Hand pose estimation and movement analysis for occupational therapy

Luciano Walenty Xavier Cejnog

December 14th, 2021

Advisor: Prof. Dr. Roberto Marcondes Cesar Jr.
Co-advisor: Prof. Dr. Teófilo Emídio de Campos
Collaborator: Profª. Drª. Valéria Meirelles Carril Elui
Project page: http://vision.ime.usp.br/~cejnog/

Overview

Overview

Context

Figure 1: *Example of hand with ulnar deviation from a patient on hand flexor tendon surgery recovery (on the right), in contrast with a normal hand of the same patient (on the left). Courtesy of Prof. Valeria Elui.*

Context

Figure 2: *Example of orthosis used on the hand, tailor-made devices made to distribute the force and leverage the effects of rheumatoid arthritis (provided courtesy by Prof. Valéria Elui).*

Context

(a) **(b)**

Figure 3: *Examples of goniometers used for hand range of motion measurements and providing objective feedback of the current state of the disease (provided courtesy by Prof. Valéria Elui).*

Research question

- **Can 3D hand pose estimation be used to measure accurately hand angles?**
- **This PhD research is part of the project "Hand tracking for occupational therapy" (proc. FAPESP 14/50769-1), that aims to study computer vision techniques capable of providing support to hand flexor tendon surgery recovery.**[1]

[1] The project is a collaboration with Professor Teófilo E. Campos (UnB), Professor Adrian Hilton (CVSSP), Professor Janko Calic (CVSSP / BBC), Professor Maria da Graça Campos Pimentel, from (ICMC/USP), and Professor Valeria Meirelles Carril Elui, (FMRP-USP).

Hypotheses

- **Computer vision methods can help diagnosis and objective feedback for patients with rheumatoid arthritis.**
- **Current state-of-the-art hand pose estimation methods can generalize to hands with rheumatoid arthritis.**

Goals

- **Main goal: to contribute to the development of a computer vision-based framework for automatic hand range of motion measurements.**
- **Specific goal: to develop methods to estimate hand joint angles in sequences of depth images, evaluating movement patterns of flexion/extension and abduction/adduction.**

Hand joints identification

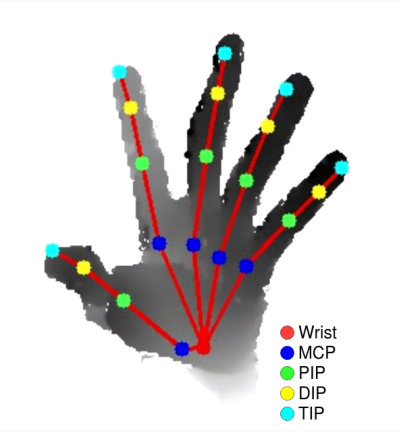


Figure 4: *Identification of hand joints. This figure was produced using the Intel Realsense® SR300 sensor, with real data from a hand with Rheumatoid Arthritis and an orthosis. The joints follow the hand model used in the HANDS17 dataset.*

Overview

Literature on hand pose estimation

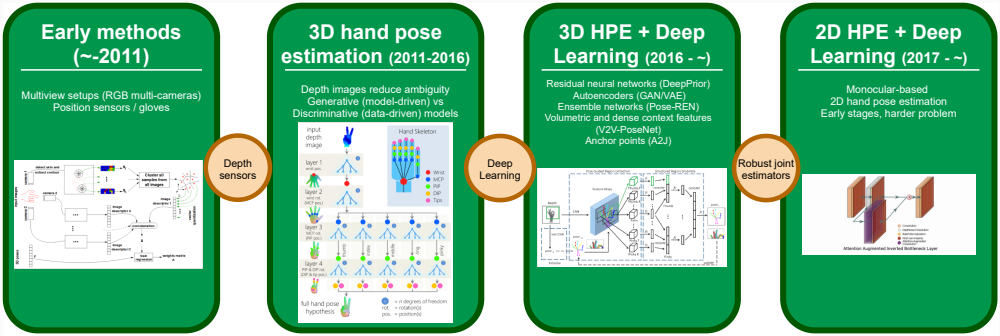
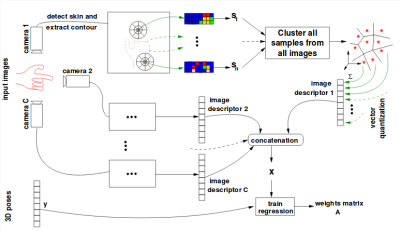


Figure 5: *Literature review.*

Early methods



(a) *Pipeline proposed by* Campos *(2006) for multiple view hand pose estimation (reproduced with permission from the author).*

(b) *CyberGlove II, reproduced from http://www.cyberglovesystems.com/cyberglove-ii/, accessed in 16/11/2017.*

Figure 6: *Early methods relied on multi-view inputs and sensor data to reduce the ambiguity generated from 2D images and the inherent hand structure.*

Depth sensors (2011 - 2016)

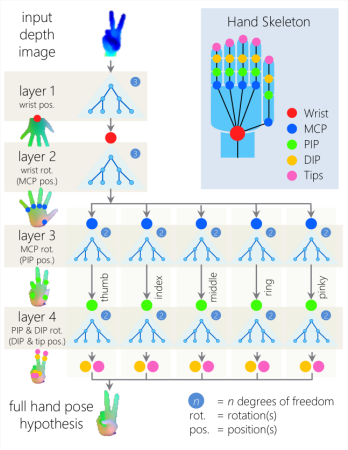


Figure 7: *Hierarchical hand pose detection pipeline, extracted from Tang et al. (2015). Copyright ©2015 IEEE.*

Deep learning on depth maps (2016 - present)

- **Many trends in deep learning have been explored by the literature:**
  - Residual neural networks: DeepPrior++ (Oberweger and Lepetit, 2017);
  - Autoencoders (Wan et al., 2017);
  - Ensemble networks: Pose-REN (Chen et al., 2019);
  - Volumetric and dense context features: V2V-PoseNet (Moon, Chang, et al., 2018; Wan et al., 2018);
  - Anchor points (Xiong et al., 2019).
- **Million-scale datasets (Yuan, Ye, et al., 2017);**

Deep learning with RGB images (2017 - present)

