the MSL alphabet from RGB images, with a 95% recognition rate. Their method employs Jacobi-Fourier moments, which are two-dimensional (2D) rotation invariants. The dataset was generated for a single user.

Reference [40] classified 24 static signs of MSL alphabet from RGB images by using normalized moments and a neural network. These moments are invariant to translation and scale but not 2D rotation. The dataset was generated for a single user. Their method achieved a recognition rate of 95.83 %.

Reference [4] classified 249 dynamic MSL words in a controlled environment from 22 people using black cloth and a black background. The images were processed, and the hands were segmented to extract 743 geometric, textural, and color features. Feature selection was performed using a genetic algorithm. Training and test classification were performed with a support vector machine to achieve an average accuracy of 97 %.

Reference [41] present a real-time MSL recognition system. It operates indoors without controlled background or clothes. The dataset was generated by ten participants performing each word five times. In total, 33 dynamic words were classified, with 86% sensitivity and 80% specificity.

Reference [16] recognized five letters and five numbers, with an F1 score of 95%. A database was generated by 100 participants who performed each sign once. This method works with 3D point clouds and uses 3D Haar features.

Reference [14] used data time warping to recognize 20 dynamic words from Kinect data, achieving an accuracy of 98.57%. The dataset was generated by 35 participants, who performed each sign once.

Reference [11] present a system to recognize 249 dynamic MSL words from 17 semantic categories. The dataset was generated by 11 people and producing 2480 videos. The acquisition environment was controlled using black background land, and black cloths were used to aid image segmentation. An average precision of 96.27% was obtained using geometric features and an SVM. For the nine words, the accuracy was less than 70%.

Reference [35] developed a system to recognize 75 dynamic words from nine categories and obtained an average accuracy of 94.9%. Some important features of their work are that they extract information from both hands, the body, and facial expressions. In addition, they interpreted 20 sentences from a medical context and obtained an average accuracy of 94.1% for this case. Their medical context was related to hospital emergencies. The dataset was generated by six participants, and the vocabulary had an average of 35 samples per word. Their method uses the MediaPipe and OpenCV libraries to extract and standardize geometric features from the body, hands, and facial expressions. Then, three 2D convolutional neural networks (CNN) are used for encoding, followed by concatenation and a fully connected layer. Finally, an Hidden Markov Model (HMM) was applied.

III. CONTRIBUTION

Most studies on sign language recognition identify static gestures belonging to alphabetic letters. Moreover, the Kinect sensor has been widely used for sign language recognition because it can identify several body joints and hand positions. Nevertheless, Kinect does not recognize the fingers individually, which is required to identify the signs belonging to MSL.

The main contributions of this study are:

- Our system recognizes 31 static and 51 dynamic MSL gestures, including medical signs, letters, and numbers (Table 1, FIGURE. 12 and FIGURE 13). These signs were selected as the basic and necessary words in a general medicine consultation service by a group of three primary care physicians.
- Recognition has been focused on a semantic field of importance, which is the vocabulary used in a primary care consultation service. This is a significant contribution to the recognition of MSL, because this context has not been studied in MSL.
- Our system includes the synthesis of MSL, which allows the doctor to type text, which is converted to animated MSL using a signing avatar. Consequently, a bidirectional patient-doctor communication is obtained for this semantic context. No other work has reported the recognition and synthesis of sign language in a general medicine consultation service.
- The use of an avatar that performs the signs reinforces communication because it is difficult for many deaf people in Mexico to identify words written in Spanish; they identify the signs better because it is what they know and handle to communicate between them.
- The proposed solution is highly economical because it is installed on a computer that the doctor already has in his office, and only an inexpensive sensor that is easy to acquire is added. Unlike approaches such as the one presented in [31], where they only translate the ASL alphabet in speech, sign languages are not just the alphabet; it is necessary to analyze other elements such as hand movements while making the sign, the body area where the sign is made to express whole words not only fingerspell them, spelling is not practical; it is only used as the last resort when the sign for a word is not known.

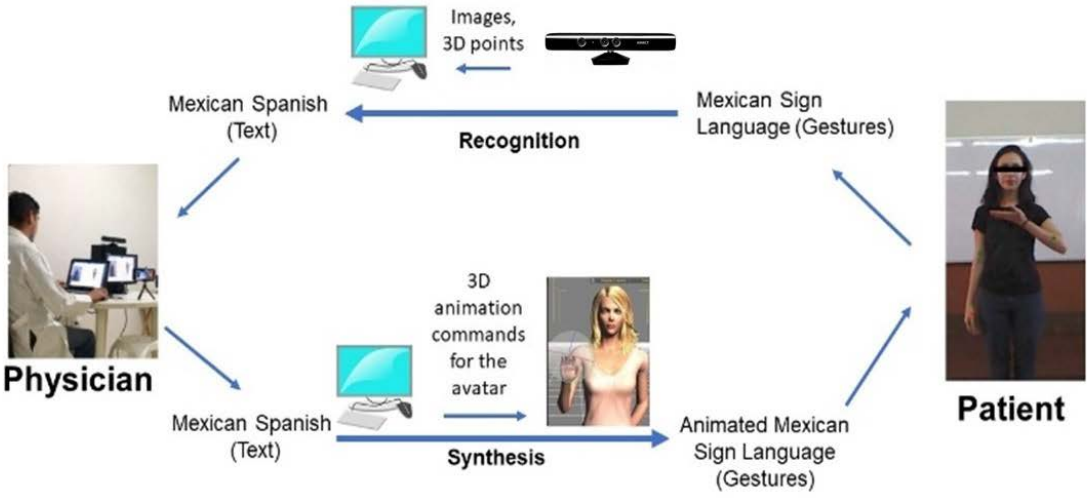

FIGURE 1. Diagram showing the two communication paths between patient and physician: 1) The MSL performed by the patient is recognized and converted to Mexican Spanish text, which is read by the physician; and 2) The physician writes Mexican Spanish text, which is converted into MSL and performed by a signing avatar that is understood by the patient.

IV. APPARATUS

The Microsoft Kinect v1 device was used in this study because of its advantages of low cost, easy data acquisition, software libraries, depth cameras, and RGB cameras.

The Kinect device developed by Microsoft is a sensor that can be used as a signal receiver to collect video and audio data. This depth sensor is known for its excellent performance in detecting each human joint position represented by a corresponding three-dimensional data point, which is used for easy human gesture recognition [7], [8], [9], [15], [45].

Our study uses point clouds captured by the Kinect depth sensor. A point cloud is a set of vertices in a three-dimensional coordinate system. These vertices are usually identified as X, Y, and Z coordinates and are representations of the external surface of an object. Point clouds are typically created using a three-dimensional laser scanner. Based on the 3D points, an object can be tracked. The points captured in time represent the sequence of movements made by the subject's hands. In our study, Kinect for Windows SDK v1.8, and Microsoft Visual Studio 2015 were used to process the data.

The experimental environment for implementing and testing the recognition system consisted of the following hardware:

- Computer 1: A notebook with a processor Intel Core i7-2630QM. A memory capacity of 6 GB of RAM and a Windows 7 operating system with a 64-bit architecture, Video Card Radeon HD 6770M with a total graphics memory of up to 1024 MB.
- Computer 2: Notebook with Intel Core i5-5200U processor A memory capacity of 6 GB of RAM and Windows 8.1 operating system with a 64-bit architecture. Intel HD Graphics 5500 with a total graphics memory of up to 3036 MB.
- Microsoft Kinect v1 was used as the real-time image capture technology. It generates 30 fps and provides RGB and depth data.

V. PROPOSED APPROACH

FIGURE 1 presents our system, which is composed of two main modules: a synthesis module, which converts a text input into a dynamic MSL using a signing avatar, and a recognition module, which interprets the sign language of the patient and generates a text output composed of words, letters, or numbers read by the doctor.

The fundamental aspect of the application of hidden Markov models (HMMs) is the use of a pattern structure model. In this way, the knowledge that has a priori on sign language structure can be incorporated into the modeling process, allowing a deeper analysis.

In addition, HMM's are easily implementable and constitute a highly flexible modeling tool that was initially used in the field of automatic speech recognition, which has found numerous applications in diverse scientific and technical areas, highlighting the possibilities they can offer for the analysis of complex spatial patterns, since they allow the incorporation of a priori information on the analyzed system into the modeling process.

The HMM was also selected because when performing a sign in space, the hands move through certain milestone positions. In addition, several signs share the initial movement but then differ. This situation is similar to that of written and spoken language understanding.

A. SYNTHESIS MODULE: SIGNING AVATAR

Deaf people have different ways of communicating with one another. They employ space, hands, and body in a different linguistic syntax than that of Spanish. Therefore, it is important to configure a signing avatar - an animated 3D model of a virtual human that presents messages in sign language with the same characteristics. Signing avatars requires extremely fine motor movements and minimal collision-avoidance routines.

