A FRAMEWORK FOR WEBCAM-BASED HAND REHABILITATION EXERCISES

Rui Liu, Burkhard C. Wünsche, Christof Lutteroth, Patrice Delmas

Department of Computer Science, University of Auckland, Private Bag 92019, Auckland, New Zealand
rliu030@aucklanduni.ac.nz, burkhard@cs.auckland.ac.nz, lutteroth@cs.auckland.ac.nz, p.delmas@cs.auckland.ac.nz

Keywords: Hand tracking, hand segmentation, skin classifiers, hand rehabilitation

Abstract: Applications for home-based care are rapidly increasing in importance due to spiraling health and elderly care costs. An important aspect of home-based care is exercises for rehabilitation and improving general health. In this paper we present a framework for demonstrating and monitoring hand exercises. The three main components are a 3D hand model, a high-level animation framework which facilitates the task of specifying hand exercises via skeletal animation, and a hand tracking program to monitor and evaluate users' performance. Our hand tracking solution has no calibration stage and is easily set-up. Segmentation is performed using a perception-based colour space, and hand tracking and motion estimate are obtained using novel variations to a CAMSHIFT and contour analysis algorithms. The results indicate that the robust tracking along with the demonstration and reconstruction of hand exercises provide an effective platform for hand rehabilitation.

1 Introduction

Reduced or complete loss of control of hand and fingers is a catastrophic event which considerably reduces patients' quality of life and their ability to participate in society. Causes include injuries (muscles, tendons, and/or bones), inflammatory and autoimmune diseases (arthritis), degenerative muscle diseases (Welander distal myopathy and myotonic dystrophy type), overuse syndromes (Carpal Tunnel Syndrome, repetitive strain injury) and neurological damage and diseases (stroke, Multiple Sclerosis, Parkinson's disease) (Fischer et al., 2007; Wessel, 2004). For many of the above diseases surgical or drug treatment does not exist, is expensive, or only partially effective. Finger and hand exercises are a very effective treatment and can assist with diagnosing and preventing some of these diseases (Wessel, 2004). Empirical evidence suggest that hand exercises can even improve motor functions of the upper extremities (Fischer et al., 2007; Boian et al., 2002).

A major problem with exercise training is that a qualified instructor must be present in order to train and supervise the patient and evaluate the success of the exercises over time. While patients are encouraged to exercise at home, many patients lack motivation and there is no supervision and monitoring. This reduces the effectiveness of exercises (Boian et al., 2002) and prevents the evaluation of the patient status.

We propose to use common consumer-level equipment, i.e., personal computers and web-cams, in combination with animation technology to teach patients the correct exercises and to monitor and evaluate them. Such a training platform can also be used to organise rehabilitation groups where patients meet in virtual space and encourage each other and compete to improve their health status. The suggested framework fits well into the move towards home telehealth, which is becoming an integral part of government health policies, for example in New Zealand and the UK.

Section 2 reviews some of the literature about hand rehabilitation and markerless hand tracking. Section 3 presents our 3D hand model and animation framework. Section 4 and 5 present our algorithms for detecting the hand in web-cam images and for analysing its shape and motion. An evaluation and discussion of our results in section 6 is followed by our conclusion in section 7.

2 Background

The human hand is composed of 27 bones. The joints connecting the bones have different numbers of degree-of-freedom (DOF) defining their range of possible movements. In order to describe and monitor hand exercises, the anatomy of the hand needs to be modeled. Various works have been done in this area, and a summary can be found in (Liu, 2010).

Hand rehabilitation exercises have been devised for hand strength, grip, wrist, fingers and dexterity (Handexercise.org, 2010) and their effectiveness has been demonstrated for a wide variety of diseases, e.g., rheumatoid arthritis (Wessel, 2004). Various ideas have been suggested to make exercises more interesting and motivate patients (Jack et al., 2001). Simple, but effective tools are demonstration slides and exercise records (Health Information Translations, 2010). For speed oriented exercises, a virtual butterfly was projected onto the hand and the user has to scare it away by closing the hand as fast as possible, and a virtual piano was used for finger fractionation for individually active and passive fingers (Jack et al., 2001). We have chosen two of the most common hand exercises for testing our framework: touching the thumb with the other fingers and abduction/adduction of fingers. Originally, those exercises were advised for arthritis patients. However, they are also being widely used for other purposes such as rehabilitation after stroke and Parkinson's disease.

