Hand Tracking and Trajectory Analysis for Physical Rehabilitation

Giulia Boato, Nicola Conci, Mattia Daldoss, Francesco G.B. De Natale and Nicola Piotto

DISI, University of Trento
Via Sommarive 14, 38123 Trento - Italy
{boato, conci, daldoss, piotto}@disi.unitn.it
denatale@ing.unitn.it

Abstract—In this work we present a framework for physical rehabilitation, which is based on hand tracking. One particular requirement in physical rehabilitation is the capability of the patient to correctly reproduce a specific path, following an example provided by the medical staff. Currently, these assignments are typically performed manually, and a nurse or doctor, who supervises the correctness of the movement, constantly assists the patient throughout the whole rehabilitation process. With the proposed system, our aim is to provide medical institutions and patients with a low-cost and portable instrument to automatically assess the rehabilitation improvements. To evaluate the performance of the exercise, and to determine the distance between the trial and the reference path, we adopted the Dynamic Time Warping (DTW) and the Longest Common Sub-Sequence (LCSS) as discriminating metrics. Trajectories and numerical values are then stored to track the history of the patient and appraise the improvements of the rehabilitation process over time. Thanks to the tests conducted with real patients, it has been possible to evaluate the quality of the proposed tool, in terms of both graphical interface and functionalities.

I. INTRODUCTION

Before man could use spoken language, communication between human beings was only possible using gestures [1]. In fact, the visual characterization of a concept through the use of hands is usually self-explanatory and can significantly improve the interaction quality among subjects in a very natural form. For this reason gesture recognition, hand detection and tracking are well-known challenging research topics [2][3].

In particular, the capability of recognizing a hand and its posture can be applied to a wide range of applications, from simple text typing, to the manipulation of complex 3D scenes [4], such as immersive telepresence or interactive gaming. More in general, the exploitation of gesture recognition allows seamless interaction with a computer, pushing towards a new paradigm, in which the cooperation between the user and the machine can be achieved in a natural and non-intrusive manner.

The first attempts to adopt hands for computer interaction applications resulted in very uncomfortable devices that required the use of particular gloves or additional (mainly cabled) sensors capable of translating human movement into an electric signal [5][6]. Nowadays, the use of touchscreens is probably the most widespread solution to replace the input keyboard and it allows a direct interaction between the user and the machine, keeping the complexity of the whole architecture hidden. Unfortunately, they can only be applied to medium-small devices, since their cost is still reasonably high. Such drawbacks can be avoided by exploring vision-based solutions.

The work proposed in [7] can be considered as a pioneering application, revealing that not only hand recognition and tracking play a major role in dealing with augmented reality scenarios; also gestures can be interpreted and exploited to provide additional functionalities [8], when used as input peripherals, and as a replacement for common devices like mouse and/or keyboard. A recent implementation of a hand and gesture recognition tool has been presented in [9], where the authors adopt a vision-based system to develop a multi-purpose virtual blackboard, to be adopted in home automation, interactive gaming or to handle virtual objects projected on a screen.

Exploiting these concepts, we present in this work a tool that has been adopted as a rehabilitation instrument to improve the sensibility of hands and arms in recovering from physical accidents or cognitive disabilities. The target is to provide patients with a low-cost and reliable tool that allows carrying out the rehabilitation process at home, in an automatic way, without requiring the assistance of the medical staff relieving the patient from the bother of going to the hospital. The application contains exercises representing different shapes with different complexity, and the performance of the users are stored to evaluate the rehabilitation process. Through a remote interface, the medical personnel can access the exercises databases of each user and determine the improvements through numerical and graphical results.

The paper provides an overview of the implemented system in Section II, while in Section III the metrics used for evaluation of the exercises are described. Section IV discusses the experimental setup, together with a selection of numerical results to assess the reliability of the proposed approach.
II. THE PROPOSED REHABILITATION TOOL

The proposed rehabilitation tool aims at supporting both patients and medical staff, and exploits the hand as a replacement of the mouse, thus achieving a natural interaction with the machine. During physical rehabilitation one crucial target consists in recovering the capability to correctly reproduce a specific trajectory, following an example provided by the doctors. The goal of this work is to provide an efficient framework to automatically assist patients while training, by providing ad-hoc exercises, together with quality evaluations, and information about the rehabilitation history to evaluate the patients’ improvements.