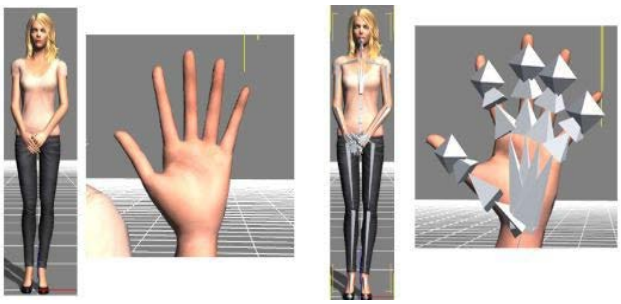

FIGURE 2. 3D Model of the signing avatar. The left image shows the avatar and one of her hands. The right image shows the bones separated by the joints that will allow movements with the same degrees of freedom as the human body.

To provide a natural appearance to the signing avatar, we used a 3D model and its bone structure, which allowed accurate folding (FIGURE 2). The visible part of the signing avatar consists of a three-dimensional polygonal mesh with associated skin, clothing, and hair textures to create a realistic representation of a human being.

Regarding the automatic translation of written text into sign language without human intervention, motion capture (MoCap) was employed to capture body, limb, and head movements in 3D space. Specifically, MoCap technology enables the recording of natural and accurate descriptions of signs and sentences performed by a signer. However, MoCap equipment is not capable of tracking finger movements properly. Consequently, these movements must be corrected during the postproduction process.

We used the motions captured by Kinect to animate the arms of the virtual character. To refine the finger movements, we made additional animations by moving each finger bone to the required sign position (FIGURE 3).

B. MSL GRAMMAR

All languages have grammar to structure words and make sentences. MSL has special grammar; for example, verbs are used in the infinitive. The structure of a sentence depends on the type of the verb used. The most common structure is Subject-Verb-Object (SVO) [1], [6], [22]. However, depending on the type of verb, other constructions are possible, such as OSV, VOS, VSO, OVS and SOV [6]. In this study, we concentrate on basic structures in the medical context and with a small vocabulary. FIGURE 4 presents the following sentence in MSL: I speak only a little Spanish.

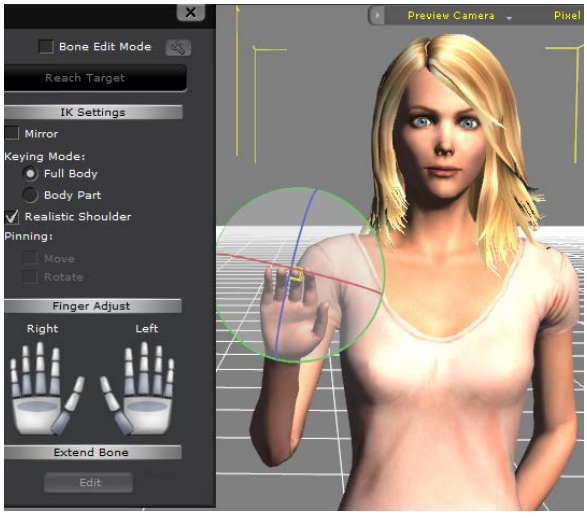

FIGURE 3. Interface to adjust finger positions of the signing avatar for each sign.

FIGURE 4. Translation of the sentence: I speak only a little Spanish. From Spanish into MSL using its specialized grammar. I only speak Spanish little (Gesture images are interpreted continuously from left to right and from top to bottom).

C. MEXICAN SIGN LANGUAGE SYNTHESIS

The MSL synthesis is composed of two main modules:

1. The translation module analyzes the input text and converts it into intermediary representation. This representation uses glosses to represent the signs accompanied by information associated with inflection. The intermediary representation provides a straightforward mapping to the signs and defines animation commands that are interpreted by the animation module to control the movements of the avatar.

The input text was then decomposed into words. If a word is not part of the vocabulary, it is expressed letter by letter in MSL. If a word is recognized, the information is communicated to the animation module to perform the associated MSL sign. The translation module was coded in the C# language using MS Visual Studio.

2. The animation module is responsible for animating the avatar and presenting the outcome on a device screen.

The avatar/virtual human was created in iClone, a tool that facilitates virtual human animation. Each letter and some medical context vocabulary were modeled by moving the body of the virtual character. Word models were created for vocabulary from [3] and [36]. (FIGURE 5 and 6).

Once the animations were modeled, we decided to use Unity, because it is a game development platform that allows us to easily export the system to Windows, Android, and iOS.

FIGURE 5. Medical context vocabulary.

To import the animations to Unity, they were first exported to the FBX format using 3DXchange.

Then, we imported a virtual human into Unity. To automate the system, we created an animated state machine (see FIGURE 7). This machine contains an animation alphabet controlled by a C# script.

When the user types words, the virtual human displays the corresponding sign letters or words in MSL.

D. MSL RECOGNITION, SIGNAL ACQUISITION

Two inputs were obtained from the two cameras of Kinect: i) the X, Y, and Z coordinates were obtained from the hand joints provided by the Kinect's infrared (IR) camera, and ii) a color picture of each hand was obtained from Kinect's RGB camera. These two inputs were taken simultaneously (FIGURE 8).

The contour of the hands was obtained to recognize its shape and compare it with the database. To obtain trajectory coordinates from the sign made with the hands, the user's initial position is saved immediately before starting the capture of coordinates. Users were asked to stand in front of the sensor, with their arms relaxed at their sides without moving.

The images were saved in a jpg format of 95 $\times$ 95 pixels. Depth filtering is obtained from the joints identified by skeletal tracking, saving the image whose coordinates of depth are smaller than those corresponding to the articulation of the wrists, and making a precise segmentation of the hands.

It should be mentioned that hand pictures used to detect dynamic signs are the front view poses of the hands. Although a sign could not correspond to the image of alphabet letters because the sign can be performed with the letter posture starting at different positions (e.g., downwards, upward, vertical, or horizontal). To overcome this limitation, different databases were acquired from different hand angles to identify alphabetic letters and medical context words.

For feature extraction, the images were preprocessed, and the dimensionality of the data was reduced.

E. MSL RECOGNITION, PREPROCESSING

First, the images were captured in RGB color. These images were then binarized (black and white) to detect the blobs.

The hand posture contour is represented as a set of X and Y coordinates (FIGURE. 9). The top and bottom edge methods were used to scan each column of the blob and find the upmost and lowest points, which were added to the lists. These points (X and Y coordinates) are the inputs to a hidden Markov model for classifying the hand shape involved in the sign.

Additionally, the X, Y, and Z coordinates of the centroid and the left and right hands describing the signal movement were saved in XML format for retrieval during the training and testing processes (FIGURE. 10).

As the heights and arm lengths of the volunteers who participated in the experiments were different, the coordinates were normalized in a range between 0 and 1 to reduce the variance in the measurements produced by each person. Subsequently, the coordinates corresponding to the signal were smoothed to reduce the noise caused by involuntary movements performed by the users.

A locally estimated scatterplot smoothing (LOESS) algorithm was used to fit smooth surfaces to the data. It uses locally weighted linear regression to smoothen the data. This process is weighted because a regression weight function is defined for the data points contained within the span. Larger smoothing values (h) produce the smoothest functions that move the least in response to fluctuations in data. The smaller the value of h, the closer is the fit of the regression function to the data. Using too small a value for the smoothing parameter is undesirable because the regression function will start to capture the random error in the data. Useful values of the smoothing parameter are generally in the range of 0.25 to 0.5 for most LOESS applications. Testing values between 0.25 and 0.5, the best fit was reached at 0.25, to maintain the general form of the sign.

F. MSL RECOGNITION, CLASSIFICATION

To interpret sign language automatically from coordinate sequences, we observe N frames $\{x_n\}_{n=1}^{N}$ of a scene sequence. Specifically, we want to infer the M discrete variables $\{w_i\}_{i=1}^{M}$ that encode the sign that is present in each of the N frames. The data at time n reveal something about the sign; however, this may be insufficient to specify it accurately. Moreover, dependencies between adjacent states are modeled, signs are more likely to appear in a certain order, and we use this knowledge to reduce the ambiguity in the sequence. These dependencies have the form of a chain model because we model probabilistic connections only between adjacent states in a time series.

The model describes the joint probability of a set of continuous measurements $\{x_n\}_{n=1}^{N}$ and a set of discrete states $\{w_i\}_{i=1}^{M}$. The tendency to observe the x_n measurements given the w_i state takes the k value.

This value is encoded in the likelihood $\Pr(x_i|\,(w_i = k))$ The prior probability of the first state $\{w_1\}$ is explicitly encoded in the discrete distribution $\Pr(w_1)$ We assume that this is uniform. Each remaining state depended on the previous state. This information is captured in the distribution $\Pr(w_i|w_{i-1})$ known as Markov assumption.

