The Computer Vision Box & Block Test in Rehabilitation Assessment

Orestis N. Zestas
CNA Lab - Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
o.zestas@go.uop.gr

Dimitrios N. Soumis
CNA Lab - Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
dimsoumis@go.uop.gr

Kyriakos D. Kyriakou
CNA Lab - Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
kyriakou@uop.gr

Kyriaki Seklou
CNA Lab - Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
kseklou@uop.gr

Nikolaos D. Tselikas
CNA Lab - Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
ntsel@uop.gr

***Abstract*—Post-stroke patients very commonly present upper limb deficits, while their rehabilitation comprises regular monitoring and kinematic assessments to evaluate motor recovery. One of the most used and recommended tools to objectively measure upper limb dexterity is the Box and Block Test (BBT). However, the test itself is time consuming and very cumbersome.**

After briefly presenting a literature review, this paper proposes a computer vision (CV-BBT) approach to the traditional BBT, as a virtual alternative of the real world procedure. Our CV-BBT integrates all the original BBT's guidelines and procedures into an interactive computer vision experience that utilizes bleeding edge technologies such as MediaPipe Hands for hand and finger tracking. This innovative tool require neither any additional computer peripherals (smart gloves, VR headsets) nor any kind of extra physical equipment (wooden box, blocks), but works instead with just a mid-range PC and a camera. Our system can be deployed in residential spaces and the test results can be sent remotely to any physician or rehabilitation expert. The application implementation is also demonstrated and conferred in-depth.

Finally, we shortly discuss some technical issues of our computer vision approach, essentially being the hand's pose prediction accuracy and processing times, as well as present some future directions regarding our tool's score normalization of healthy patients against those achieved with the original BBT.

***Index Terms*—Box & Block Test, Computer Vision, Rehabilitation Assessment, Upper-limb Rehabilitation, Telemedicine**

I. Introduction

Stroke is a prevalent and disabling global healthcare problem, the rehabilitation of which is a major part of patient care. Upper limb hemiplegia and impaired finger dexterity are some common dysfunctions that affect nearly 80% of post-stroke patients [1]. Thus, clinicians recommend systematic motor assessments and consistent hand function monitoring to patients that suffered a stroke, in order to evaluate their manual dexterity in every-day tasks. While many post-stroke rehabilitation processes have already started to take place at patients' home, clinical evaluation by visiting the physician remains crucial. Modern technology helps in this direction by introducing terms such as "telemedicine" and "home healthcare technology", which not only enable the patient to be distantly monitored, but also induce him/her to participate in contemporary neurorehabilitation practices, such as virtual reality and computer vision approaches of standardized real-world tests and exercises.

In this paper, we examine an innovative processing tool, the Computer Vision Box and Block Test (CV-BBT). Our approach aims to provide a fast, portable and lightweight solution which will not only equip caregivers with a tool for distant monitoring and accurate rehabilitation assessment, but will also present patients with a means to relearn basic hand and finger movements, from the comfort of their own home.

The Box and Block Test (BBT) is one of the most used and recommended tools to evaluate unilateral manual dexterity [2]. The BBT consists of a wooden box and one hundred and fifty blocks. The box is separated in the middle by a divider of about three blocks in total height, splitting the area into two equal compartments. The blocks are placed in one compartment of the box, facing the patient's dominant hand. The patient has exactly one minute to move as many blocks as possible from one compartment of the box to the other, while picking up and transporting over the partition divider only one block at a time. The same process restarts for the non-dominant hand. An example demonstration of BBT can be seen in Figure 1. Ideally, the whole procedure should be recorded, in order to be carefully examined later by the

979-8-3503-9958-5/22/$31.00 ©2022 IEEE

physician and reach more accurate conclusions, making it all the more tedious and time consuming.

Our computer vision approach eliminates the original BBT's drawbacks, while incorporating the entire standardized procedure along with accurate hand and finger tracking (using MediaPipe Hands), bundled in a single, fast application. No extra equipment is needed, such as a wooden box, blocks, smart gloves, VR headsets, etc, apart from a mid-range PC and a single camera (built-in/external).