- **Robust 3D joint detection methods applied to monocular RGB image inputs.**
- **Application-wise important for accessibility (nature of input).**
- **Early stages of development: the absence of the depth dimension makes the problem much harder.**

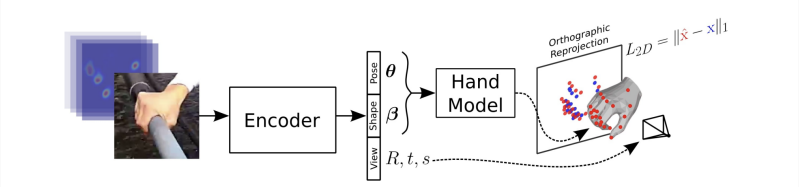


Figure 8: *Architecture used on Boukhayma et al. (2019).*

Discussion

- Depth-based methods have been considered a better choice for data acquisition from patients with rheumatoid arthritis.
- Method chosen: Pose-REN (Chen et al., 2019). This method is competitive in all datasets and can be executed "in the wild" in real-time.

Overview

Proposed pipeline

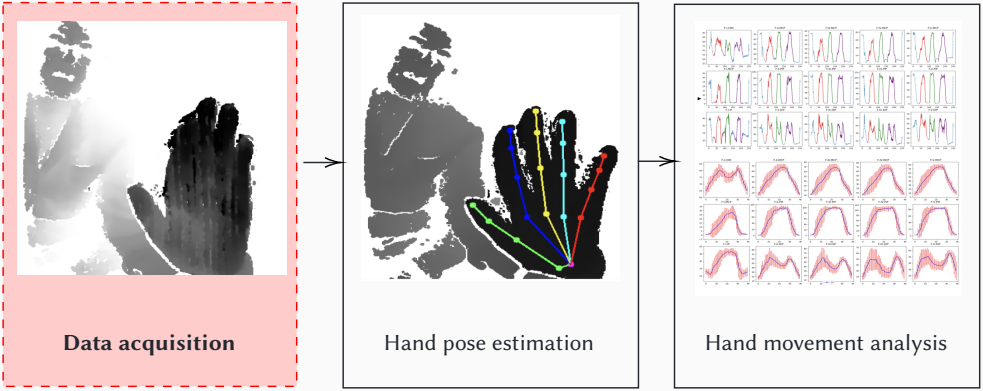
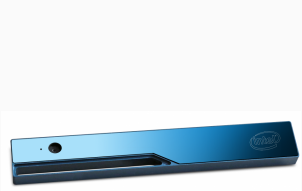


Figure 9: *Proposed pipeline.*

Dataset formation

- **Data acquisition was made in collaboration with Professors Valeria Elui and Daniela Goia at FMRP-USP.**
- **Main goals:**
  - To design a baseline setup for data acquisition;
  - To study different depth sensors;
  - To acquire data from patients in recovery of flexor tendon surgery.

Dataset formation



(a) *R200 (medium range)*

(b) *SR300 (short range) - chosen for acquisition*

(c) *Leap Motion (hand tracking for HCI)*

Figure 10: *Sensors used on the initial setup.*

Final Setup

Figure 11: *Setup used for data acquisition, with the Intel RealSense® SR300 acquiring depth image sequences in a frontal view.*

Dataset summary

Summary	
Patients with rheumatoid arthritis	8
Number of people in the control set	12
Patient Sequences	79
Control Sequences	108
Patient clips	310
Control clips	581
Total clips	891
Total number of frames	85755
Frames used on clips	60192
Percentage of frames used	70.2%
Size (GB)	482

Dataset discussion

- First dataset for hand pose estimation to contain data from Rheumatoid Arthritis patients.
- Challenging for current state-of-art pose estimation methods.
- Real-time hand pose estimation during capture.
- Main limitation: dataset does not contain hand joint annotations per frame.

