Hand Gesture Recognition for Deaf People Interfacing

Isaac García Incertis
CARTIF Foundation, Spain
isagar@cartif.es

Jaime Gómez García-Bermejo
University of Valladolid, Spain
jaigom@eis.uva.es

Eduardo Zalama Casanova
University of Valladolid, Spain
ezalama@eis.uva.es

Abstract

In this paper, an approach for deaf-people interfacing using computer vision is presented. The recognition of alphabetic static signs of the Spanish Sign Language is addressed. The proposed approach combines a number of norms to evaluate the distance of the current sign, to the sign models stored in a database (a dictionary). This solution leads to a largely selective criterion. The method is simple enough to provide real-time recognition, and works suitably for most letters.

1. Introduction

Hand-gesture recognition has become an important issue in pattern recognition [1, 2], since it has a large number of interesting applications within man-machine interfacing (e.g. domotics, ambient assisted living, easy pointing and selecting in graphic interfaces, or object handling in augmented reality systems [7]). In the present work, the use of hand-gesture recognition for deaf people interfacing is addressed [3, 4, 5]. Deaf-people interfacing is a very challenging issue, due to its cardinal social interest, and its inherent complexity: (i) the multichannel nature of the information (hand pose and spatial location, hand movement, movement speed, face expression, body pose); (ii) the signing variability from one to other person; (iii) and the presence of disturbing elements (clothes, surrounding furniture). Furthermore, sign language from different regions differs significantly. The present work focuses the static gesture case corresponding to the alphabet letters in the Spanish Sign Language (LSE).

Hand-gesture recognition has been approached in [6] and [7], but only for counting the number of stretched fingers (without taking care of finger identification). Gesture recognition has been also addressed in [5], but instrumented gloves are used which simplifies the recognition problem but complicates the hardware. The general approach proposed in [8] could be adapted for hand-gesture recognition, but requires solving a time-consuming, optimisation problem. Gesture identification upon image moments is proposed in [3], but this solution is only expected to work suitably for small gesture sets. Geometric histograms are proposed in [9, 10]. Curvature approaches are tackled in [11, 12]. Moment invariants, simple descriptors and chain code histograms methods are described in [13, 14, 15] respectively. Also particle approaches have been achieved [16]. Finally, a 3D recognition approach requiring a multicamera, thus a costly set-up, is proposed in [17]. A simpler 3D approach is proposed in [4], which in turn requires a 3D model with more than twenty degrees-of-freedom, as well as a continuous, error-free tracking of the hand (for updating 2D to 3D correspondence).

In the present work, a simple, low-time consuming approach is presented which uses only one camera (a common colour webcam), and is able to discriminate most signs in the spanish sign language.

The paper is organized as follows: The proposed approach is fully described in Section 2. Experimental results are shown in section 3. Concluding remarks are summarized in section 4

2. Method description

The overall approach works as follows: First, hand region and the corresponding contours are extracted through appropriate colour segmentation. Then, the obtained contour is sampled every a given arc distance, and resulting points are compared to those of a target gesture in a dictionary. This comparison is performed upon four distance criteria, so that a suitable recognition is finally achieved.
In concrete, RGB image is first converted into the HSV colour space (figure 1), since this space is more appropriate for arbitrary colour detection, and segmentation in HS plane is more robust to changes in the lighting conditions. Furthermore Hue origin is selected so that the interest region is about 180° (127 in byte representation). This way, the target colour region is roughly centered in the H circle. Then, points out of the target HS region are filtered out, in combination with some morphological operations which provide further noise reduction (figure 2). Then, the large region contours are extracted and fed into the gesture recognition process (figure 3).

Recognition is performed in the following way. Hand contour is sampled every a given arc distance. This distance is calculated upon the desired final points number, which is selected so that it is large enough to represent complex shaped geometry (and small enough to speed-up the recognition process), say one hundred points. The obtained point list is then converted to polar coordinates centred at the region gravity centre (figure 4). Furthermore, the radius are normalized so that some size invariance, thus signing person-to-camera distance invariance, is achieved.

Matching is performed by comparing points in the current gesture diagram to points in the model diagram. This comparison requires some point-to-confidence band distance calculation, which could be performed upon a given norm criterion. However, a much more selective criterion have been obtained by combining four norms (figure 6):

a) $L_0$ norm of this distance, i.e. number of points of the current gesture out of the confidence band around the gesture model.

b) $L_1$ norm of the same distance.
c) $L_2$ norm.

d) $L_{\infty}$ norm.