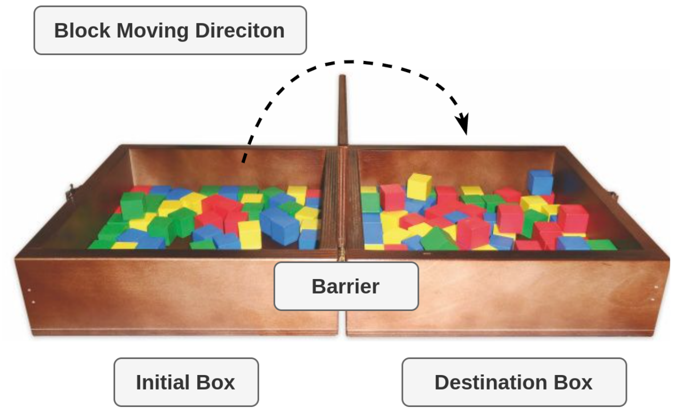


Fig. 1. Demonstration of the original BBT.

The rest of this paper is organized as follows. Section II provides a literature review regarding related virtual BBT approaches. In Section III, we analyze and demonstrate our computer vision based BBT application. Section IV summarizes some technical issues regarding the implemented CV-BBT application, as well as presenting our ultimate goal from our ongoing study regarding CV-BBT. Lastly, the paper is wrapped up in Section V.

II. Literature Review

At the time of writing, the literature contains only five studies that have implemented a virtual BBT. One approach takes advantage of non-immersive VR technologies and a depth sensing camera [3]. Though, while the results are promising when compared to the original BBT, the equipment needed is expensive, not easily transportable, and does not provide accurate visualization of the patient's hand. Another study suggests a virtual BBT comprised of active hand tracking along with an immersive VR headset [4]. On average, patients here scored around thirty five blocks less than the real-world equivalent scenario, according to the metrics. Two additional published papers propose similar immersive VR implementations to assess unilateral manual dexterity using controllers, such as working with a Leap Motion Controller (Ultraleap) [5], or an Oculus Quest controller [6]. The first one compares results captured from the virtual BBT to those of the original BBT in patients with Parkinson's disease, while the second assesses the concurrent validity among healthy and post-stroke patients, while providing hand kinematic analyses. Finally, a digital BBT was proposed in 2013 [7], utilizing a depth-sensing camera (Microsoft Kinect), an existing wooden box and blocks. The camera connects to a host computer to detect and record the motion data during a test session, in order to go beyond the original BBT's scheme.

In essence, the above studies require either extra equipment, which is often expensive and reduce the overall application portability, or powerful enough computer hardware to construct and simulate the testing environment.

III. Computer Vision Approach

Our Computer Vision Box and Block Test (CV-BBT), goes beyond the basic functionality of the original BBT, making the entire procedure virtual, without needing any kind of specialized computer peripherals nor substantial computing power, but a mid-range PC and camera. It bundles a high quality implementation of modern computer vision and image recognition algorithms, in a portable but lightweight standalone application. Figure 2 illustrates the high level architecture of our proposed implementation.

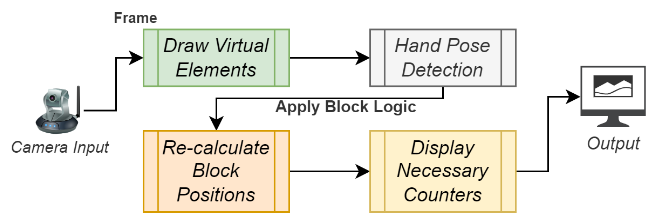


Fig. 2. The proposed architecture.

As long as the camera is positioned about 50cm from the hand and there is enough natural light present in the testing environment, the system is able to perform in its optimal state and successfully track the patient's hand and fingers.