Overview

Proposed pipeline

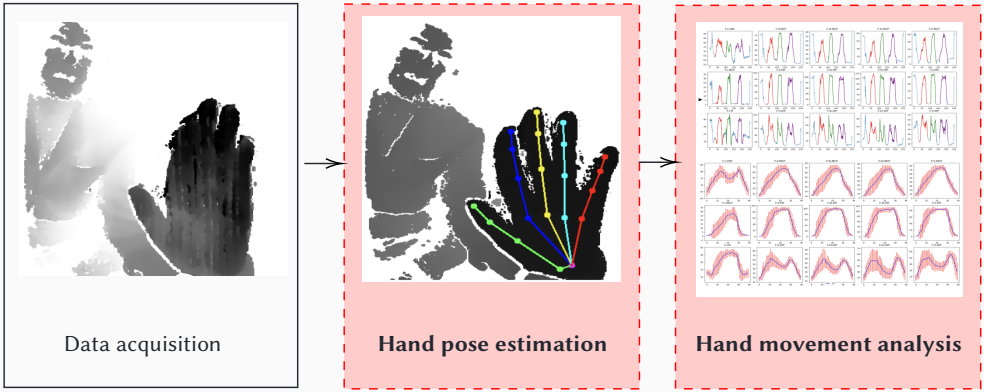


Figure 12: *Proposed pipeline.*

Hand pose estimation

- **Pose-REN trained with HANDS17 dataset.**

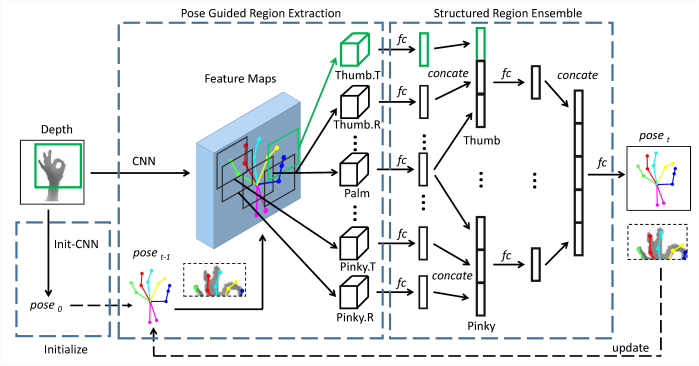
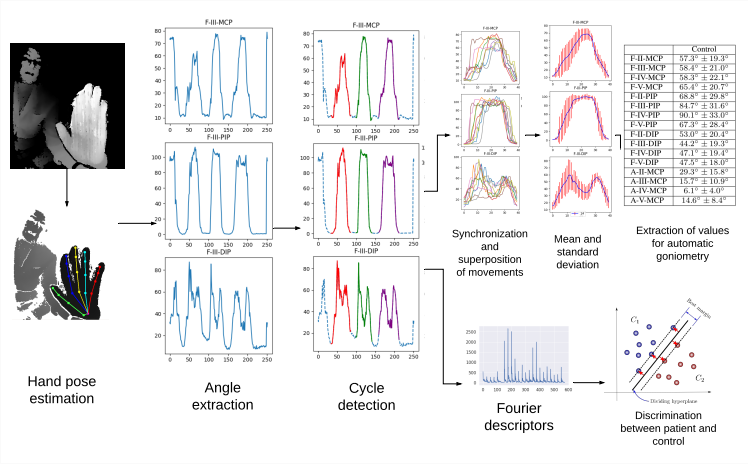


Figure 13: *Pipeline used on Pose-REN hand pose estimation method. Extracted from Chen et al. (2019) (Copyright license nr. 4918240801176).*

Hand analysis



	Control
F-II-MCP	$57.3° \pm 19.3°$
F-III-MCP	$58.4° \pm 21.0°$
F-IV-MCP	$58.3° \pm 22.1°$
F-V-MCP	$65.4° \pm 20.7°$
F-II-PIP	$68.8° \pm 29.8°$
F-III-PIP	$84.7° \pm 31.6°$
F-IV-PIP	$90.1° \pm 33.0°$
F-V-PIP	$67.3° \pm 28.4°$
F-II-DIP	$53.0° \pm 20.4°$
F-III-DIP	$44.2° \pm 19.3°$
F-IV-DIP	$47.1° \pm 19.4°$
F-V-DIP	$47.5° \pm 18.0°$
A-II-MCP	$29.3° \pm 15.8°$
A-III-MCP	$15.7° \pm 10.9°$
A-IV-MCP	$6.1° \pm 4.0°$
A-V-MCP	$14.6° \pm 8.4°$

Figure 14: *Hand movement analysis pipeline.*

Hand analysis

Computing angles

For the finger x, the flexion angles from the joints MCP, PIP and DIP are defined respectively as:

$$\widehat{\mathrm{FMCP}}_x = \arccos\left(\overrightarrow{\mathrm{MCP}_x - W} \cdot \overrightarrow{\mathrm{PIP}_x - \mathrm{MCP}_x}\right) \tag{1}$$

$$\widehat{\mathrm{FPIP}}_x = \arccos\left(\overrightarrow{\mathrm{PIP}_x - \mathrm{MCP}_x} \cdot \overrightarrow{\mathrm{DIP}_x - \mathrm{PIP}_x}\right) \tag{2}$$

$$\widehat{\mathrm{FDIP}}_x = \arccos\left(\overrightarrow{\mathrm{DIP}_x - \mathrm{PIP}_x} \cdot \overrightarrow{\mathrm{TIP}_x - \mathrm{DIP}_x}\right) \tag{3}$$

The abduction measurement is computed by the distance between two consecutive fingertips:

$$\mathrm{ATIP}_x = \left\|\mathrm{TIP}_{x-1} - \mathrm{TIP}_x\right\|_2 \tag{4}$$

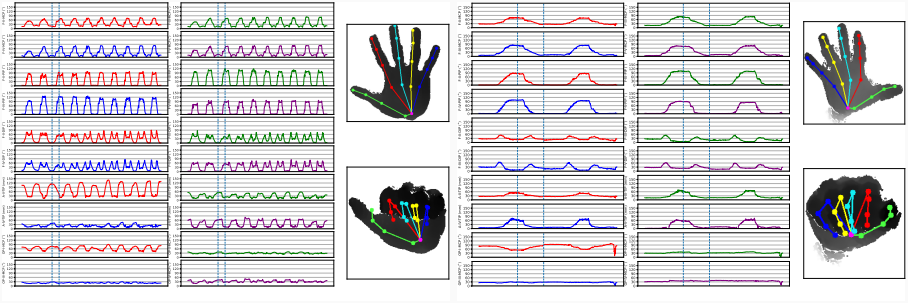
Overview

Experiment 1: Pipeline validation

- **Goal: Visual validation of the pipeline.**
- **Execution of the angle processing pipeline for each clip and patient sequence.**
- **Show results of each step, identifying and analysing patterns.**

Experiment 1: Pipeline validation

Visual validation of angle measurements:



(a) *Control* **(b)** *Patient*

Figure 15: *Angle evaluation of an individual in control group.*

Experiment 1: Pipeline validation

Visual validation of clip extraction:

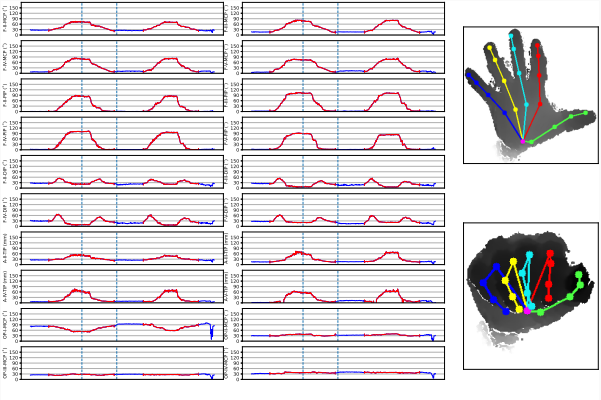


Figure 16: *Manual annotation of movement intervals in the angle sequence described in Figure 16. Extracted clips are marked in red.*