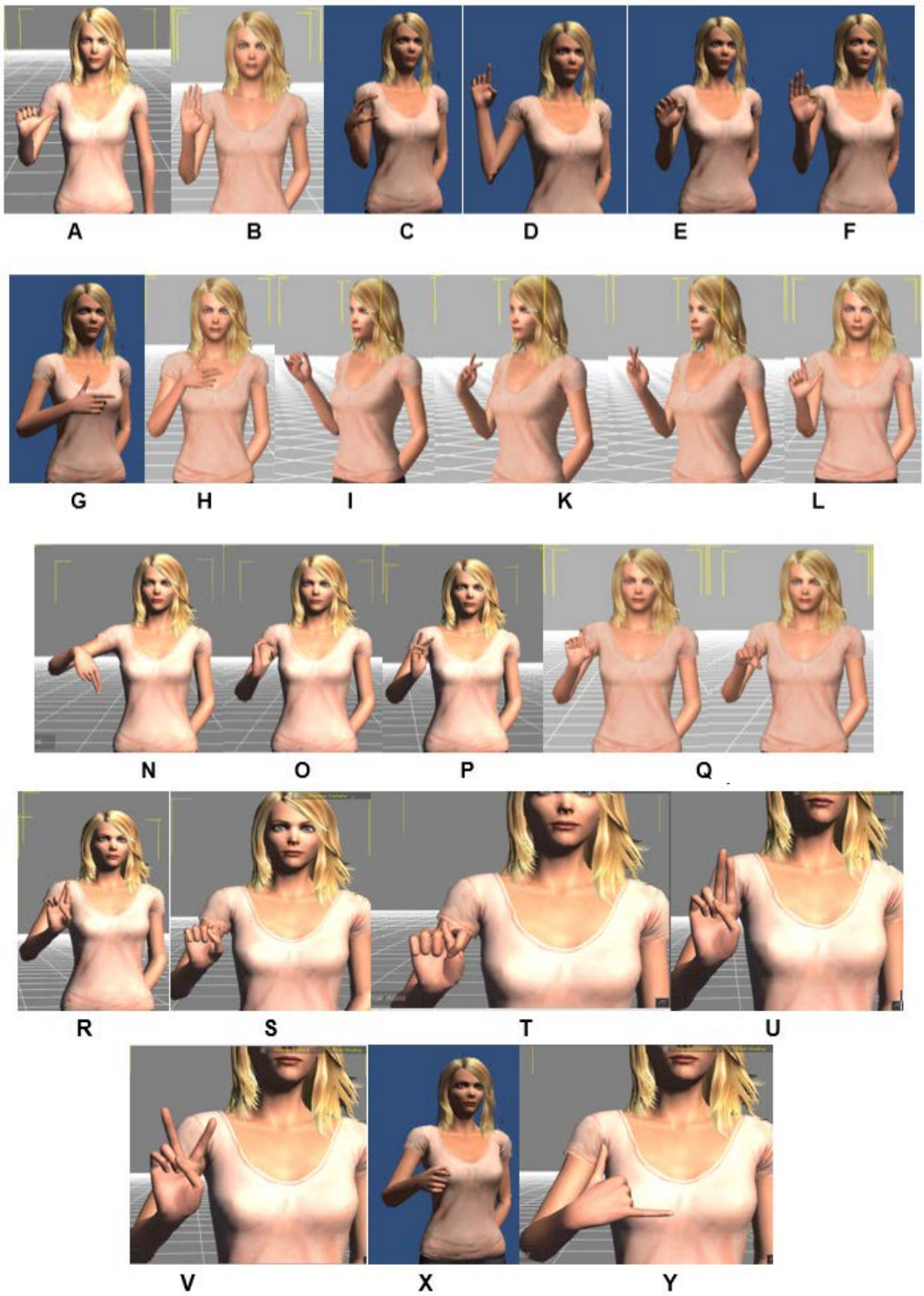

FIGURE 6. Some Mexican sign language (MSL) animated alphabet letters performed by the avatar. Letters that show more than one frame are dynamic.

The overall joint probability was factorized. This model is known as the Hidden Markov Model (HMM). The states $\{w_i\}_{i=1}^{M}$ in the directed model have the form of a chain and the overall model has the form of a tree [33].

To discriminate and identify a sign, classification (FIGURE 8) was performed on two data types: images and X, Y, and Z coordinates. Specifically, an HMM was applied to the images, and another HMM was performed on the X, Y, and Z coordinates.

One of the related problems of an HMM is to find a model μ that maximizes the probability of a sequence of observations $O = (o_1, o_2, \ldots, o_T)$; that is, to determine the model that

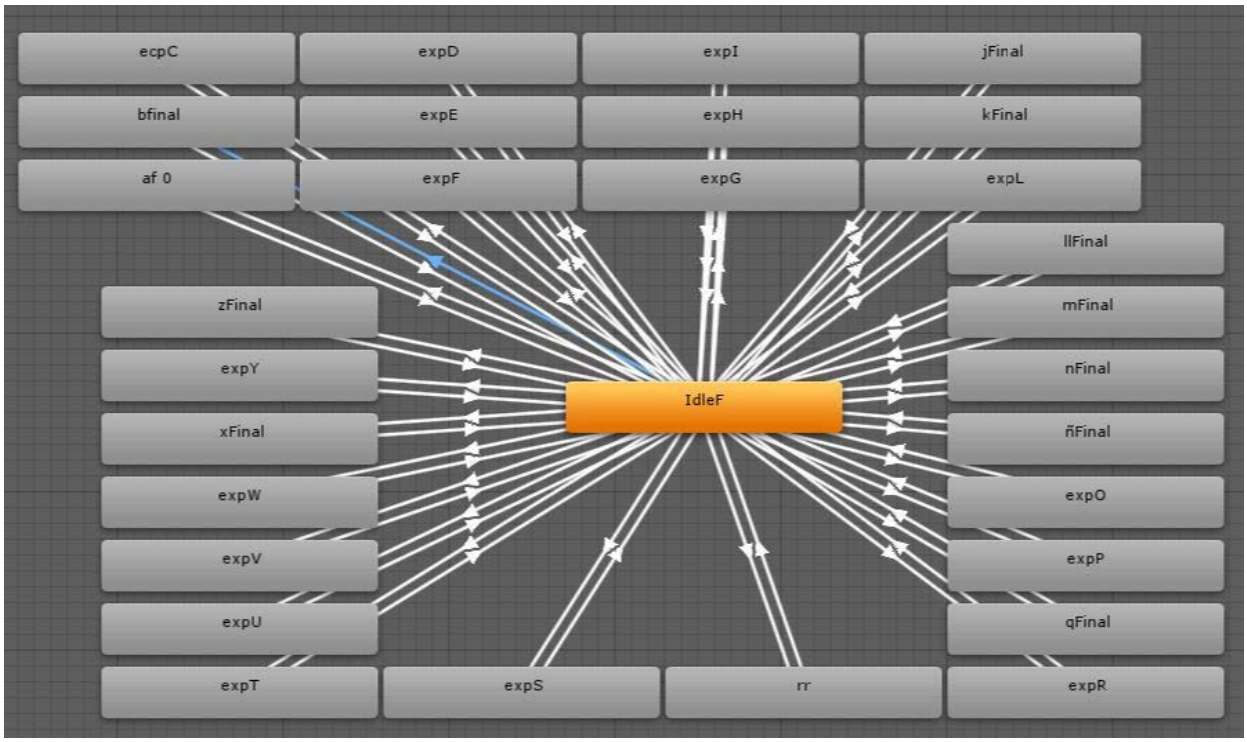

FIGURE 7. Animation state machine. The orange rectangle is the initial state. The arrows are transitions between states. When two states are connected, Unity computes a linear interpolation between the last pose of the previous state and the first pose of the next state. Each gray rectangle is an animation of a letter or an action.

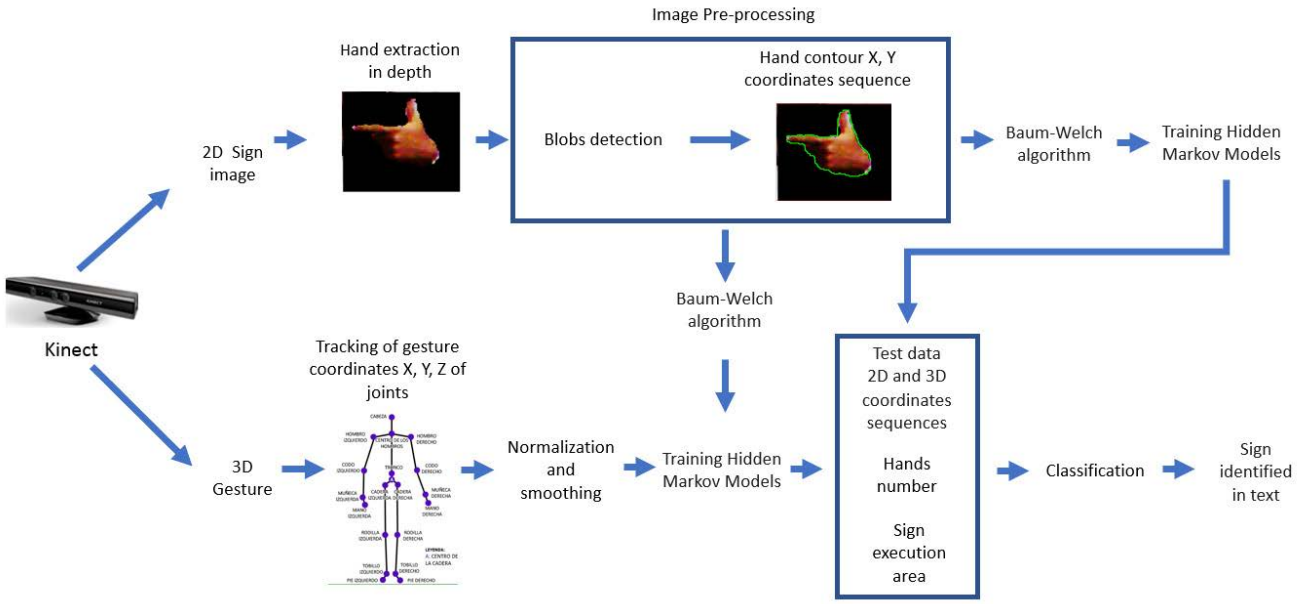

FIGURE 8. Data flow for the recognition of Mexican sign language using Kinect data.