Distance of the current gesture to every gesture model in the dictionary is calculated upon the first norm above. A positive identification (under this norm) is assumed for the closest model, whenever the corresponding norm value is neatly smaller than the norm value corresponding to the next best-matching model. Otherwise, an uncertain result is assumed.

Distance of the current gesture to every gesture model in the dictionary is calculated upon the first norm above. A positive identification (under this norm) is assumed for the closest model, whenever the corresponding norm value is neatly smaller than the norm value corresponding to the next best-matching model. Otherwise, an uncertain result is assumed.

Figure 6. Sign Recognition by four criteria in close frame and full-trunk frame.

The same process is repeated for every norm above. A final positive identification is assumed whenever three or four norms lead to a positive identification for the same gesture in the dictionary.

3. Experimental results

The described procedure has been implemented and tested in a 3 GHz, Pentium IV personal computer. A Logitech colour webcam has been used. Signing person may be reasonably placed from the webcam (figure 6).

Using Intel OpenCV libraries, a 12 fps processing ratio has been achieved, thanks to the low computational cost of the proposed approach.

Table 1. Recognition ratios (static alphabetic signs)

<table>
<thead>
<tr>
<th>Letter</th>
<th>Recognition ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>99.3%</td>
</tr>
<tr>
<td>B</td>
<td>99.7%</td>
</tr>
<tr>
<td>C</td>
<td>96.3%</td>
</tr>
<tr>
<td>CH</td>
<td>95.7%</td>
</tr>
<tr>
<td>E</td>
<td>99.3%</td>
</tr>
<tr>
<td>G</td>
<td>85.0%</td>
</tr>
<tr>
<td>H</td>
<td>55.3%</td>
</tr>
<tr>
<td>I</td>
<td>97.7%</td>
</tr>
<tr>
<td>K</td>
<td>86.0%</td>
</tr>
<tr>
<td>L</td>
<td>98.0%</td>
</tr>
<tr>
<td>M</td>
<td>91.7%</td>
</tr>
<tr>
<td>N</td>
<td>95.7%</td>
</tr>
<tr>
<td>O</td>
<td>97.7%</td>
</tr>
<tr>
<td>P</td>
<td>99.0%</td>
</tr>
<tr>
<td>Q</td>
<td>92.0%</td>
</tr>
<tr>
<td>R</td>
<td>93.3%</td>
</tr>
<tr>
<td>S</td>
<td>62.0%</td>
</tr>
<tr>
<td>U</td>
<td>97.7%</td>
</tr>
<tr>
<td>W</td>
<td>98.7%</td>
</tr>
</tbody>
</table>

Recognition results obtained on some static (or statically identifiable) alphabetic signs of the Spanish Sign Language are shown in Table 1. The different gestures has been signed a number of times, under sign language commonly-expected variability conditions. The signing persons did not have any feedback about the sign recognized (or not) by the system. The correct sign identification ratios are reported. It is shown that the procedure works suitably for most letters, even reaching a 99% for some of them. Most no reported cases correspond to moving gestures, not addressed in the present paper. F and T letters are not reported because their contours are too similar (figure 7), so a further region based analysis would be required.

It should be noted that results in Table 1 correspond to the worst case: there is no feedback from the system to the signing person (the signing persons does not know if the sign is being correctly recognized). However, in practice the signing person is expected to be in front of a computer screen (or other display device), so that the user can notice the recognition result and thus modify the hand pose on-the-fly, thus the sign is correctly recognized. The tests we have performed in this situation have shown that a 100% recognition ratio can be easily approached.

Figure 7. Letters ‘F’ and ‘T’ in LSE.

4. Conclusion

Deaf-people interfacing is a very challenging issue, due to its cardinal social interest, and its inherent complexity. In the present paper an approach has been presented, which works suitably for static letters of the Spanish Sign Language. It is a low time-consuming approach, so that a real-time recognition ratio is easily achieved. Furthermore, a 100% recognition ratio can be easily approached by providing some feedback from the system to the signing person. In fact, this would be the case in most practical situations, where the user signs in front of a computer screen.

Figure 8. Letter ‘Z’ in LSE.
Next research will focus in dynamic signs, where simultaneous gesture-shape and hand trajectory recognition is required (figure 8). Moreover, of course a vocabulary analysis would improve the results (e.g. will allow F and T to be discriminated). Finally, hand segmentation without gloves would be desirable, but is disturbed by face, arm and either some clothes.

5. Acknowledgement

This work has been partially supported by the Spanish Profit program (Project Nr. FIT-350300-2004-22).

6. References