The developed tool consists of two main building blocks: i) the detection and tracking of hands and of ii) the performance evaluation of a specific action. This section will concentrate on the first aspect, identifying the key features required to develop a flexible, portable and reliable tool for rehabilitation. It is worth noting that in this case we need to provide the user with a fully-functional device, and therefore well-known and consolidated systems have been considered during the deployment phase.

After acquiring the image, the data flow for the video processing is composed by the following steps:

A: - Color adjustment;
- Background suppression (Fig. 1(right))
- Extraction of skin-like areas;

B: - Image classification to detect the presence of hands in the scene;
- Hand detection and tracking (Fig. 1(left)).

In the next subsections all phases are described in detail.

A. Pre-processing

The pre-processing routines typically aim at improving the image quality for a more reliable processing of the visual contents. In this case the scope is to remove all the spurious information that might affect the detection and tracking of the hands. Often, hand tracking algorithms can be fault-tolerant, but in this case we must reduce the influence of external noise in order to be able to track the centroid of the hand; we have therefore introduced few additional algorithmic steps.

A background suppression procedure has been implemented, which is performed in two steps. First of all the background is removed adopting a simple differencing with an empty scene model. This simplification is reasonable since the workspace conditions are typically stable throughout the exercise and do not affect the background area, apart from the presence of shadows that are properly removed using the routine proposed in [10] and shown in (1), where $SM(x, y)$ is the shadow mask, $I(x, y)$ is the current frame, while $B(x, y)$ is the background model. The apices $H$, $S$, and $V$ refer to the hue, saturation and value of the HSV color space, respectively. The routine has been originally developed for visual surveillance applications, but the appropriate setting of the thresholds allows for a seamless integration in our framework with no particular drawback. The values we have chosen depend on the illumination conditions of the environment, but are in general set according to Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.05 - 0.85</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.90 - 0.95</td>
</tr>
<tr>
<td>$\tau_S$</td>
<td>0.15 $\times$ max($S$)</td>
</tr>
<tr>
<td>$\tau_H$</td>
<td>60 - 120</td>
</tr>
</tbody>
</table>

The remaining areas are analyzed to extract only the portions of the image that contain pixels recognized as “skin”. Since the color of the skin among different ethnic groups typically differs only in the saturation, we have only considered the hue component in the HSV domain.

B. Hand recognition and tracking

Even though the tracker itself is out of the scope of this work, we report anyway the main features of the algorithms we have adopted to detect and track the hand in the image plane. Skin detection is obtained using a basic algorithm exploiting the hue of the image as a discriminating feature. As far as the classification process is concerned, hand detection is performed adopting AdaBoost [11] as a boosting algorithm, thanks to its capability of managing real-time applications like the one presented in this work. The boosting algorithm, returns the area belonging to the hand, by calculating the integral image using the Haar features, making it possible to successively activate the tracker (Fig. 1(left)).

The tracking block relies on a cascade of two different modules: the Lucas-Kanade optical flow [12] and CAMSHIFT (Continuously Adaptive MeanSHIFT) [13]. During the initialization phase, AdaBoost detects the hand of the patient. This is performed by placing the hand in the central part of the screen in a pre-defined area: the system can then acquire both the color distribution (tracked by CAMSHIFT), and the spatial features (corner points), which are tracked by Lucas-Kanade algorithm. The program performs then a periodic re-initialization that can be forced also in the case the tracker fails. In case of failure the hand is then searched in the whole
\[ SM(x, y) = \begin{cases} 
\alpha \leq \frac{I^V(x, y)}{B^V(x, y)} \leq \beta \land |I^S(x, y) - B^S(x, y)| < \tau_S \land |I^H(x, y) - B^H(x, y)| < \tau_H \\
I(x, y) \text{ otherwise} 
\end{cases} \]  

(1)