best explains the sequence. However, it is not possible to find such a model analytically; therefore, the Baum-Welch algorithm was used to estimate the parameter μ of an HMM model that maximizes the probability of a sequence of observations $P(O|\mu)$. Also known as the forward-backward algorithm, the Baum-Welch algorithm is a dynamic programming approach and a special case of the expectation-maximization algorithm. Its purpose is to tune the parameters of the HMM, namely the

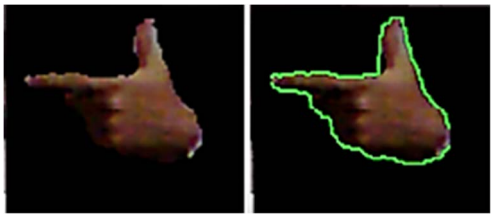

FIGURE 9. Hand contour.

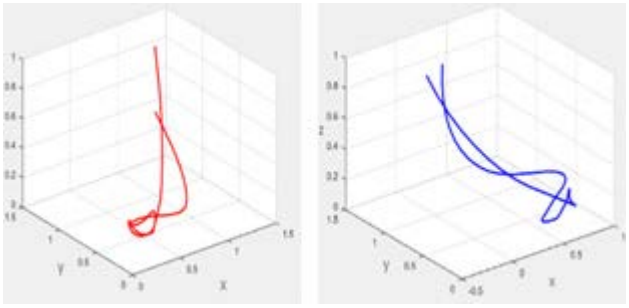

FIGURE 10. Examples of hand trajectories described by the kilogram sign. (a) right hand (red curve), (b) left hand (blue curve).

state transition matrix A, the emission matrix B, and the initial state distribution π_0, such that the model is maximally similar to the observed data [17].

The HMM uses five states per sign applied to coordinates X, Y, and Z in the classification phase.

The outputs from the coordinates and hand posture were compared against the scheme of the body area where the sign was made. Four main areas were identified in the process of capturing the signs: the head, left body side, right body side, and central body part (FIGURE 11). Within these areas, location is determined by where a person performs a sign. Consequently, this was provided as an input to recognize the signs. Moreover, there were signs that required movement from both hands or only the dominant hand; therefore, the number of hands involved in the sign was used as the input.

G. MSL RECOGNITION, TRAINING PROCESS

For training, the coordinates are stored in XML databases with the sign trajectory and sequences describing hand contours to create the model of each sign. There are three databases: alphabet letters from MSL finger spelled (29 signs) (FIGURE 12), numbers from 0 to 9 (10 signs) (FIGURE 13), and medical context words (43 signs) (Table 1).

Twelve volunteers (ten listeners and two deafs) participated in the training process. Each user performed each sign ten times. Model creation took four minutes. The task was conducted offline. In contrast, test classification was performed in real time.

H. MSL RECOGNITION, TEST PROCESS

Ten volunteers (eight listeners and two deafs) participated in this test. These were different from those of the participants in the training phase. During the test phase, participants performed each sign once. Consequently, a test set of 430 signs of medical context words, 290 signs from alphabet letters, and 100 signs from numbers were obtained.

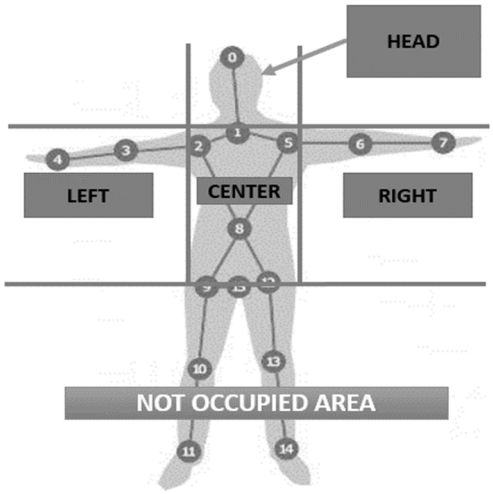

FIGURE 11. Body areas used for sign recognition.

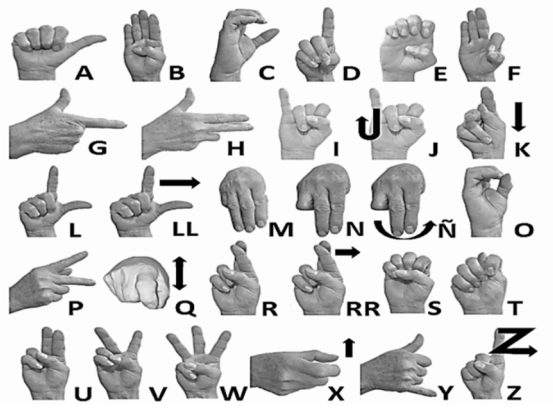

FIGURE 12. Alphabet letters from MSL.

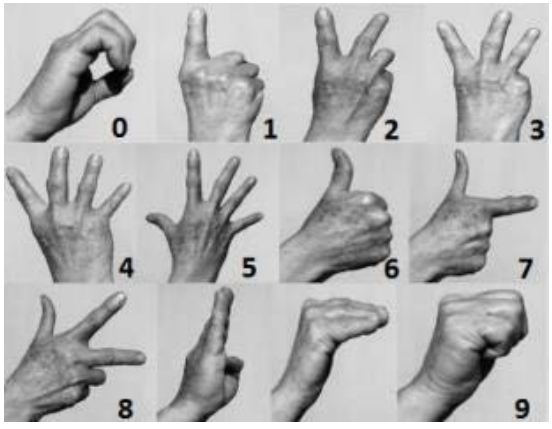

FIGURE 13. Decimal digits for MSL.

TABLE 1. Words, hands number and areas involved in each sign recognition.

Word	Hands	Area
Abortion	Both	Left
Allergy	Both	Head
Years	Both	Center
Burn	Both	Left
Fine	One	Head
Head	Both	Head
Tiredness	Both	Center
Itching	One	Head
Heart	One	Center
Body	Both	Center
Weak	Both	Center
Day	Both	Center
Diarrhea	Both	Center
Pain	Both	Center
Sleep	One	Head
Pregnancy	Both	Left
Stomach	One	Center
Excrement	One	Center
Fever	One	Head
Throat	One	Head
Gastritis	Both	Center
Flu	One	Head
Wound	One	Center
Hello	One	Head
Kilogram	Both	Center
Bad	One	Head
Dizziness	Both	Head
Menstruation	One	Head
Month	Both	Center
Much	Both	Center
Molar/Tooth	One	Head
Nervous	Both	Center
No	One	Right
Pee	One	Center
Few	One	Center
Scorch	Both	Center
So-so	One	Right
Week	One	Right
Yes	One	Right
Cough	One	Center
You	One	Center
See/Sight/Eyes	Both	Head
Vomit	Both	Center

The test phase was performed in real time by placing a person in front of the sensor to capture the corresponding data. Four sets of data were obtained to determine which sign was made: sign trajectory, image contour, body area where the sign was made, and the number of hands involved in the sign performance (See Table 1). The data for each sign were compared to the sign database to discard incorrect signs. Sign-trajectory identification was performed in seconds and displayed in the text at the interface. To recognize the sign, the probability of each sign given the observations was obtained and the model with the highest probability was selected.

I. MEDICAL CONTEXT GRAPHICS USER INTERFACE (GUI)

The system interface is composed of three modules:

1. The doctor-patient module (FIGURE. 14). This allows interaction between doctors and patients. Specifically, this module has five areas: a) the area where the doctor writes the question. It is important to mention that the doctor requires prior training in the words available in the system to ask his questions; b) an area that shows the video of the patient answering the doctor's question in real time; c) an area with an avatar where the translations from Spanish into MSL of the doctor's questions are displayed; d) an area displaying text with the same message as the avatar to ensure deaf patient comprehension. This helps in the case that the deaf person is able to understand written language; and e) an area displaying the patient's sign response in terms of written Spanish for the doctor. To use this interface, the doctor was placed in front of the computer and the patient was placed in front of the sensor (FIGURE 15). There are two screens to display the same interface to the patient and doctor, so that the doctor and patient can see the doctor's question and the patient's answer at the same time.
2. Alphabet module. This module helps to spell words (FIGURE 16) and is accessed from the doctor-patient interface. Specifically, this module displays images of the hand postures corresponding to each letter. This module is also useful for including new words finger-spelled by users in a dataset.
3. The numbers module (FIGURE 17) works in a similar way to the alphabet module and can also be accessed from the doctor–patient interface. This is activated when it is necessary to express numerical quantities. It only identifies numbers from zero to nine. To express two or more digits, one must spell digits by digit until the end of the complete number.