A. Software Modules

CV-BBT is developed using OpenCV [8] and MediaPipe's Hands solution [9]. We took advantage of the OpenCV library to process each individual frame in near real-time, and perform the necessary drawings of all the virtual elements. Then, using MediaPipe Hands, which implements a solid model and pipeline architecture, the application provides fast and high prediction quality of human hands. The development of our system can be broken down into three key sections: Setting up and modeling the computer vision environment, detecting and tracking the hand plus fingers, and finally applying the operation logic for each captured frame. CV-BBT is open sourced at *https://gitlab.com/o.zestas/cv_box_and_block*.

B. Environment Modeling

Starting off, we need to construct the computer vision environment of the BBT. Since we make use of a single RGB camera, we can begin by capturing each frame repetitively, and feeding it as input to our processing algorithm. For our testing, we used a 60fps (frames per second) with a resolution of 720p camera (1280x720), connected to our host PC. We begin by initially applying horizontal flipping of the image, with the subsequent horizontal shift and absolute difference calculation to check for a vertical-axis symmetry, using OpenCV. We continue by drawing some basic components like the title, timer, score counter and partition divider. In order to avoid cluttering the screen and keep a low memory footprint, each one of the 150 blocks is being generated one-by-one, meaning that after the starting block generation, each new block will be spawned on each successful block movement by the patient. This way, the objective stays clear throughout the test and drawing numerous elements on each frame is not necessary. A move is considered successful when the block is moved from its starting compartment to the other, without touching the partition divider in the middle, following the original BBT's guidelines. These successfully transferred blocks persist on each frame draw (though there is a limit how many are visible, to avoid unnecessary memory hogs) for the sake of promoting positive reinforcement. The color of each block can be one of four; red, green, blue or yellow, along with a transparency filter applied. This is done in favor of keeping the background moderately visible, and not letting it obstruct with the patient's point of view and hand/finger movements.

C. Hand Detection & Operation Handling

MediaPipe's bleeding edge hand tracking library utilizes an ML pipeline consisting of two models working together:

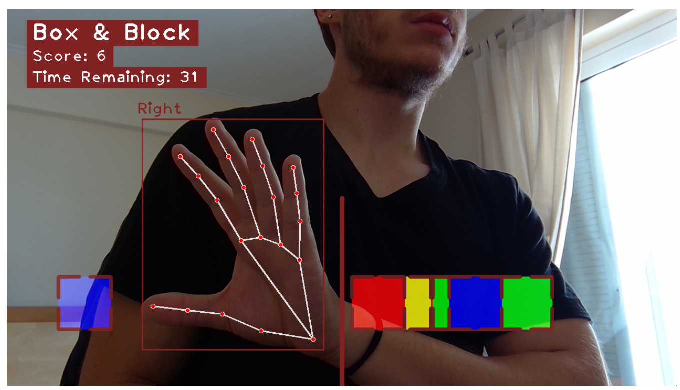
- A palm detector that uses a full input image to search for palms and locates them using an oriented hand bounding box.
- A hand landmark model that uses the cropped hand bounding box provided by the palm detector to generate high-quality 2.5D landmarks.

Doing so highly reduces the need for data augmentation algorithms, such as scaling and translating the image, while enabling the system to devote all of its computational power towards producing accurate hand and finger landmarks. Since we are adopting real-time hand tracking and speed is of the essence, the bounding box that is used in the previous frame (iteration) of the landmark prediction is now being fed as input to the current frame, which means that the detector is not being applied every time. Instead, MediaPipe's hand solution applies the detector only on the first given frame, or when the hand is lost from the box.