A large variety of hand tracking algorithms has been proposed. A good survey is given in (Mahmoudi and Parviz, 2006). Two important categories are marker-based and marker-less methods. Our tracking system employs a marker-less method. Without markers the easiest way to identify (potential) hand shapes is by using a skin colour classifier. A large amount of literature exist on this topic and some of the more recent surveys and comparative studies include (Kakumanu et al., 2007; Vassili et al., 2003). A common drawback of skin colour classifiers is their sensitivity to changes in the illumination. Recently a perception-based colour space has been proposed (Chong et al., 2008). The authors argue that the separation of the foreground and background objects will not be affected by the global illumination changes. This property supports our goal of calibration free approaches which are also able to deal with the real-world lighting conditions and backgrounds.

Once a (potential) hand shape has been identified based on skin colour it must be verified and its 3D position and orientation must be determined. A popular way to achieve this is by using a 3D hand model and searching for a mapping between it and the perceived hand shape subject to the model's inherent constraints (e.g., joint constraints and rigidity of bones). An example is the technique by (Stenger et al., 2001), who estimate the pose of the hand model with an Unscented Kalman filter (UKF). The opposite approach can also be taken by matching a hand image to a set of hand templates (Stenger et al., 2006). A different approach is to perform the matching between hand features rather than the entire shape. Several authors use Haar-like features, which describe the ratio between the dark and bright areas within a region (Chen et al., 2007). Advantages include fast matching and rapid elimination of wrong candidates.

3 Hand Modeling and Animation

The open source 3D humanoid character modeling tool "MakeHuman^{TM}" (MHTeam, 2010) was chosen for creating our hand model. The animation is achieved using skeletal animation and skinning, considering anatomical constraints (Liu, 2010). Vertices are bound to one or several bones using a local *bone coordinate system* and vertex weights. If a bone moves then the mesh positions with non-zero weights for the corresponding bone move with it. Bones are connected by joints and hierarchical structured. A bone's position is defined relative to its parent bone.

We developed a high level control interface based on the specific components of hand exercises. This simplifies the specification of hand exercise animations and facilitates hand tracking by reducing the number of transitions from one hand configuration to another.

The *individual control library* contains natural movements of hand components defined as

$$Movement_i = \{Name_i, Components_i, Prerequisite_i,$$

$$StartState_i, EndState_i, Range_i\}$$

where *Name* is a unique identifier and *Components* are the parts of the hand model affected by the movement. A *Prerequisite* specifies a hand configuration necessary for starting a motion. The *StartState* and *FinalState* indicate the component positions, defined as 3D transformation matrices, at the start and end of the motion. *Range* is a percentage value describing the extend of a motion, e.g., 80% abduction of the index finger. Examples are shown in figure 3.

The *pre-baked control library* contains hand movements which are difficult to compute by using individual controls such as a pinch of two fingers. In addition it stores postures frequently used in different hand exercises, e.g., a full abduction posture of the

hand or a fist. These postures are pre-calculated rather than computed at run time. Examples are shown in figure 3.

4 Hand Segmentation

Hand region segmentation is achieved using pixel-based region-growing. A seed point is obtained by displaying a target (e.g., hand template) in the center of the screen and requiring the user to move their hand until it matches the target. The seed point is added to the currently detected hand region and all its neighbours are stored in a queue data structure. In each iteration a pixel is removed from the queue, tested whether it belongs to the hand region, and if yes its not yet tested neighbours are added to the queue.

In order to decide whether a pixel belongs to the hand region we use a perception-based colour space (Chong et al., 2008) and compute the current pixel's colour distance to the mean value of the hand region detected so far. Two pixels are considered to belong to the same region if their colour distance is within a given threshold. Finding a suitable value for thresholding is difficult. Making the inclusion of pixels only dependent on comparison with its direct neighbours is error prone because of "colour leaking". Gradient based methods are common in other segmentation applications, but work unsatisfactory because of local strong illumination changes over the hand surface. We combine instead a colour distance criteria with edge detection. Further improvements are achieved by combining local and global thresholds and using multiple seeds for region-growing. Inappropriately positioned seed points are avoided by deploying histogram matching (Bradski and Kaehler, 2008). More details are given in (Liu, 2010).