VI. EVALUATION

A. WORDS USED IN PRIMARY CARE CONSULTATION SERVICE

Words belonging to the semantic field of a primary care consultation service were used in the system. These words were chosen based on the advice of three qualified primary care physicians, an MSL interpreter, and a deaf person regarding the most commonly used medical questions posed by doctors in a general medicine consultation service. As a result, 43 words were selected for the experiments (Table 1).

B. PARTICIPANTS

The experiments were performed according to the Declaration of Helsinki and Nuremberg Code. All human subjects who used sign language for data acquisition (images and 3D scans) participated voluntarily. The subjects signed a consent form to participate in the data acquisition experiments.

The research protocol for this study was reviewed by a Ph.D. Research Program Committee of the Research Center in Artificial Intelligence of the University of Veracruz prior to the initiation of the research. In addition, the participants were informed of the research objectives. Nobody was harmed, and

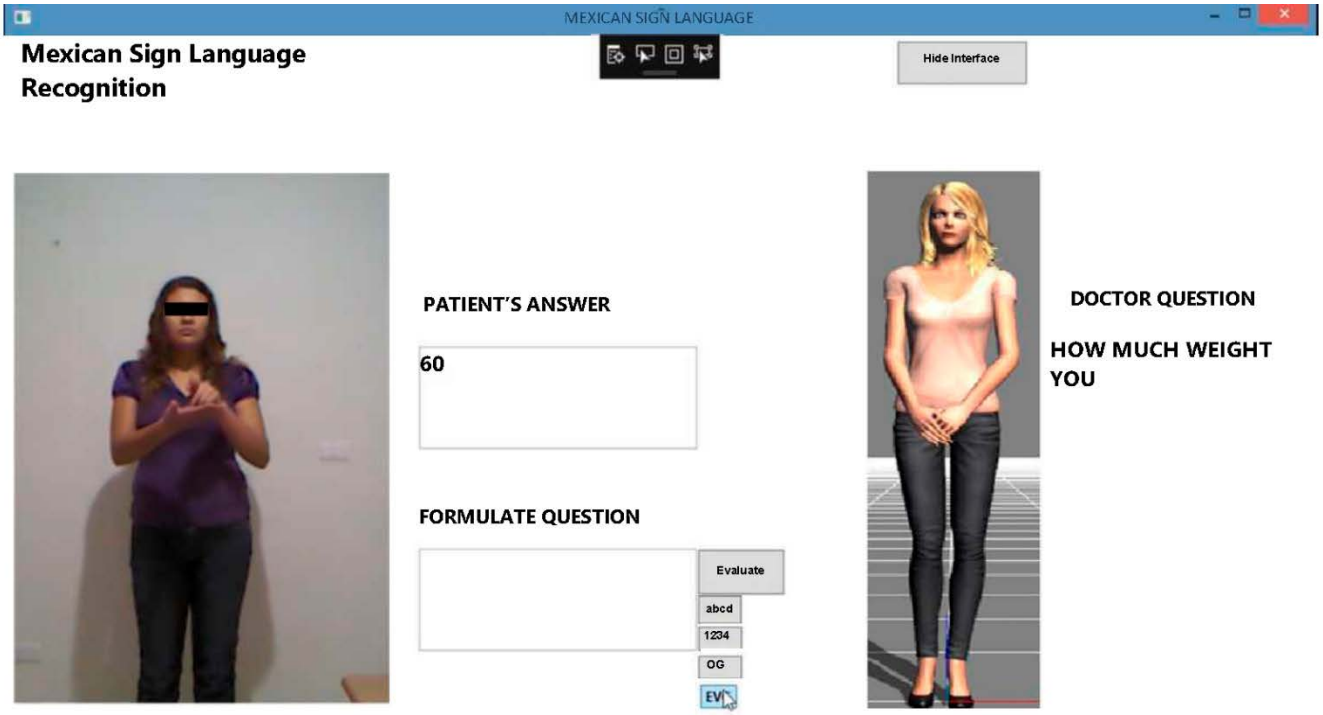

FIGURE 14. Doctor-Patient Interface. The Figure shows doctor's question to deaf patient (in "DOCTOR QUESTION" section): how much weight you (how much do you weigh?) in MSL and patient's response to the physician (in "PATIENT'S ANSWER" section) recognized in real time is 60 (60 kilograms). Within its section the physician has the controls to go to the interface to spell letters and numbers when its required. Yellow boxes indicate physician and deaf patient interface sections in the figure.

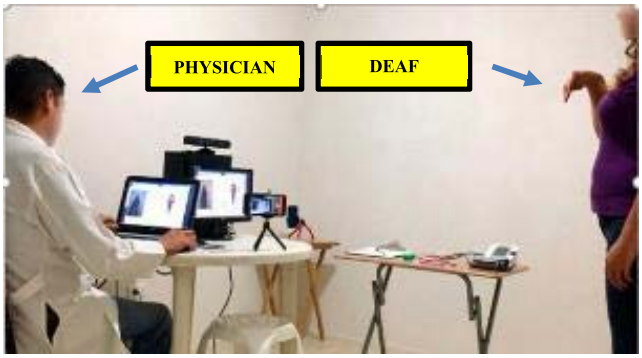

FIGURE 15. Real-time system test (tripods with cellphones shown in the figure were used only to record a demo video from different angles of the scene but they are not part of system implementation.)

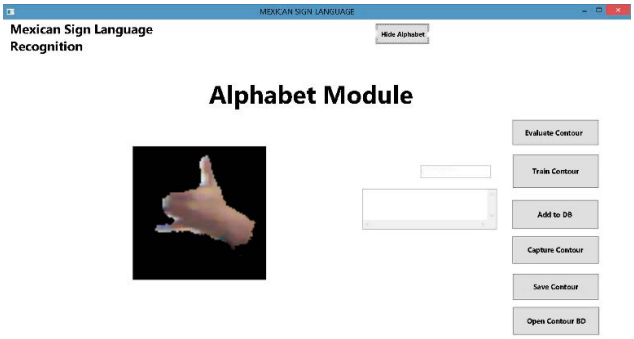

FIGURE 16. Interface of the alphabet fingerspell module (Alphabet module) The physician can use alphabet module when the patient needs to express words not recognized by the system by fingerspelling them out, the controls include the options to capture the signal and evaluate it and even add new fingerspelled words.

each participant could withdraw from the experiments at any moment if they wished.

Twenty-two volunteers (18 listeners and 4 deaf, mean age of 22 years, standard deviation of 7.15 years) participated in the experiments, all of whom used the same dialect of MSL. The volunteers were divided into 12 in the training phase and 10 in the test phase. The participants did not have any previous experience with interaction with a computer system developed for this task. They were given previous training for data collection, explaining how the experiment would be performed, and how to interact with the system. Each participant practiced three times before starting data capture.

C. EXPERIMENTS

Prior to performing the experiments, the users were informed about the objective of the experiment, steps to perform the experiment, interaction with the system, storage of the outcome data, and repetitions to be performed for each signal. Moreover, all participants were instructed on the words that

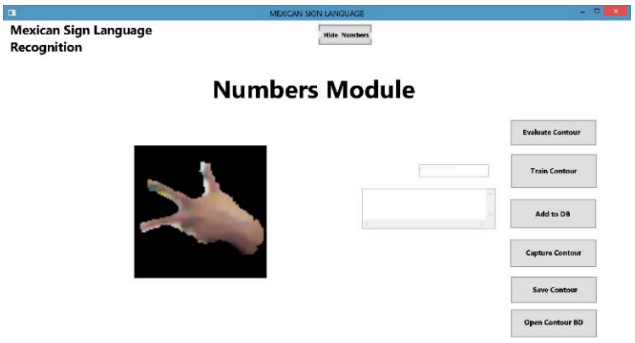

FIGURE 17. Number fingerspell module interface (Numbers Module). It shows the detected hand posture and shows the number to which it corresponds, as it only recognizes numbers from 0 to 9, the figures must be fingerspelled.

FIGURE 18. System users signing.

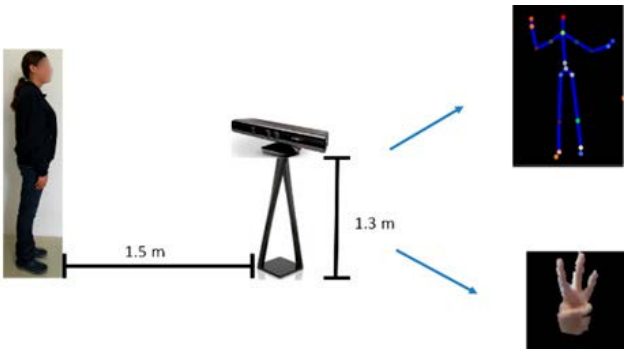

FIGURE 19. Experimental design for data acquisition.

to be performed in MSL, even if they had prior knowledge of MSL. They practiced 82 signs, five times each. (FIGURE 18)

The Kinect sensor was located at a height of 1.3 meters and at 1.5 meters from the person. This was due to the average height (0.093 height standard deviation) of the participants (FIGURE 19). A mark was placed on the floor to indicate to the participants the distance of 1.5 meters apart from the sensor.