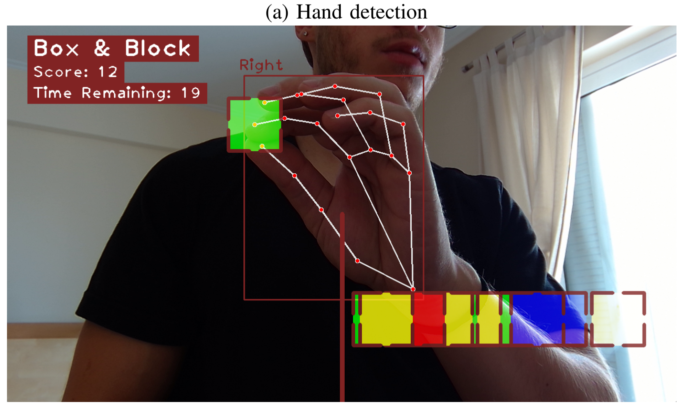
Now that a solid means of hand detection and tracking is implemented, we continue by applying the algorithm to each frame captured by the camera. For our CV-BBT, only one hand needs to be recognized and tracked throughout the process. If more than one hand is visible, then the first one to be detected is the one that the patient will work with, while the other will be ignored. As performed in the original BBT, the user must grab one block with his hand before moving it to the other side of the box. Hence, in the CV-BBT we need to initially estimate the hand's pose and calculate if at least the thumb and index fingers are closed, in order to begin emulating a "grabbing" behavior. This means that the distance between these two fingers must be lower than a preset threshold, a number that we adjusted to work best when the hand is about 50cm from the camera. Calculating two-finger distances is a process of applying the Pythagorean theorem to the triangle that's formed by the fingers themselves, where a single edge is considered the palm to finger-tip distance. Thus, given the index and thumb finger-tips, we can compute their distance on every frame. In order for a patient to be able to hold a virtual block, we must also calculate the hand's position relative to the camera. When its location resides within the region of a block, and its pose has the fingers closed, then the targeted block is considered movable. Since frames are captured consecutively for the whole duration of the test, we are able to calculate each next position of a single virtual block on each frame, and follow the course of the patient's hand.

Continuing, we apply collision checking only for the moving block, such that it's not able to pass through the partition divider, and only on top of it. This is relatively easy to implement, since we have already calculated the real hand's position we are able to instantly check if the block collides with the middle divider in the current frame.

Once a virtual block has passed on top of the solid divider and it's been released from the patient's hand, "gravity" is immediately attached to it and begins to fall down into the respective compartment of the box. Since this is considered a successful block move, the patient is rewarded with a point. We have also added the necessary out-of-bounds checks, so that the blocks remain within the virtual box's region.



(a) Hand detection



(b) Moving a virtual block

Fig. 3. Demonstration of the proposed CV-BBT.

D. Example Demonstration

Figure 3 demonstrates an example test run of the CV-BBT. When a hand is detected by the system, an annotation is displayed on top of the bounding box, indicating whether it is the left or right hand, as seen in Figure 3(a). The hand landmarks remain visible throughout the test. In Figure 3(b), the process of grabbing and moving a virtual block is also presented, while showing that the active hand tracking is accurate, even on challenging hand pose predictions.

From our testing, the CV-BBT works best with cameras that can capture frames at a rate of at least 60 per second, and a working resolution of 720p. Nonetheless, the test can perform decently even when running on a common laptop's hardware. In that case, the CV-BBT's resolution must be lowered to 480p. This is as easy as editing the script file and changing the target width and height to the preferred numerals. The system auto-calibrates itself on new resolution changes, along with properly adjusting the position of each element.

Launching the CV-BBT does not have any software requirements on the host PC, other than a working Python 3 installation. The system comes with a pre-configured set of all the necessary libraries and tools in a separate file, which makes the software management easier and highly adaptable. In addition, the implementation is easy to expand upon if needed, in order to facilitate extra functionality.

IV. Discussion & Further Development

Programming an interactive computer vision application, which aims to provide accurate real-world rehabilitation assessments and test emulations, can be a very tedious process. The detection of the patient's hand is vulnerable to many environmental factors such as room lighting conditions, polychormatic background, or even the user's hands themselves. In addition, the test's accuracy and results need to be comparable to the corresponding real world scenario, as performed with a physiotherapist's supervision, thus it is critical that every captured frame is processed in near real time. Note that apart from hand detection, each frame needs to be post-processed in order to draw the necessary virtual elements and apply their logic. As a result, such an application can be computationally very expensive. Our approach yielded very promising results even when running on mid-range hardware, while using MediaPipe in combination with the OpenCV library (30-40 milliseconds process time per frame).