Experiment 1: Pipeline validation - Summarization and individual results

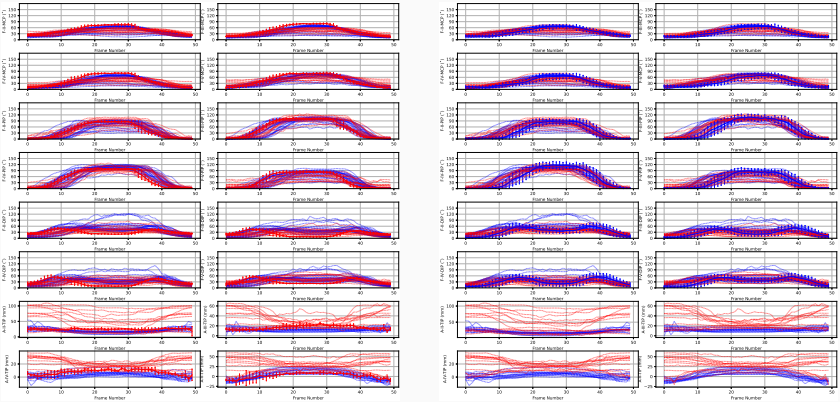


Figure 17: *Summarization in terms of mean and standard deviation of all trajectories extracted from clips from the same person: patient (left) and control (right).*

Experiment 1: Pipeline validation - Summarization and individual results

	Finger	2		3		4		5	
		min	max	min	max	min	max	min	max
P07 - L	MTC (°)	15.74	**44.05**	10.56	**47.77**	4.91	**47.51**	6.76	**58.84**
	IFP (°)	0.74	85.21	1.23	98.16	1.37	91.04	1.08	71.82
	IFD (°)	19.45	47.15	17.70	43.76	16.36	48.84	13.13	54.07
	abd (cm)	2.52		1.66		1.69		2.50	
P07 - R	MTC (°)	23.63	**76.20**	14.88	**80.43**	9.05	**81.23**	7.89	**80.91**
	IFP (°)	0.94	88.63	2.76	103.09	1.49	99.76	1.74	83.98
	IFD (°)	20.45	48.65	6.66	45.11	8.06	54.33	18.23	56.34
	abd (cm)	2.87		2.48		1.34		0.98	

Table 2: *Patient measurements extracted during the data acquisition session.*

Classification experiment

- **Pipeline application: to classify sequences into patients or control.**
- **We defined three types of classification experiments:**
  - 80-20% split
  - Leave-one-person-out
  - Leave-one-person-out with sample synthesis (LOO + SS).

Classification experiment - parameter design

- **Descriptors: Fourier descriptor vs Baseline descriptor(min and max for each angle).**
- **Classification algorithms (scikit-learn package):**
  - AdaBoost
  - Decision Tree
  - Gaussian Process
  - Linear SVM
  - Naive Bayes
  - Nearest Neighbors
  - Neural Net
  - QDA
  - Random Forest
  - RBF SVM

Results (Leave-One-Person-Out)

Experiment (%)	Control (%)	Patient (%)	General (%)
Fourier Linear SVM	**94.33** ± 10.53	**81.57** ± 31.34	**89.63** ± 21.67
Fourier Neural Net	92.89 ± 12.01	73.11 ± 36.33	85.60 ± 25.85
Baseline AdaBoost	89.54 ± 15.54	74.07 ± 30.09	83.84 ± 23.28
Baseline Linear SVM	89.08 ± 20.29	74.57 ± 33.44	83.74 ± 26.85
Baseline Neural Net	87.35 ± 22.97	73.91 ± 35.87	82.40 ± 29.14

Table 3: *Best performing classifiers on the leave-one-person-out experiment.*

Experiment 2: Classification

Leave-one-person-out with sample synthesis (LOO + SS)

- **Random gaussian noise to balance train-test + leave-one-person out.**
- **Different levels of noise and train set sizes have been tested.**

Experiment	Control (%)	Patient (%)	General (%)
LOO	94.66% ± 8.45%	83.00% ± 31.69%	90.37% ± 21.14%
LOO + SS, $\sigma = 1$	88.41% ± 18.90%	72.80% ± 36.20%	82.66% ± 27.66%
LOO + SS, $\sigma = 2$	88.63% ± 18.36%	73.51% ± 35.96%	83.06% ± 27.25%
LOO + SS, $\sigma = 4$	87.29% ± 18.09%	73.86% ± 34.85%	82.35% ± 26.38%

Table 4: *Accuracy comparison (in %) between the Linear SVM with sample synthesis using different values of σ (in mm) with the result obtained in the Leave-one-person-out experiment.*

Classification

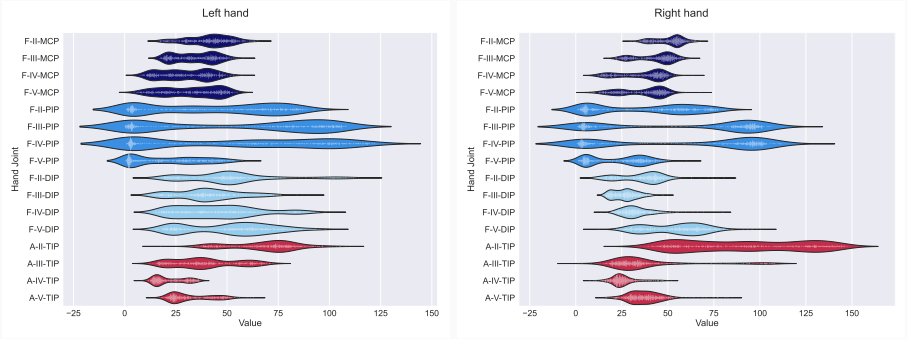
Main findings:

- **Without noise, we are able to reach a good accuracy score for classification between control and patients, even with scenarios of unseen shapes.**
- **With the presence of noise, the accuracy score is lower especially in patients. The training set size has little influence on the accuracy, and the use of Fourier descriptors does enhance the results.**

Experiment 3: Comparison with goniometer

- Goal: to compare patient measurements obtained by the sensor with reliable goniometer measurements.
- Ground-truth: Range of motion goniometer measurements for flexion and abduction of five patients.
- Maximum and minimum value for each flexion angle, and the maximum distance between tips for abduction.