Moreover, the following conditions were met in the experiments: i) participants stayed indoors away from windows, ii) there was a single light source coming from the ceiling and projecting downwards, and iii) the distances to which the sensor was placed were met.

To verify that the sign was correctly performed, it was rehearsed before starting the capture. Additionally, there was a person who reminded participants how to perform the sign and told them to perform it twice for its correct validation before capturing it.

Each participant performed each sign 10 times. Each repetition was stored in database.

D. METRIC

Accuracy, sensitivity, specificity, and F1 Score metrics were used to measure the validity of our proposal in terms of MSL sign classification.

Where:

True positive (TP) = number of cases in which the sign is detected when that sign is made.

False positive (FP) = number of cases where the sign is detected when another sign is made.

True negative (TN) = the number of cases in which the sign was not detected when another sign was present.

False negative (FN) = the number of cases in which the sign was not detected when that sign was made.

Accuracy refers to the closeness of a measured value to an actual or true value. The accuracy metrics (Eq. 1) assesses the ability to correctly classify a test sign as true.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \cdot 100 \tag{1}$$

Sensitivity or recall is the rate of true positives (Eq. 2). The greater the sensitivity of the test, the more the signs are properly identified.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \cdot 100 \tag{2}$$

The specificity of the test was the true negative rate (Eq. 3). The higher the specificity of the test, the lower the FP rate.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \cdot 100 \tag{3}$$

Precision refers to how precise/accurate the model is out of the predicted positive, how many of them are actually positive (Eq. 4).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \cdot 100 \tag{4}$$

F1 Score is a balance between Precision and Sensitivity (Eq. 5).

$$\text{F1} = 2 \cdot \frac{\text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \cdot 100 \tag{5}$$

VII. RESULTS

Table 2 shows the values obtained in the metrics (accuracy, sensitivity, specificity, precision, and F1 Score) for the experiments performed using primary care consultation service words (Table 1). In terms of specificity, a mean specificity rate of 99.80% was obtained for the general medicine consultation service words. Specifically, most of the words (36 of 43) obtained an accuracy rate of 100%, whereas ''Tiredness'' word had the lowest specificity rate (98%). Similarly, a mean accuracy of 99.5% was achieved. Many of the words (26 of 43) reported an accuracy rate of 100%, whereas ''Tiredness''

word obtained the lowest accuracy rate (98%). The precision metric also reported a mean rate of >90%. Over half of the words (24 of 43) reported a 100% precision rate; however, the lowest precision rate was 53% (''Tiredness'' word).

The sensitivity and F1 score metrics reported mean rates below 90 (87.90% for sensitivity and 88.60% for F1 score). Specifically, half of the words (21 of 43) had a 100% sensitivity rate. In contrast, ''scorch'' and ''so-so'' words reported the lowest sensitivity rate (60%). On the other hand, only 16 of 43 words reported 100% as F1 score rate, whereas ''Tiredness'' word reported the lowest F1 score rate (64%). It is important to note that the ''tiredness'' word obtained the lowest rates for four of the five metrics.

For all the dynamic signs that obtained a low score, this was because the hand contour was not always clearly visible, but only the wrist or back of the hand. In other cases, the sign was performed too quickly. For improvement and future work, we plan to use a sensor with a higher resolution and temporal sampling rate and also incorporate additional RGBD cameras to obtain complementary views of the signs that are partially occluded from a frontal view. From a machine learning point of view, the acquisition of more examples from more users will help our models to generalize better.

Table 3 presents the results for the alphabet letter metrics. The specificity achieved the highest mean (99.70%), followed by the accuracy metric (99.50%). Most letters achieved 100% specificity (22 of 29 letters) and accuracy rates (18 of 29 letters). Conversely, 98% and 99% were reported as the lowest rates for accuracy (letter: s) and specificity (letters: d, e, f, g, m, n, s), respectively.

The precision, F1-score, and sensitivity metrics achieved mean rates of 92.80%, 92.50% and 92.40%, respectively. More than half of the letters achieved a 100% precision rate (17 of 22 letters), F1 score rate (13 of 22 letters), and sensitivity rate (15 of 22 letters). These three metrics reported 70% as the lowest rate in letters.

This is an exploratory study because there are no previous MSL studies on patient-doctor interaction; similar studies focus only on alphabet fingerspelling, and this is not practical for communication in daily life.

Regarding doctor-patient communication in an interview with three primary care physicians, they expressed that the system improves direct communication with the deaf patient, since in many cases deaf people have no way to communicate with the doctor and they have to depend on a relative who tells the doctor the symptoms and who decides the treatment which is not ideal but the patient must be informed of all implications and give consent for the treatment of his illness. The doctors commented that deaf people do not go to medical visits daily for communication difficulties that if those barriers were broken patient consultations would surely increase between deaf patients, and a doctor mentioned that a patient read the lips and thus communicated, but that is not a skill widespread that deaf people possess, which is why the community deaf is limited in accessing a medical consultation by the existing communication difficulties.

TABLE 2. Metric results for general medicine consultation service words.

Word	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1 Score (%)
Abortion	100	80	100	100	89
Allergy	100	80	100	100	89
Bad	100	90	100	90	90
Body	99	70	100	78	74
Burn	99	90	100	82	86
Cough	99	70	100	100	82
Day	100	100	100	100	100
Diarrhea	100	100	100	100	100
Dizziness	100	100	100	100	100
Excrement	100	100	100	91	95
Fever	100	100	100	100	100
Few	100	100	100	100	100
Fine	99	80	100	80	80
Flu	99	80	100	89	84
Gastritis	100	90	100	100	95
Head	99	100	99	63	77
Heart	99	70	99	70	70
Hello	100	100	100	100	100
Itching	100	90	100	90	90
Kilogram	100	100	100	100	100
Menstruation	99	70	99	70	70
Molar/Tooth	99	80	99	73	76
Month	100	100	100	100	100
Much	100	80	100	100	89
Nervous	100	100	100	100	100
No	100	100	100	91	95
Pain	100	100	100	100	100
Pee	100	100	100	100	100
Pregnancy	100	100	100	83	91
Scorch	99	60	100	86	71
See/Sight/Eyes	99	70	100	100	82
Sleep	99	70	100	100	82
So-so	99	60	100	100	75
Stomach	100	100	100	100	100
Throat	99	80	99	73	76
Tiredness	98	80	98	53	64
Vomit	100	100	100	100	100
Weak	99	70	100	78	74
Week	100	100	100	100	100
Wound	99	70	99	70	70
Years	100	100	100	100	100
Yes	100	100	100	91	95
You	100	100	100	100	100
MAXIMUM	100	100	100	100	100
MEAN	99.5	87.9	99.8	90.7	88.6
MINIMUM	97.9	60.0	98.3	53.3	64.0
STANDARD DEVIATION	0.6	13.6	0.4	12.7	11.5

Table 4 presents the metrics of these numbers. It can be noticed that specificity achieved the highest mean rate (99.70%) followed by accuracy (99.40%). The remaining metrics achieve similar mean rates (97%).

Only numbers one and six reported mean rates between 80% and 99% for the five metrics. The remaining numbers were 100% for all metrics.

VIII. DISCUSSION

Our recognition method achieved a precision of 99% and a F1-Score average of 88%. Obtaining a higher precision (99%) than similar works using Kinect, such as [20] and [30].

Our proposal has the following advantages:

- The Kinect sensor involved in the interaction is non-intrusive; consequently, people simply place it in front

TABLE 3. Metric results for alphabet letters.

Letter	Accuracy	Sensitivity	Specificity	Precision	F1 Score
A	100	90	100	100	95
B	99	80	100	89	84
C	100	100	100	100	100
D	99	90	99	75	82
E	99	90	99	75	82
F	99	90	99	82	86
G	99	90	99	82	86
H	100	100	100	100	100
I	100	90	100	100	95
J	100	100	100	100	100
K	100	100	100	100	100
L	99	80	100	89	84
LL	100	100	100	100	100
M	99	80	99	80	80
N	99	80	99	80	80
Ñ	100	100	100	100	100
O	100	100	100	100	100
P	100	100	100	100	100
Q	100	100	100	91	95
R	99	80	100	89	84
RR	100	100	100	100	100
S	98	70	99	70	70
T	99	80	100	100	89
U	100	100	100	100	100
V	100	100	100	100	100
W	100	100	100	100	100
X	100	90	100	100	95
Y	100	100	100	91	95
Z	100	100	100	100	100
MAXIMUM	100	100	100	100	100
MEAN	99.5	92.4	99.7	92.8	92.5
MINIMUM	97.9	70.0	98.9	70	70
STANDARD DEVIATION	0.6	9.1	0.4	9.7	8.7

TABLE 4. Metric results for numbers.