In the near future, we intent to deploy this technically innovative approach as a valid and reliable tool, ready to be used by physiotherapists or other rehabilitation experts. As such, a minimum requirement should be the normalization of the CV-BBT's expected score values of healthy people, opposed to those performed using the legacy BBT, as presented for example in [10]. Though, such a task can prove to be rather difficult, since it requires a really large group of volunteers, whose ages must be uniformly distributed. Still, we have already started with this kind of study in our city (at the University campus of UoP), and our sample is sufficient only for young ages, ranging from 20-24 to 25-29 years of age. We have made an interesting observation regarding the CV-BBT's score normalization, and that is the mean value's ratio consistency between the original BBT and our implementation. Thus, we have come to the conclusion that our results seem to be quite promising, and stand as a first indication whether it's reasonable to continue with our study for all ages, or not.

V. Conclusions

We have described a tool to reliably measure the unilateral manual dexterity, using a computer vision approach of the validated test, the BBT. This implementation requires neither any additional physical equipment nor any kind of specialized computer peripheral, thus providing ease of access. We have also utilized a robust way of detecting and tracking the patient's hand and fingers, as well as keeping the application running at decent speed, even on mid-range hardware. We also plan to implement score normalization for 20+ years of age, since our initial indications point to very promising results. Our main objective is to provide a valid and reliable tool ready to be used by clinicians for accurate rehabilitation assessments.

Acknowledgment

Part of this work has been co-financed by the European Regional Development Fund of the EU and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (Project: Rehabotics, code: 5131369).

References

[1] Lawrence ES et al; Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. Stroke. 2001 Jun;32(6):1279-84. doi: 10.1161/01.str.32.6.1279. PMID: 11387487.

[2] Winstein CJ, et al., "Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart/Stroke Association". Stroke, 47 (6), June, 2016.

[3] Cho S, Kim WS, Paik NJ, Bang H. Upper-Limb Function Assessment Using VBBTs for Stroke Patients. IEEE Comput Graph Appl. 2016 Jan-Feb;36(1):70-8. doi: 10.1109/MCG.2015.2. Epub 2015 Jan 8. PMID: 25585413.

[4] Gieser, S.N., Gentry, C., LePage, J., Makedon, F. (2016). Comparing Objective and Subjective Metrics Between Physical and Virtual Tasks. In: Lackey, S., Shumaker, R. (eds) Virtual, Augmented and Mixed Reality. VAMR 2016. Lecture Notes in Computer Science(), vol 9740. Springer, Cham. https://doi.org/10.1007/978-3-319-39907-2_1

[5] Oña ED, Jardón A, Cuesta-Gómez A, Sánchez-Herrera-Baeza P, Cano-de-la-Cuerda R, Balaguer C. Validity of a Fully-Immersive VR-Based Version of the Box and Blocks Test for Upper Limb Function Assessment in Parkinson's Disease. Sensors (Basel). 2020 May 13;20(10):2773. doi: 10.3390/s20102773. PMID: 32414177; PMCID: PMC7285781.

[6] Everard, G., Otmane-Tolba, Y., Rosselli, Z. et al. Concurrent validity of an immersive virtual reality version of the Box and Block Test to assess manual dexterity among patients with stroke. J NeuroEngineering Rehabil 19, 7 (2022). https://doi.org/10.1186/s12984-022-00981-0

[7] Chih-Pin Hsiao, Chen Zhao and E. Y. -L. Do, "The Digital Box and Block Test Automating traditional post-stroke rehabilitation assessment," 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013, pp. 360-363, doi: 10.1109/PerComW.2013.6529516.

[8] Open CV. [Online]. Available: http://opencv.willowgarage.com/wiki/. [Accessed: 13-Oct-2010].

[9] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." arXiv preprint arXiv:2006.10214 (2020).

[10] Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the Box and Block Test of manual dexterity. Am J Occup Ther. 1985 Jun;39(6):386-91. doi: 10.5014/ajot.39.6.386. PMID: 3160243.