Experiment 3: Comparison with goniometer



(a) *Left hand*

(b) *Right hand*

Figure 18: *Sensor angle observations from a patient.*

Experiment 3: Comparison with goniometer

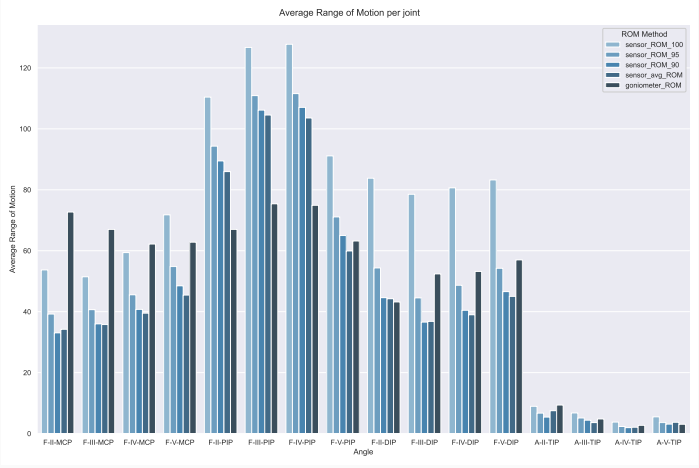


Figure 19: *Average range of motion per strategy, comparing with the goniometer.*

Experiment 3: Comparison with goniometer

- **Big error magnitudes.**
- **Error sources mapping: Pearson correlation between all measurements.**

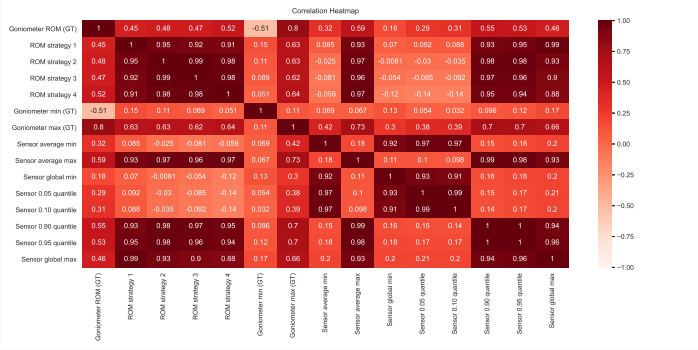


Figure 20: *Correlation heatmap between observations.*

Experiment 3: Comparison with goniometer

- Lower correlation for minimum values.
- This result shows that the sensor ROM measurements are still inaccurate for practical scenarios.

Overview

Conclusion

- This thesis sought to evaluate the viability of an automatic pipeline for patients with rheumatoid arthritis, using state-of-the-art hand pose estimation methods.
- The proposed method is able to accurately estimate skeleton angles and range of motion measurements from control and patients, even with the 3D hand pose estimation algorithm being trained in a completely different dataset of healthy hand movements.

Conclusion

The main findings of the experiments are:

- **The angles extracted by the hand analysis pipeline encode correctly flexion and abduction movements, characterizing visually each movement in terms of angle variation.**
- **A simple classifier and motion descriptor is able to distinguish between control and patient classes, even with unseen subjects.**
- **When compared to real goniometer range of motion measurements, the error magnitude is still high, indicating that the there is a lot of room for improvement in the application for real patient assessments.**

Acknowledgements

This work received funding from CAPES and FAPESP (#16/13791-4, #15/22308-2, #14/50769-1).

Thank you!

Overview

References i

- Adnane Boukhayma et al. (2019). "3D hand shape and pose from images in the wild". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10843–10852.

- E. Burnaev et al. (Dec. 2015). "Influence of resampling on accuracy of imbalanced classification". In: *Eighth International Conference on Machine Vision (ICMV 2015)*. Ed. by Antanas Verikas et al. Vol. 9875. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 987521. DOI: 10.1117/12.2228523. arXiv: 1707.03905 [stat.ML].

- T E de Campos (2006). "3D Visual Tracking of Articulated Objects and Hands". PhD thesis. Department of Engineering Science, University of Oxford, Trinity Term.

References ii

- L. W. X. Cejnog, T. E. de Campos, V. M. C. Elui, and R. M. Cesar Jr. (2019). "Hand range of motion evaluation for Rheumatoid Arthritis patients". In: *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)*, pp. 1–5.

- L. W. X. Cejnog, T. E. de Campos, V. M. C. Elui, and Roberto Marcondes Cesar Jr. (2021). "A framework for automatic hand range of motion evaluation of rheumatoid arthritis patients". In: *Informatics in Medicine Unlocked* 23, p. 100544. ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2021.100544. URL: https://www.sciencedirect.com/science/article/pii/S2352914821000344.

- Xinghao Chen et al. (2019). "Pose guided structured region ensemble network for cascaded hand pose estimation". In: *Neurocomputing*.

References iii

- Edward R Dougherty et al. (2002). “Inference from clustering with application to gene-expression microarrays”. In: *Journal of Computational Biology* 9(1), pp. 105–126.
- Ali Erol et al. (2007). “Vision-based hand pose estimation: A review”. In: *Computer Vision and Image Understanding* 108(1), pp. 52–73.
- Guillermo Garcia-Hernando et al. (2018). “First-person hand action benchmark with RGB-D videos and 3D hand pose annotations”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409–419.

References iv

- Daniela Nakandakari Goia et al. (Mar. 2017). “A new concept of orthosis for correcting fingers ulnar deviation”. en. In: *Research on Biomedical Engineering* 33, pp. 50–57. ISSN: 2446-4740. DOI: 10.1590/2446-4740.02516. URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2446-47402017000100050&nrm=iso.

- J. Gutiérrez-Martínez et al. (2014). “System to measure the range of motion of the joints of the human hand .”. In: *Revista de investigacion clinica; organo del Hospital de Enfermedades de la Nutricion* 66 Suppl 1, S122–30.

- Amélia Pasqual Marques (1997). *Manual de goniometria*. Editora Manole.

References v

- Clifton G. Meals et al. (2018). "Viability of Hand and Wrist Photogoniometry". In: *HAND* 13(3). PMID: 28391753, pp. 301–304. DOI: 10.1177/1558944717702471. eprint: https://doi.org/10.1177/1558944717702471. URL: https://doi.org/10.1177/1558944717702471.

- Gyeongsik Moon, Juyong Chang, et al. (2018). "V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

- Gyeongsik Moon, Shoou-I Yu, et al. (2020). "InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image". In: *European Conference on Computer Vision (ECCV).*

References vi

- Licia Maria Henrique da Mota et al. (2013). "Guidelines for the diagnosis of rheumatoid arthritis". In: *Revista Brasileira de Reumatologia (English Edition)* 53(2), pp. 141–157. ISSN: 2255-5021. DOI: https://doi.org/10.1016/S2255-5021(13)70019-1. URL: http://www.sciencedirect.com/science/article/pii/S2255502113700191.