5 Hand Tracking

Hand tracking is achieved using a continuously adaptive mean-shift (CAMSHIFT) algorithm (Bradski, 1998). By tracking the local extrema in the growing mask, the seed points for the region growing algorithm are extracted for each subsequent frame. Rather than computing the probability distribution of the histogram for the *histogram back projection* step, we recenter the mean window according to the region growing result. Therefore the ability of tracking the hand is enhanced by filtering out the region with lower probability before growing. In addition, the perceptual distance between hand and other skin-colour interferences helps preventing wrong segmen-

tations. With this approach, the previous calculation for the center of the mean shift window can be rewritten as the accumulation of the pixels which are recognised in the region and are under the mean window as shown in equations 1–3.

$$M\prime_{00} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} R(x,y) \tag{1}$$

$$M\prime_{10} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} xR(x,y) \tag{2}$$

$$M\prime_{01} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} yR(x,y) \tag{3}$$

N and M are the x and y-coordinate range of the searching window, while $R(x,y)$ indicates the pixels detected within the region mask generated from hand segmentation results. The probability centroid is turned into the center of mass of the current detected region as shown below.

$$C\prime(x_c,y_c) = (\frac{M\prime_{10}}{M\prime_{00}}, \frac{M\prime_{01}}{M\prime_{00}}) \tag{4}$$

After the approximated center is retrieved, the new seed point in the frame will be taken from this point and the region growing will start. In order to avoid distractions when trying to relocate the mean window, the seed point will be located within the hand region in the next frame. Therefore a successive tracking result can be achieved. By taking only the seed with the highest probability of lying in the hand region, the tracking procedure will recovery quicker from segmentation interferences (e.g., temporary occlusions).
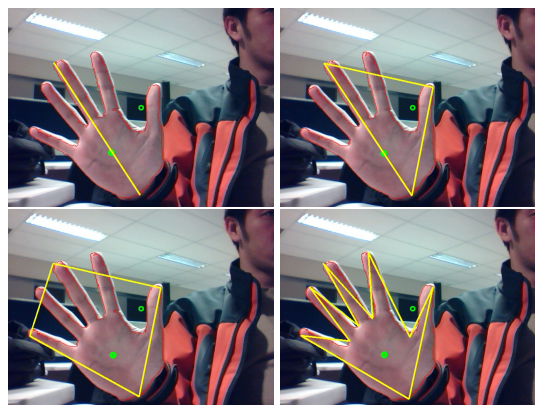

Figure 1: The red thin contour is the boundary of the segmented hand region. The yellow thick contour is the result of the Douglas-Peucker algorithm with the error threshold decreasing from top-left to bottom-right.

5.1 Hand Motion Estimation

Hand motion is estimated in four steps: contour extraction, contour approximation, finger tip tracking, and finger motion estimation. The process is simplified by using the anatomical constraints of the hand (Liu, 2010). We define the hand contour as the boundary of the segmented hand region which ideally coincides with the real silhouette of the hand. This is subtly different from the definition in (Bradski and Kaehler, 2008), where the contours were organised as a tree structure containing both exterior and hole contours. We tried to obtain a continuous representation of the contour using an active contour models (Kass et al., 1988), but the method failed to recover all the concavities in the hand shape. We therefore convert the hand contour to a polyline by using the Douglas-Peucker algorithm (Douglas and Peucker, 1973). Initially, the contour will be approximated as the line connecting two extremal points on the contour (top-left image in figure 1). In each step the furthest point on the segmented region boundary will be added until the distance between the contour and boundary is less than a pre-defined threshold. The procedure is illustrated in figure 1.