III. Exercise Evaluation

After the hand is correctly detected, the user performs the exercise and the resulting curve is matched with the template to evaluate its accuracy. The test and target trajectories obtained through the tracking module of the interface are smoothed via spline interpolation to reduce the noise introduced by the tracker itself and by the presence of outliers. In order to assess the quality of the exercises, we introduced two trajectory-matching schemes: the former is based on the Dynamic Time Warping (DTW) technique [14], while the latter is based on the Longest Common SubSequence (LCSS) algorithm [15]. The enabling factor of these techniques is the capability of warping the sequences locally in order to accommodate the best alignment. The stretch may be operated in the temporal or spatial domain, providing time warping in DTW and spatial warping in LCSS.

DTW is a distance measure used in 1-D time-series comparison [16], initially applied to speech signals. It relies on a classical distance operator (e.g., \( L_p \)-norm) and on a particular matching procedure that finds the optimal alignment between the query and the target series allowing "temporally warped" matches between samples. The matching score is given by the cumulative distance between all samples in the query sequence and the corresponding nearest samples in the target. In particular, the distance between two generic series \( X \) and \( Y \) is related to the optimum warp path that minimizes the overall warping distance:

\[ D(X, Y) = \min \left\{ \frac{1}{K} \left[ \sum_{k=1}^{K} w_k \right] \right\} \]  

(2)

where \( w_k \) is the minimum distance between 2 samples indexes (one from \( X \), one from \( Y \)) in the \( k^{th} \) elements of the warp path.

As far as spatial warping is concerned, given a sequence pair, LCSS discovers the maximally long subsequence between the query and the target. In this case, the concept of subsequence provides a more flexible structure, not requiring consecutive samples. This allows to effectively cope with different sampling rates and to handle problems related to noise and outliers. In [15] the authors define a spatio-temporal region over the query samples and simply verify whether the aligned samples of the target fall within it (match) or not (mismatch). In particular this region is defined by two threshold \( \epsilon \) and \( \delta \), in space and time respectively. Given two time series \( X \) of length \( N \) and \( Y \) of length \( M \), their distance is evaluated by dynamically computing \( LCSS_{\epsilon, \delta} \) as in (3), where \( H(X_i) = \{ x_1, x_2, \ldots, x_{i-1} \} \) and \( H(Y_j) = \{ y_1, y_2, \ldots, y_{j-1} \} \). The LCSS algorithm outputs the length of the longest common subsequence between the series. A similarity score in the range [0-1] is then defined as follows:

\[ S(X, Y, \epsilon, \delta) = 1 - \frac{LCSS_{\epsilon, \delta}(X_M, Y_N)}{\max\{M, N\}} \]  

(4)

Fig. 3 sketches the difference between the alignments of the same sequence pair, obtained through a temporal warping (a) and a spatial warping (b), respectively. The LCSS algorithm leads to a more significant alignment, since it allows excluding some samples from the matching process; on the other hand, the DTW approach requires to match every sample of the query, thus causing possible one-to-many correspondences (as in Fig. 3 (a)).

Both techniques have been employed in order to reliably face time/space warp. Indeed, due to the need of reproducing
\[ LCSS_{\epsilon, \delta}(X_i, Y_j) = \begin{cases} 
0 & \text{if } i = 0 \text{ or } j = 0 \\
1 + LCSS_{\epsilon, \delta}(H(X_i), H(Y_j)) & \text{if } |x_i - y_j| < \epsilon \text{ and } |i - j| < \delta \\
\max \{ LCSS_{\epsilon, \delta}(H(X_i), Y_j), LCSS_{\epsilon, \delta}(X_i, H(Y_j)) \} & \text{otherwise} 
\] (3)

A target path with the best accuracy, a sensible difference in the total number of samples can be noticed between the target path (proposed by the medical staff and inserted in the collection of exercises) and the output generated by the patient. To tackle this different sampling rate and obtain a reliable matching score without the need of resampling the whole path, we adopted matching algorithms specifically designed to handle such fluctuations.