- Mei-Ying Ng et al. (2021). "An enhanced self-attention and A2J approach for 3D hand pose estimation". In: *Multimedia Tools and Applications*, pp. 1–16.

- Markus Oberweger and Vincent Lepetit (2017). "Deepprior++: Improving fast and accurate 3D hand pose estimation". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 585–594.

References vii

- Adriana Garcia Orfale et al. (2005). "Translation into Brazilian Portuguese, cultural adaptation and evaluation of the reliability of the Disabilities of the Arm, Shoulder and Hand Questionnaire". In: *Brazilian Journal of Medical and Biological Research* 38(2), pp. 293–302.

- Javier Romero et al. (Nov. 2017). "Embodied Hands: Modeling and Capturing Hands and Bodies Together". In: *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*. URL: http://doi.acm.org/10.1145/3130800.3130883.

- Nicholas Santavas et al. (2020). *Attention! A Lightweight 2D Hand Pose Estimation Approach*. eprint: 2001.08047.

References viii

- David L Scott et al. (2010). "Rheumatoid arthritis". In: *The Lancet* 376(9746), pp. 1094–1108. DOI: 10.1016/s0140-6736(10)60826-4.
- Toby Sharp et al. (2015). "Accurate, robust, and flexible real-time hand tracking". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, pp. 3633–3642.
- Srinath Sridhar et al. (2013). "Interactive markerless articulated hand motion tracking using RGB and depth data". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2456–2463.

References ix

- Björn Stenger et al. (2006). "Model-based hand tracking using a hierarchical Bayesian filter". In: *IEEE transactions on pattern analysis and machine intelligence (PAMI)* 28(9), pp. 1372–1384.
- Xiao Sun et al. (2015). "Cascaded hand pose regression". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 824–832.
- Andrea Tagliasacchi et al. (2015). "Robust Articulated-ICP for Real-Time Hand Tracking". In: *Computer Graphics Forum*. Vol. 34. 5. Wiley Online Library, pp. 101–114.
- Siamak Bashardoust Tajali et al. (2016). "Reliability and validity of electro-goniometric range of motion measurements in patients with hand and wrist limitations". In: *The open orthopaedics journal* 10, p. 190.

References x

- Danhang Tang et al. (2015). "Opening the Black Box: Hierarchical Sampling Optimization for Estimating Human Hand Pose". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3325–3333.

- Anastasia Tkach et al. (2016). "Sphere-meshes for real-time hand modeling and tracking". In: *ACM Transactions on Graphics (TOG)* 35(6), p. 222.

- Chengde Wan et al. (2017). "Crossing nets: Dual generative models with a shared latent space for hand pose estimation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 7.

- Chengde Wan et al. (2018). "Dense 3D regression for hand pose estimation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5147–5156.

References xi

- Fu Xiong et al. (2019). "A2J: Anchor-to-Joint Regression Network for 3D Articulated Pose Estimation from a Single Depth Image". In: *Proceedings of the IEEE Conference on International Conference on Computer Vision (ICCV).*

- Shanxin Yuan, Guillermo Garcia-Hernando, et al. (2018a). "Depth-based 3D hand pose estimation: From current achievements to future goals". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2636–2645.

- Shanxin Yuan, Guillermo Garcia-Hernando, et al. (2018b). "Depth-based td3D hand pose estimation: From current achievements to future goals". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

References xii

- Shanxin Yuan, Qi Ye, et al. (2017). "Bighand2. 2m benchmark: Hand pose dataset and state of the art analysis". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4866–4874.

- Christian Zimmermann and Thomas Brox (2017). "Learning to estimate 3D hand pose from single RGB images". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4903–4911.

Hand Pose Estimation

- Hand pose estimation is an important task in the computer vision field.
- Challenges include the high dimensionality of the hand structure, self-occlusions and ambiguities on the model and the similarity between the fingers.
- Recent development of consumer-level 3D depth cameras and advances in computer vision and deep learning allow different fields of application to benefit from the latest advances on hand pose estimation.

Applications

- Several applications in areas such as human-computer interface, augmented reality, sign language recognition and robotics.
- A potential application field for hand pose estimation is hand surgery recovery and occupational therapy, in particular in the treatment of hand degenerative diseases.

Context

Rheumatoid Arthritis (RA)

- **Rheumatoid Arthritis is an autoimmune chronic disease that leads to joint deformities.**
- **Findings of population-based studies show RA affects 5 to 10% of adults in developed countries.**
- **Three times more frequent in women than men, and 50% of risk of developing RA is attributable to genetic factors (Scott et al., 2010).**

Diagnosis and treatment

- Recent advances in occupational therapy research indicate that the first 12 months with RA symptoms stand out as an acknowledged “window of therapeutic opportunity” (Mota et al., 2013).
- Identifying the disease in its early stages is fundamental in preventing its progression.

Tools used in diagnosis and treatment

- A common tool used in the treatment is the design of orthoses for injured hands.
- Orthoses are external devices tailor-made for patients and applied to any part of the body to stabilize it or immobilize it, prevent or correct deformities, protect against injury, maximize function and reduce the pain caused by deformity (Goia et al., 2017).
- For the hand case, the orthosis acts like a lever system distributing the force applied to the ulnar deviation.

(a) *Parts of an orthosis. Courtesy of Prof. Valéria Elui.*

(b) *Prototype of 3D printed orthosis made on CAD software (extracted from Goia et al. (2017)).*

Figure 21: *Examples of orthoses used on the hand, tailor-made devices made to distribute the force and leverage the effects of rheumatoid arthritis.*

Diagnosis and treatment

- **The evaluation of hand function is fundamental for the therapist to plan the treatment.**
- **A widely used metric for measuring the hand movement capacities is range of motion (ROM). (Marques, 1997)**
- **The range of motion is defined as the quantity of movement of an articulation.**
  - **Active range of motion** refers to movement without interference of external factors.
  - Passive range of motion refers to movement only by external factors.