The finger positions can be estimated from the contour with reduced number of vertices by computing the convex hull and the convexity defects of the convex hull (Homma and Takenaka, 1985). Figure 2 illustrates the *depth* of a convexity defect, which we use to identify finger tip positions and the hand state. The finger tips are defined as the center positions between the start point of the ith defect and the end point of the $(i \pm 1)th$ defect. The $\pm$ depends on the order in which the contour vertices are traversed (clockwise or anticlockwise). The above described algorithm assumes that the patient exercise starts from an abduction posture to detect the finger. This could be generalised by using colour markers or template matching to improve identification of finger positions and hand gestures. Along with the finger tips, we also retrieve the position of the palm by using the points on the hand contour with the largest distance to the convex hull. Together with the finger tips, the system is able to compute the relative position of the finger tips with respect to the palm. The finger motion can be estimated based on distance changes between finger tips and between finger tips and the palm.

6 Results

We evaluated our hand segmentation method for different illumination conditions, backgrounds and
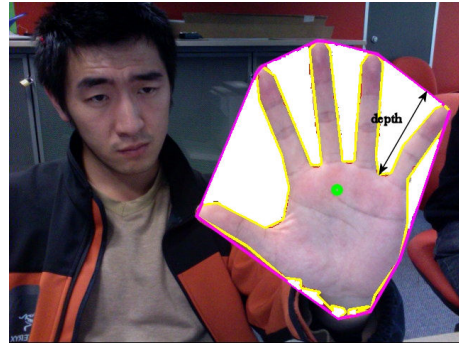

Figure 2: The convex hull (purple line), the convexity defects (white regions) and the depth of one convexity defect.

skin colours, using 9 university students as test subjects. We found that it is sufficiently stable and forms a suitable foundation for low-cost hand tracking applications. We have compared our perceptual colour space skin classifier with traditional skin classifiers and found that it is superior (Liu, 2010). The distortion of the perceptual colour space makes the segmentation sensitive to colour leaking problems in bright regions, e.g., highlights on the skin grow into bright background regions. Using multiple seed points and hue histogram back projection improved segmentation results considerably and reduced occurrences of the colour leaking problem. Another problem occurred for users with strongly visible skin wrinkles, which resulted in the loss of finger segments due to the edge feature criteria. The problem was alleviated by tuning the high-low ratio of the Canny edge detector.

To evaluate hand tracking, three subjects were asked to move their hand, rotate it, and move their fingers. As we expected, the finger positions are estimated well when the hand is posed at the initial position. Limitations became apparent when rotating the hand, but such rotations are not required for our hand exercises. When the fingers were bend like in a fist motion, then the finger tip could not be tracked, but the corresponding knuckle was usually correctly identified. False tracking results occurred occasionally when the hand and face overlapped.

We performed a small user study with three students testing the performance of our framework for tracking and correctly reconstructing hand configurations. The correct reconstruction is important for evaluating exercise performance and for patient collaboration in a virtual space. The subjects were asked to imitate animations of hand exercises as closely as possible. Figure 3 shows the results for a female Chinese student performing exercise #1. The two left most columns display the hand animation shown to

the user, the center column the web-cam images captured of the user performing the exercises, and the two right most columns show the hand animation obtained by tracking the hand motion on the web-cam images. The user easily matched the hand with the initial seed point. During the hand exercises, some further instructions about the orientation of the hand had to be given. Since the demonstration animation was very slow, the subject finished touching each finger earlier than the 3D hand model, i.e., the user predicted the required finger movement from the animation without following it precisely. This caused some mismatches of the demonstration and generated results.

The feedback from the three subjects involved shows a satisfactory recognition of the comparison between the animation they saw and the animation generated. None of the subjects involved in the tests complained about the intensity of the exercises and the length of them. No user reported discomfort or pain. The results indicate that typical hand exercises can be recognised and tracked with moderately slow finger and hand movements. The tracking algorithm can achieve satisfactory performance with respect to different illumination conditions and real world background using low-cost webcam input. The application can be initialised intuitively and exercises demonstration can be followed easily.