Number	Accuracy	Sensitivity	Specificity	Precision	F1 Score
0	100	100	100	100	100
1	97	90	98	82	86
2	100	100	100	100	100
3	100	100	100	100	100
4	100	100	100	100	100
5	100	100	100	100	100
6	97	80	99	89	84
7	100	100	100	100	100
8	100	100	100	100	100
9	100	100	100	100	100
MAXIMUM	100	100	100	100	100
MEAN	99.4	97.0	99.7	97.1	97.0
MINIMUM	97.0	80.0	97.8	81.8	84.2
STANDARD DEVIATION	1.3	6.7	0.7	6.4	6.4

of the sensor and perform the signs without wearing any clothing or device. The Kinect is an easily affordable device. After the experiments, it was observed that it was only necessary to save the coordinates corresponding to the hands.

Body joints (i.e., elbows) did not provide key data for recognition and could provide noise to the data because each person moved their arms and placed them at different positions when they performed a sign. In our case, it was also possible to use an RGBD camera instead of a Kinect sensor.

TABLE 5. Comparative table of works related to signs in medical field.

Reference	Sign Language	Accuracy	Amount of medical signs	Data
[10]	American Sign Language	93%	5 signs from medical field and healthcare	Video Clips
[27]	Filipino sign language	80.55%	30 signs used in assessing health	Data glove (Flex sensor)
[35]	Mexican Sign Language	94.9%.	10 signs from an emergency medical vocabulary	Camera
Proposed work	Mexican Sign Language	99%	43 signs from medical vocabulary	Kinect Sensor

- Our prototype system showed that communication between the patient and doctor, due to the translation from Spanish to MSL, and vice versa, is feasible. In the future, we plan to conduct usability tests to confirm successful communication.
- Our study combines the features of both sensors, the depth sensor when retrieving the X, Y, and Z coordinates, and the RGB camera for capturing images of the hands when signing. Some signs describe the same trajectory, and what differentiates them is hand posture.

It is important to note that the performance of our system can be affected by the use of the Kinect sensor. Specifically, the performance of the Kinect sensor may be affected by the following factors:

- Sunlight: Consequently, to obtain the best results, our system should only be used indoors.
- It is used for several hours (approximately 6 h of uninterrupted use). Consequently, the body could not be accurately identified. This led we stopped the experiments and switched off the sensor to cool it.

Regarding the comparison with other works that have been used to test medical signs (see Table 5), [10] reported a case study of the recognition of five medical signs in American sign language with 93% accuracy. Reference [35] presented a system that recognized 73 MSL words with a precision of 94.9%. Of these words, 10 were from emergency medical vocabulary. [27] developed a method to combine a hand glove with a computer vision system to translate Filipino sign language for medical purposes. Their vocabulary consisted of 26 alphabet letters, 10 decimal digits, and 30 words used in healthcare, with 80.55% accuracy. In the case of these three studies, our method recognizes 82 different signs with an average accuracy of 99%.

IX. CONCLUSION

Sign language recognition through computational vision is not a trivial task as it is for people who, despite the variations between subjects or noise added to the signal, can easily distinguish the signs. Incorporating a bimodal interface and

combining the two inputs (depth and RGB data) yields the best results for both sensors to obtain fast results that are susceptible to being used in real time, which is necessary for communication.

The main contribution of this work lies in the support of deaf people's communication in the medical consultation of a primary care physician. This can be implemented in health services and opens the door to providing in Mexico a functional prototype of MSL interpretation focused on the health context. Additionally, this implementation is not limited to the digital alphabet.

The chosen sensor and developed software fulfilled the task of identifying the proposed signs with 99% precision, despite its limitations. It should be mentioned that with more training data, the system will become more robust and can be used for other sign languages only by storing databases with the signs of the corresponding language.

The resulting average F1 Score of 88% could be improved by strengthening with more cases of sign training. In future work, we intend to add more words used in the medical context to improve communication for deaf people and hearing people, as well as to increase the size of the database by inviting more test subjects.

The low metrics in the recognition of signs such as fatigue, week, burns, menstruation, and heart are related to the adaptation of the space for capturing the signs, which must be far from windows and sunlight that allows the sensor to avoid a bad communication that could cause the doctor to write a bad prescription or not see some symptoms due to the mistranslation.

In addition, the failure is related to the heating of the sensor owing to continuous hours of use, which can be solved by turning off the sensor when it is not used to guarantee its proper functioning when interviewing a patient.

In future work, we propose a percentage evaluation of cases where the doctor could communicate correctly with the patient and thus correctly diagnose the disease and avoid poor selection of drugs for treatment.

In addition, other sensors such as the Leap Motion Controller (small optical USB hand-tracking module designed to be placed on a physical desk, facing up) and Empatica E4 (a portable wireless multisensor device for data and computerized biofeedback in real-time acquisition. It has four built-in sensors: a photoplethysmograph (PPG), electrodermal activity (EDA), 3-axis accelerometer, and temperature.) to enhance the effectiveness of the proposed solution.

DECLARATION

Conflicts of Interest/Competing Interest: Not applicable.

Availability of Data and Material: The data can be made available on request.

Authors' Contributions: Candy Obdulia Sosa-Jiménez: conception, MSL recognition, design, data acquisition, data analysis, software programming, and paper writing; Homero Vladimir Ríos-Figueroa: conception, design, data analysis, supervision, and paper writing; Ana Luisa Solís-González-Cosío: conception, MSL synthesis, design, programming, data analysis, supervision, and paper writing.

REFERENCES

[1] M. T. Calvo-Hernandez, M. de L. Acosta-Huerta, E. D. Maya-Ortega, E. Sanabria-Ramos, and G. A. Zeleni, "Spanish dictionary—Mexican sign language. (DIELSEME)," Special Educ. Board, Ministry Educ., Mexico, (in Spanish), 2004.

[2] S.-O. Caballero-Morales and F. Trujillo-Romero, "3D modeling of the Mexican sign language for a speech-to-sign language system," *Comput. Sistemas*, vol. 17, no. 4, pp. 593–608, 2013.

[3] M. T. Calvo-Hernandez, *Mexican Sign Language Dictionary*. Mexico City, Mexico: Ministry of Education, 2010.

[4] J. Cervantes, F. García-Lamont, L. Rodríguez-Mazahua, A. Y. Rendon, and A. L. Chau, "Recognition of Mexican sign language from frames in video sequences," in *Intelligent Computing Theories and Application. ICIC 2016* (Lecture Notes in Computer Science), vol. 9772, D. S. Huang and K. H. Jo, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-42294-7_31.

[5] X. Chai, G. Li, X. Chen, M. Zhou, G. Wu, and H. Li, "VisualComm: A tool to support communication between deaf and hearing persons with the kinect," in *Proc. 15th Int. ACM SIGACCESS Conf. Comput. Accessibility (ASSETS)*, Oct. 2013, pp. 1–2.

[6] M. Cruz-Aldrete, "Mexican sign language grammar," Ph.D. thesis, College Mexico, Mexico City, Mexico, 2008.

[7] I.-J. Ding and C.-W. Chang, "An adaptive hidden Markov model-based gesture recognition approach using kinect to simplify large-scale video data processing for humanoid robot imitation," *Multimedia Tools Appl.*, vol. 75, no. 23, pp. 15537–15551, Dec. 2016.

[8] I.-J. Ding and C.-W. Chang, "An eigenspace-based method with a user adaptation scheme for human gesture recognition by using kinect 3D data," *Appl. Math. Model.*, vol. 39, no. 19, pp. 5769–5777, 2015.

[9] P. Doliotis, A. Stefan, C. Mcmurrough, D. Eckhard, and V. Athitsos, "Comparing gesture recognition accuracy using color and depth information," in *Proc. 4th Int. Conf. Pervasive Technol. Rel. Assistive Environ. (PETRA)*, Crete, Greece, 2011, pp. 1–7.

[10] R. Elakkiya and K. Selvamani, "Extricating manual and non-manual features for subunit level medical sign modelling in automatic sign language classification and recognition," *J. Med. Syst.*, vol. 41, no. 11, Nov. 2017, Art. no. 175, doi: 10.1007/s10916-017-0819-z.

[11] J. Espejel-Cabrera, J. Cervantes, F. García-Lamont, J. S. R. Castilla, and L. D. Jalili, "Mexican sign language segmentation using color based neuronal networks to detect the individual skin color," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115295.

[12] R. Galicia, O. Carranza, E. D. Jiménez, and G. E. Rivera, "Mexican sign language recognition using movement sensor," in *Proc. IEEE 24th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2015, pp. 573–578.

[13] E. Gani and A. Kika, "Albanian dynamic dactyls recognition using kinect technology and DTW," in *Proc. 8th Balkan Conf. Informat.*, Sep. 2017, pp. 1–7.

[14] G. García-Bautista, F. Trujillo-Romero, and S. O. Caballero-Morales, "Mexican sign language recognition using kinect and data time warping algorithm," in *Proc. Int. Conf. Electron., Commun. Comput. (CONIELECOMP)*, 2017, pp. 1–5.

[15] L. Jaemin, H. Takimoto, H. Yamauchi, A. Kanazawa, and Y. Mitsukura, "A robust gesture recognition based on depth data," in *Proc. 19th Korea-Japan Joint Workshop Frontiers Comput. Vis.*, Jan. 2013, pp. 32–127.