Tools used in diagnosis and treatment

- Qualitative evaluation: *Disabilities of the Arm, Shoulder and Hand* (*DASH*) questionnaires (Orfale et al., 2005) are used to assess hand function during the recovery process.
- Quantitative evaluation: Hand-finger goniometer measures the active range of motion of each joint.

Treatment

- With a specific hand/finger goniometer the therapist can access objectively and reliably the range of motion measurements.
- Such devices are widely used due to their simplicity and low cost.
- The procedure, however, requires a trained therapist that follows the protocols, is time consuming and requires a careful setup and patient positioning.

Alternatives in occupational therapy

- **Electric sensors (Tajali et al., 2016; Gutiérrez-Martínez et al., 2014).**
- **Digital photogrammetry: determination of angles in hand images.**
- **Limitation: result is not immediately assessed (Meals et al., 2018).**
- **Future possibilities of using $3D$ scanning and video capture technology to the development of an automatic goniometer for the hand.**

Contributions

- **We propose an original end-to-end hand pose estimation and movement analysis approach for occupational therapy;**

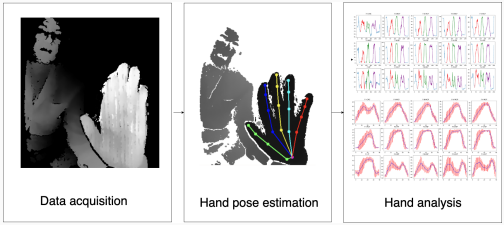


Figure 22: *Pipeline proposed.*

Contributions

- **We apply a state-of-the-art hand pose estimation method (Chen et al., 2019) for automatic range of motion evaluation of patients;**

Figure 23: *Estimative of joints obtained by the pose estimation algorithm on a RA patient.*

Contributions

- **We propose hand movement analysis tools based on the estimated angles and range-of-motion measurements from skeletons [2];**

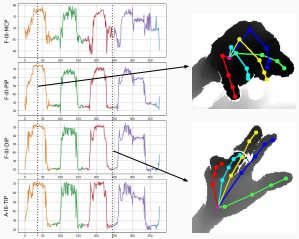


Figure 24: *Estimative of joints obtained by the pose estimation algorithm on a RA patient.*

[2] Results presented in Cejnog, de Campos, Elui, and R. M. Cesar Jr. (2019)

Contributions

- **We propose a dataset acquisition protocol and report the main decisions, difficulties of the process and the final acquisition protocol;**

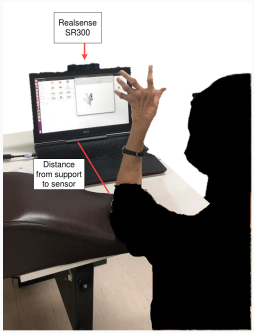


Figure 25: *Setup used for data acquisition, with the Intel RealSense® SR300*

Contributions

- **We present a new dataset of depth maps and hand tracking results obtained using from patients of Rheumatoid Arthritis being treated at the Hospital das Clínicas in the Faculdade de Medicina Ribeirão Preto / University of São Paulo;**
- **We perform experiments with the dataset, comparing with measurements obtained by goniometers** [3]

[3] Results presented in Cejnog, de Campos, Elui, and Roberto Marcondes Cesar Jr. (2021).

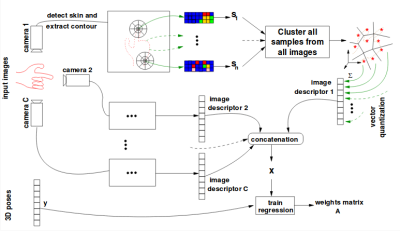
Literature on hand pose estimation

- **Literature revision objective: to find a suitable state-of-the-art method for application in the pipeline.**
- **In the thesis we describe the methods subdivided in a *historical cut*:**
  - Early methods (1997-2007)
  - Depth sensors (2011 - 2016)
  - Deep learning on depth maps (2016 to state-of-the-art)
  - Deep learning on RGB images (2017 to state-of-the-art).

Early methods

- First solutions: hand gesture recognition, sensors (gloves), multi-view setups in order to reduce ambiguity.
- The use of such devices limit the applicability of hand tracking to unnatural interactions.
- Tracking of single RGB sequences, applying general object tracking models: dependent of the number of poses on the training set.
- Markerless image-based hand pose estimation (Stenger et al., 2006): many solutions have been proposed with single-viewpoint and multi-viewpoint input devices (Erol et al., 2007; Campos, 2006).

Early methods



(a) *Pipeline proposed by Campos (2006) for multiple view hand pose estimation (reproduced with permission from the author).*

(b) *CyberGlove II, reproduced from http://www.cyberglovesystems.com/cyberglove-ii/, accessed in 16/11/2017.*

Figure 26: *Early methods relied on multi-view inputs and sensor data to reduce the ambiguity generated from 2D images and the inherent hand structure.*

Depth sensors (2011 - 2016)

- With the development of low-cost depth sensors, most methods started to use depth maps as input.
- The use of depth sensors reduce the data ambiguity without the necessity of configuring and calibrating a multiple view setup.
- Generative (Sharp et al., 2015; Tagliasacchi et al., 2015; Tkach et al., 2016) (or model-driven) vs discriminative (Sridhar et al., 2013; Sun et al., 2015; Tang et al., 2015) (or data-driven) methods.

Depth sensors (2011 - 2016)

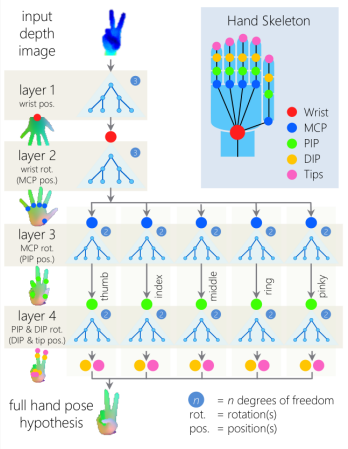


Figure 27: *Hierarchical hand pose detection pipeline, extracted from Tang et al. (2015). Copyright ©2015 IEEE.*

Deep learning on depth maps (2016 - present)

- **HANDS in the million 2017 challenge on 3D pose estimation (Yuan, Garcia-Hernando, et al., 2018a): results published in a survey (Yuan, Garcia-Hernando, et al., 2018b).**
  - Nature of input: Depth images vs **voxel grid**;
  - **Detection-based** (probability density maps) vs regression-based method;
  - **Hierarchical** vs holistic detection;
  - **Cascaded** vs one step training;
  - Generalization capacity of discriminative methods is still an issue.