7 Conclusion

We have presented a framework for hand rehabilitation exercises and have demonstrated its potential with a user study. In order to realise this framework we developed a novel hand segmentation approach which is calibration-free and illumination invariant and hence well suited for unprepared home environments and real-world conditions. "Colour leaking" problems were reduced with a novel approach using probability maps and histogram back projection. Hand tracking was achieved using a novel variation of a seed point-based CAMSHIFT tracking method which utilises our segmentation results. The hand motion was estimated using a simple finger tip estimation based on the Douglas-Peucker algorithm and a convexity criterion. User testing showed that the algorithm provides stable tracking and that hand motion can usually be successfully reconstructed using the finger tip (knuckle) estimate.

We have constructed a hand model based on an existing modeling platform for human anatomy and animated it using a skeletal animation framework using predefined high-level atomic motions based on anatomical and physiological constraints. The

hand model proved suitable both to demonstrate hand tracking and to represent the hand motion analysis results. The reconstructed models had slight variations in finger positions, but the basic exercises were successfully represented. Initial user testing showed that the application is easy to use and that it would be useful for measuring the correctness and repetition of hand exercises.

REFERENCES

Boian, R., Sharma, A., Han, C., Merians, A., Burdea, G., Adamovich, S., Recce, M., Tremaine, M., and Poizner, H. (2002). Virtual reality-based post-stroke hand rehabilitation. *Studies in Health Technology and Informatics*, 85:64–70.

Bradski, D. G. R. and Kaehler, A. (2008). *Learning OpenCV, 1st edition*. O'Reilly Media, Inc.

Bradski, G. R. (1998). Real time face and object tracking as a component of a perceptual user interface. In *WACV '98: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*, page 214, Washington, DC, USA. IEEE Computer Society.

Chen, Q., Georganas, N., and Petriu, E. (2007). Real-time vision-based hand gesture recognition using haar-like features. In *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, pages 1–6.

Chong, H. Y., Gortler, S. J., and Zickler, T. (2008). A perception-based color space for illumination-invariant image processing. *ACM Trans. Graph.*, 27(3):1–7.

Douglas, D. H. and Peucker, T. K. (1973). Algorithm for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica*, 10:112–122.

Fischer, H. C., Stubblefield, K., Kline, T., Luo, X., Kenyon, R. V., and Kamper, D. G. (2007). Hand rehabilitation following stroke: A pilot study of assisted finger extension training in a virtual environment. *Topics in Stroke Rehabilitation*, Volume 14:1–12.

Handexercise.org (2010). Hand exercise: A resource for hand exercising. http://www.handexercise.org/.

Health Information Translations (2010). Active hand exercises. http://www.healthinfotranslations.org/pdfDocs/Active_Hand_Exercises.pdf.

Homma, K. and Takenaka, E.-I. (1985). An image processing method for feature extraction of space-occupying lesions. *J Nucl Med*, 26(12):1472–1477.

Jack, D., Boian, R., Merians, A. S., Tremaine, M., Burdea, G. C., Adamovich, S. V., Recce, M., and Poizner, H. (2001). Virtual reality-enhanced stroke rehabilitation. *IEEE transactions on neural systems and rehabilitation engineering*, 9(3):308–318.

Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recogn.*, 40(3):1106–1122.