IV. EXPERIMENTAL SETUP AND APPLICATION

Since the objective of this work is to provide a versatile tool to be widespread also in the patients’ homes and not only in rehabilitation institutions, we have considered the accessibility as the main constraint for the implementation. Therefore we have adopted a standard PC and a common webcam (VGA) mounted on an ad-hoc stand for the acquisition (Fig. 4). The camera used for acquisition is fixed at 60cm from the desk and the area used to perform the exercise is 60x50cm. The current implementation relies on the OpenCV software library [17].

![Fig. 4. Experimental setup.](image)

The adopted visual interface partially reflects the implementation in [9], even though a more accurate classification process has been carried out. In particular, the major improvements can be noticed by calculating the Fitts Law [18], which efficiently measures the quality of a pointing device for visual interfaces through the so-called Index of Performance (IP). This value is calculated by taking into account the size and distance of the target for a certain pointer. We have therefore selected the buttons size and window structure in such a way to have a similar accuracy to the mouse. In our previous implementation of the virtual blackboard in [9], the average IP was 7.63, while in the current implementation we achieve an average IP = 8.48 using different distances \( D \) from the target and different size of the target itself. In Table II the numerical results for the validation of the interface using the Fitts Law are shown. The target is a square (length of the side \( W \)), and the values for both the distance from the target and the size of the target itself are expressed in terms of pixels.

<table>
<thead>
<tr>
<th>( D )</th>
<th>( W )</th>
<th>IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>8</td>
<td>14.20</td>
</tr>
<tr>
<td>64</td>
<td>16</td>
<td>14.72</td>
</tr>
<tr>
<td>64</td>
<td>64</td>
<td>14.09</td>
</tr>
<tr>
<td>128</td>
<td>8</td>
<td>4.78</td>
</tr>
<tr>
<td>128</td>
<td>16</td>
<td>7.54</td>
</tr>
<tr>
<td>128</td>
<td>32</td>
<td>11.36</td>
</tr>
<tr>
<td>128</td>
<td>64</td>
<td>13.19</td>
</tr>
<tr>
<td>256</td>
<td>64</td>
<td>7.91</td>
</tr>
<tr>
<td>256</td>
<td>16</td>
<td>5.97</td>
</tr>
<tr>
<td>256</td>
<td>32</td>
<td>5.20</td>
</tr>
<tr>
<td>512</td>
<td>64</td>
<td>6.45</td>
</tr>
<tr>
<td>512</td>
<td>32</td>
<td>5.72</td>
</tr>
<tr>
<td>512</td>
<td>16</td>
<td>4.78</td>
</tr>
<tr>
<td>512</td>
<td>8</td>
<td>4.93</td>
</tr>
</tbody>
</table>

Fig. 5. Sample trajectories used for performance evaluation. Curves (a), linear segments (b), mix of curves and linear segments (c).

As it can be noticed from the snapshot in Fig. 4, the hand is adopted as a replacement of the mouse and the user can have a feedback of his movements by looking at the display. According to the medical staff’s specifications, each exercise must take into account also the execution time and not only the spatial match. Additional information must be then provided in terms of velocity, in order to highlight the areas where the user encounters major difficulties.

Fig. 5 shows a concise set of three sample exercises, presenting different features: curves (a), linear segments (b), mix of curves and linear segments (c). In our experimental session, each exercise has been performed 10 times. Different trials of the ‘3’, ‘1’, and ‘5’ exercises are shown in Fig. 6, 7, and 8, respectively. To determine the corresponding score, the numerical results for the trajectory alignment are evaluated and averaged, as presented in Table III. In the first column of
the ID of the track is indicated. IDs are assigned on the basis of the shape of the exercise and, similarly to the examples in Fig. 6, 7, and 8, they include an almost perfect trial (a), a second test with less precision (b), and a third sample that strongly diverges from the original track. The matching scores are calculated using both LCSS and DTW. For the LCSS matching scheme, a spatial threshold of 40 pixels (to determine the pertinency of the pixel with the original track) and a time window of 5 frames are chosen. The DTW implementation uses instead the Itakura local constraints and weights uniformly all points.