[16] J. Jimenez, A. Martin, V. Uc, and A. Espinosa, "Mexican sign language alphanumerical gestures recognition using 3D Haar-like features," *IEEE Latin Amer. Trans.*, vol. 15, no. 10, pp. 2000–2005, Oct. 2017.

[17] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, May 2008.

[18] P. Kumar, R. Saini, P. P. Roy, and D. P. Dogra, "A position and rotation invariant framework for sign language recognition (SLR) using kinect," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8823–8846, Apr. 2018.

[19] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "A multimodal framework for sensor based sign language recognition," *Neurocomputing*, vol. 259, pp. 21–38, Oct. 2017.

[20] S. Lang, M. Block-Berlitz, and R. Rojas, "Sign language recognition using kinect," in *Proc. Int. Conf. Artif. Intell. Soft Comput.*, in Lecture Notes in Computer Science, vol. 7267, 2012, pp. 394–402, doi: 10.1007/978-3-642-29347-4_46.

[21] M. P. Lewis, G. F. Simons, and C. D. Fennig, *Languages of Africa and Europe*. Dallas, TX, USA: SIL International Publications, 2016.
[22] L. A. Lopez-Garcia, R. S. Rodriguez-Cervantes, M. G. Zamora-Martinez, and S. S. Esteban-Sosa, *My Hands That Talk, Sign Language for Deaf*. Mexico City, Mexico: Trillas, 2015.
[23] F. E. Luis-Pérez, F. Trujillo-Romero, and W. Martínez-Velazco, "Control of a service robot using the Mexican sign language," in *Advances in Soft Computing* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2011, pp. 419–430.
[24] Z. Ma and E. Wu, "Real-time and robust hand tracking with a single depth camera," *Vis. Comput.*, vol. 30, no. 10, pp. 1133–1144, Oct. 2014.
[25] M. Mohandes, M. Deriche, U. Johar, and S. Ilyas, "A signer-independent Arabic sign language recognition system using face detection, geometric features, and a hidden Markov model," *Comput. Electr. Eng.*, vol. 38, no. 2, pp. 422–433, 2012.
[26] *Disability in Mexico, Data as of 2014*, Nat. Inst. Statist. Geogr., Aguascalientes, Mexico, 2016. [Online]. Available: http://internet.contenidos.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825090203.pdf
[27] C. Ong, I. Lim, J. Lu, C. Ng, and T. Ong, "Sign-language recognition through gesture & movement analysis (SIGMA)," in *Mechatronics and Machine Vision in Practice 3*, J. Billingsley and P. Brett, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-76947-9_17.
[28] M. Palmeri, F. Vella, I. Infantino, and S. Gaglio, "Sign languages recognition based on neural network architecture," in *Intelligent Interactive Multimedia Systems and Services 2017* (Smart Innovation, Systems and Technologies), vol. 76, G. De Pietro, L. Gallo, R. Howlett, and L. Jain, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-59480-4_12.
[29] H. Pazhoumand-Dar, "Fuzzy association rule mining for recognising daily activities using kinect sensors and a single power meter," *J. Ambient Intell. Hum. Comput.*, vol. 9, no. 5, pp. 1497–1515, Oct. 2018.
[30] F. Pedersoli, S. Benini, N. Adami, and R. Leonardi, "XKin: An open source framework for hand pose and gesture recognition using kinect," *Vis. Comput.*, vol. 30, no. 10, pp. 1107–1122, Oct. 2014.
[31] N. Praveen, N. Karanth, and M. S. Megha, "Sign language interpreter using a smart glove," in *Proc. Int. Conf. Adv. Electron. Comput. Commun.*, Oct. 2014, pp. 1–5, doi: 10.1109/ICAECC.2014.7002401.
[32] F. P. Priego-Pérez, "Recognition of images of Mexican sign language," M.S. thesis, Dept. Comput. Sci., CIC, IPN, Mexico City, Mexico, 2012.
[33] S. Prince, *Computer Vision: Models, Learning, and Inference*. London, U.K.: Cambridge Univ. Press, 2012.
[34] R. Rastgoo, K. Kiani, and L. Escalera, "Sign language recognition: A deep survey," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113794.
[35] J. E. R. Sánchez, A. A. Rodríguez, and M. G. Mendoza, "Real-time Mexican sign language interpretation using CNN and HMM," in *Advances in Computational Intelligence. MICAI 2021* (Lecture Notes in Computer Science), vol. 13067, I. Batyrshin, A. Gelbukh, and G. Sidorov, Eds. Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-89817-5_4.
[36] M. E. S. de Fleischmann and R. Gonzalez-Perez, "Hands with voice," in *Dictionary of Mexican Sign Language*. Mexico City, Mexico: National Council for Discrimination Prevention, 2011.
[37] A. Shinde and R. Kagalkar, "Advanced Marathi sign language recognition using computer vision," *Int. J. Comput. Appl.*, vol. 118, no. 13, pp. 1–7, May 2015.
[38] S. V. J. Francisco, C. Toxqui-Quitl, D. Martínez-Martínez, and H.-G. Margarita, "Mexican sign language recognition using normalized moments and artificial neural networks," *Proc. SPIE*, vol. 9216, pp. 316–320, Sep. 2014, doi: 10.1117/12.2061077.
[39] F. Solís, C. Toxqui, and D. Martínez, "Mexican sign language recognition using Jacobi–Fourier moments," *Engineering*, vol. 7, no. 10, pp. 700–705, 2015.
[40] F. Solís, D. Martínez, and O. Espinoza, "Automatic Mexican sign language recognition using normalized moments and artificial neural networks," *Engineering*, vol. 8, no. 10, pp. 733–740, 2016.
[41] C. O. Sosa-Jiménez, H. V. Ríos-Figueroa, E. J. Rechy-Ramírez, A. Marin-Hernandez, and A. L. S. González-Cosío, "Real-time Mexican sign language recognition," in *Proc. IEEE Int. Autumn Meeting Power, Electron. Comput.*, Nov. 2017, pp. 1–6.
[42] M. M. Süzgün, H. Özdemir, N. Camgöz, A. Kindiroglu, D. Basaran, C. Togay, and L. Akarun, "Hospisign: An interactive sign language platform for hearing impaired," *Proc., Int. Conf. Comput. Graph., Animation Gaming Technol.*, vol. 11, no. 3, İstanbul, Turkey, 2015, pp. 75–92.
[43] F. Trujillo-Romero and S.-O. Caballero-Morales, "3D data sensing for hand pose recognition," in *Proc. 23rd Int. Conf. Electron., Commun. Comput.*, Mar. 2013, pp. 109–113.
[44] A. Truong, H. Boujut, and T. Zaharia, "Laban descriptors for gesture recognition and emotional analysis," *Vis. Comput.*, vol. 32, no. 1, pp. 83–98, Jan. 2016.
[45] Y. Wang, C. Yang, X. Wu, S. Xu, and H. Li, "kinect based dynamic hand gesture recognition algorithm research," in *Proc. 4th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, Aug. 2012, pp. 274–279.
[46] Z. Wang, B. Chen, and J. Wu, "Effective inertial hand gesture recognition using particle filtering based trajectory matching," *J. Electr. Comput. Eng.*, vol. 2018, pp. 1–9, Jan. 2018, doi: 10.1155/2018/6296013.
[47] World Health Organization. (2018). *Deafness and Hearing Loss*. [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
[48] H. Wu, J. Wang, and X. Zhang, "Combining hidden Markov model and fuzzy neural network for continuous recognition of complex dynamic gestures," *Vis. Comput.*, vol. 33, no. 10, pp. 1265–1278, Oct. 2017.
[49] Z. Zhang, Z. Tian, and M. Zhou, "HandSense: Smart multimodal hand gesture recognition based on deep neural networks," *J. Ambient Intell. Hum. Comput.*, 2018, doi: 10.1007/s12652-018-0989-7.

CANDY OBDULIA SOSA-JIMÉNEZ received the B.Sc. degree in computer science and the M.Sc. and Ph.D. degrees in artificial intelligence from the University of Veracruz, Mexico. She is a Professor with the Statistics and Informatics School, University of Veracruz. Her research interests include machine learning, pattern recognition, computer vision, sign language, sensors, human–computer interaction, and programming.

HOMERO VLADIMIR RíOS-FIGUEROA received the B.Sc. degree in mathematics and the M.Sc. degree in computer science from the National Autonomous University of Mexico (UNAM), Mexico, and the Ph.D. degree in computer science and artificial intelligence from the University of Sussex, U.K. He has been a Professor with the Research Institute for Artificial Intelligence, University of Veracruz, Mexico, since 2000. His research interests include artificial intelligence, computer vision, pattern recognition, and machine learning.

ANA LUISA SOLíS-GONZÁLEZ-COSíO received the B.Sc. degree in mathematics from the National Autonomous University of Mexico (UNAM), and the degree in software technology and computer graphics from the International Computing Center, Tokyo, Japan. She is a Professor with the Mathematics Department, School of Sciences, UNAM. Her research interests include visual computing, intelligent virtual environments, and human–computer interaction.