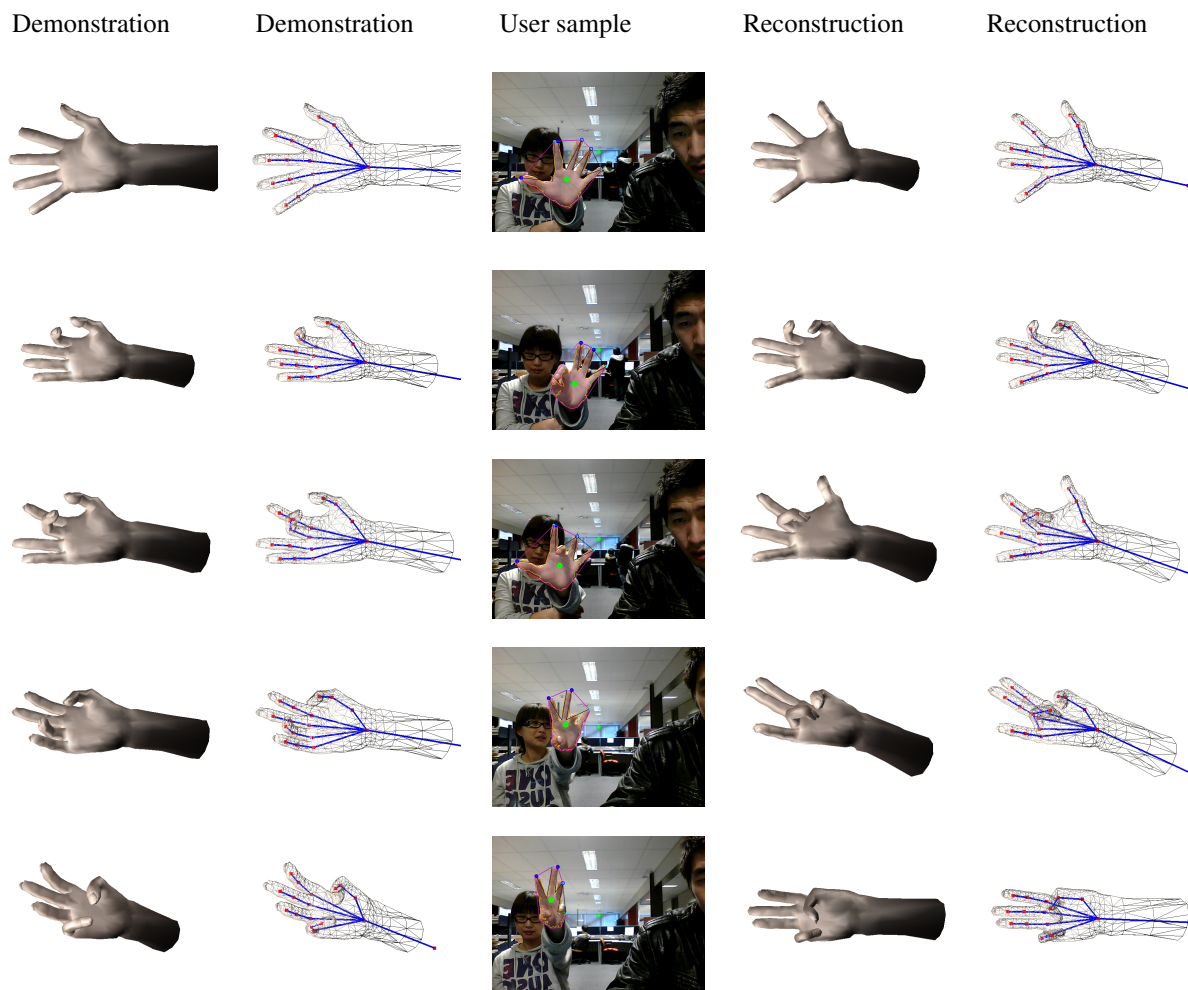| Demonstration	Demonstration	User sample	Reconstruction	Reconstruction

Figure 3: User study results of user #8 for the hand exercise #1, which was generated using pre-baked animation. The two left-most columns show screen shots of the hand model demonstrating the hand exercise. The center column shows the user performing the hand exercises. The two right most columns show the hand motions obtained by evaluating the web-cam images. Note that the participant uses the left hand and we currently mirror the webcam image in order to match it with the model of the right hand.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331.

Liu, R. (2010). A framework for webcam-based hand rehabilitation exercises. BSc Honours Dissertation, Graphics Group, Department of Computer Science, University of Auckland, New Zealand.

Mahmoudi, F. and Parviz, M. (2006). Visual hand tracking algorithms. *Geometric Modeling and Imaging–New Trends*, 0:228–232.

MHTeam (2010). Make human open source tool for making 3d characters. http://www.makehuman.org/.

Stenger, B., Mendona, P. R. S., and Cipolla, R. (2001). Model-based 3d tracking of an articulated hand. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:310.

Stenger, B., Thayananthan, A., Torr, P. H. S., and Cipolla, R. (2006). Model-based hand tracking using a hierarchical bayesian filter. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(9):1372–1384.

Vassili, V. V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. Graphicon*, pages 85–92.

Wessel, J. (2004). The effectiveness of hand exercises for persons with rheumatoid arthritis: A systematic review. *Journal of Hand Therapy*, 17(2):174–180.