In order to prove the reliability of the matching algorithms we have chosen, we also tested the rehabilitation tool by using the approximate matching technique proposed in [19]. The implementation in [19] is a modified version of the edit distance [20] originally developed to match genomic sequences [21]. The method turned out to be effective even in comparing generic strings of samples and it has been successfully applied to visual surveillance applications to compare trajectories of moving objects. The adaptation to fit the requirements imposed by this work is straightforward and it consists mainly on the proper setting of a few parameters. However, also by properly tuning the thresholds, the method has has revealed major weaknesses, and the incapability of properly disclosing exercises of similar shapes but with different accuracy. For this reason we have excluded it as a candidate for matching.

Results are shown in the last column of III. Values are normalized in [0, 1] for both LCSS and approximate matching, being 1 the perfect match. In case of DTW there is though no possibility of expressing the distance in a normalized form, since values can span in the range [0, ∞), being 0 the perfect match.

![Fig. 6. Different trials of the '3' exercise.](image)

![Fig. 7. Different trials of the '1' exercise.](image)

![Fig. 8. Different trials of the '5' exercise.](image)

![Fig. 9. Graph reporting the speed of the performance for a sample '5' exercise.](image)

**TABLE III**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>LCSS</th>
<th>DTW</th>
<th>Approx. Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 vs 3a</td>
<td>0.87</td>
<td>4.23</td>
<td>0.63</td>
</tr>
<tr>
<td>3 vs 3b</td>
<td>0.81</td>
<td>6.96</td>
<td>0.50</td>
</tr>
<tr>
<td>3 vs 3c</td>
<td>0.71</td>
<td>7.8</td>
<td>0.62</td>
</tr>
<tr>
<td>1 vs 1a</td>
<td>0.84</td>
<td>3.32</td>
<td>0.58</td>
</tr>
<tr>
<td>1 vs 1b</td>
<td>0.71</td>
<td>5.44</td>
<td>0.44</td>
</tr>
<tr>
<td>1 vs 1c</td>
<td>0.62</td>
<td>7.62</td>
<td>0.54</td>
</tr>
<tr>
<td>5 vs 5a</td>
<td>0.84</td>
<td>5.62</td>
<td>0.59</td>
</tr>
<tr>
<td>5 vs 5b</td>
<td>0.69</td>
<td>6.70</td>
<td>0.35</td>
</tr>
<tr>
<td>5 vs 5c</td>
<td>0.12</td>
<td>8.08</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Table III shows how both LCSS and DTW can be efficiently adopted for matching, since both of them are capable of highlighting the differences between the reference path and the users’ trials. To provide additional information to the medical staff about the quality of the trial, a plot referring to velocity (pixels per second) is also included (see Fig. 9 as an example) in the user interface.

It is worth noting that rehabilitation quality can not be evaluated as a single measure, i.e. a matching score, since the whole process can take months, or even years to be
accomplished. This makes the score of the exercise per se meaningless, if not compared with the patients’ history. For this reason, in our implementation we have provided the medical personnel with a visual interface that shows for each patient a full clinic overview, including the types of exercises, the score of each trial, and the history. The history is essential in determining the improvements with respect to both the previous trials, and compared with the average of all other patients.

V. CONCLUSION

In this work we have proposed a portable user interface for physical rehabilitation. The framework is designed to provide an efficient tool to automatically assist patients in recovering from different kinds of traumas. The execution of an exercise is complemented with an assessment stage that evaluates the evolution of the rehabilitation process over time, providing, needless to say, big advantages to both the patient and the doctor. To demonstrate the functionalities of the proposed system we have reported three sample exercises regarding trajectory matching, each of them dealing with different types of exercises. To evaluate the similarity between the current exercise and the reference path, the alignment scores are calculated using DTW and LCSS and both of them result to be suitable for the application scopes.

ACKNOWLEDGMENT

This work has been partially developed under the A-Cube project, funded by the Provincia Autonoma di Trento (Italy). The authors would like to acknowledge Rudi Ferrari for his precious help in the development phase and Doctor Giovanni Guandalini (Villa Rosa Hospital - Trento - Italy) for his guidelines on parameter selection and performance evaluation.

REFERENCES