Deep learning on depth maps (2016 - present)

- **The development of deep learning methods brought the necessity of larger datasets.**
- **Solutions with synthetic data generation: data augmentation;**
- **As a consequence, million-scale datasets have been made available to deal with the amount of data:**
  - BigHand2.2M (Yuan, Ye, et al., 2017);
  - First-Person Action dataset (Garcia-Hernando et al., 2018).

Deep learning with RGB images

- Many methods use synthetic hand poses for data augmentation in the model training.
- Boukhayma et al. (2019) incorporate the use of the MANO hand model (Romero et al., 2017).
- The attention method of Santavas et al. (2020) is currently the best perfirming method in most 2D hand pose estimation datasets.
- Larger datasets: Moon, Yu, et al. (2020) presented a dataset with 2.6 million images of 2D annotations on hand interactions.

Discussion

- **The popularization of depth sensors and the development of data-driven deep learning methods allowed new solutions to arise, with deep learning approaches reaching the best results to date in the standard datasets (ICVL, MSRA, NYU and HANDS17).**
- **The RGB variant of hand pose estimation is a much more difficult problem, in an earlier state of development with deep learning solutions.**

Discussion

- In the context of our work, the goal was to find a method suitable for hands with rheumatoid arthritis, as well as healthy hands.
- Preliminary experiments with the image-based method of Zimmermann and Brox (2017), whose source code was made available by the authors.
- Despite the ease of execution, we evaluated qualitatively that the results were inconsistent and very sensitive to skin tones and the presence of the orthosis.

Setup for Pose-REN

- The method required a background clutter step for preprocessing. The setup was built such that the hand is the nearest object from the camera.
- During acquisitions, we decided that the hand pose estimation should be executed in real-time during the capture, allowing the repositioning of the hand by the therapist.
- The availability of a pre-trained model in the HANDS17 dataset enhanced greatly the precision and robustness of the results.

Samples of the dataset for Pose-REN

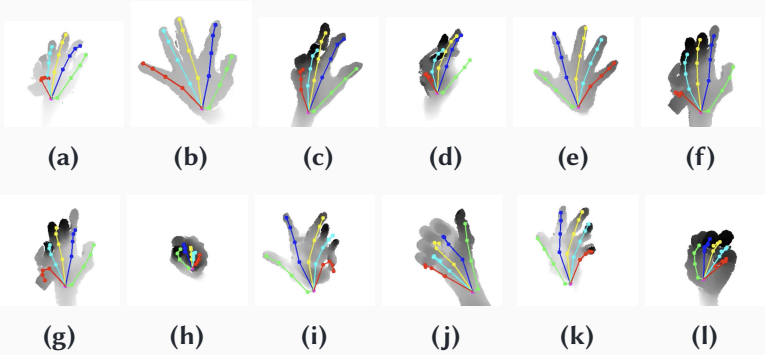


Figure 28: *Sample results obtained by applying Pose-REN model (Chen et al., 2019) trained on HANDS17 model with patients data, obtained in September 2019.*

Hand pose estimation

- Pose-REN method is based on the estimation of feature maps using Convolutional Neural Networks (CNNs).
- These feature maps are combined using an ensemble network, in order to generate a consistent hand pose.
- The skeleton used by HANDS17 dataset has 21 points of reference: the center of the wrist (W) and for each finger x the proximal interphalangeal (PIP_x), the distal interphalangeal joints (DIP_x) and the tip (TIP_x). The exception is the thumb, which is represented by the carpometacarpal joint (CMC) and a single interphalangeal joint (IP).

Hand pose estimation

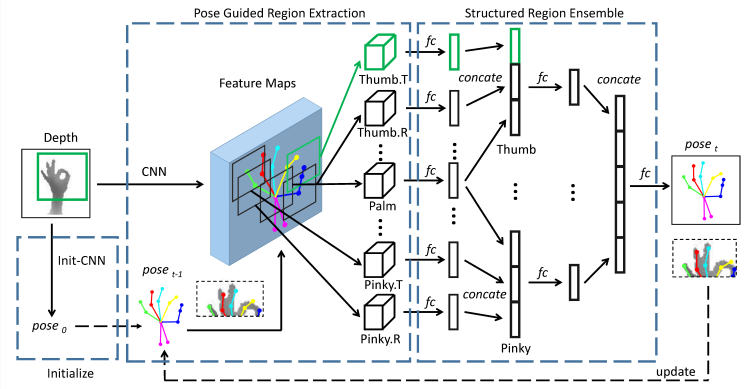


Figure 29: *Pipeline used on Pose-REN hand pose estimation method. Extracted from Chen et al. (2019) (Copyright license nr. 4918240801176)*

Hand analysis

- **Using the skeletons $\vec{S}(t)$ obtained by the hand pose estimation method, the analysis aims to obtain measurements of flexion/extension and adduction/abduction.**
- **Such measurements are computed for each frame of all sequences obtained in the acquisition.**
- **With one skeleton per frame, each recorded sequence yields a signal that is composed by time series, one for each estimated measurement. This time series is noisy and contain many movements of flexion or abduction per sequence.**
- **We will refer to each cycle of flexion or abduction inside a sequence as a *clip*.**

Validation

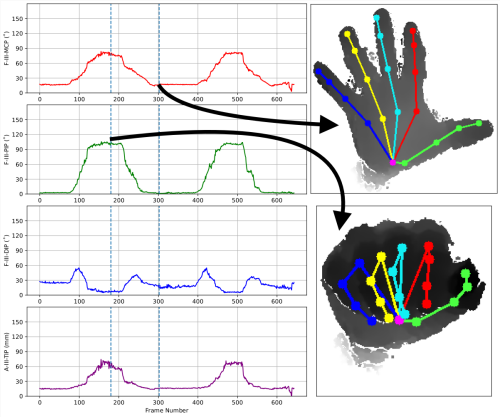


Figure 30: *Angle estimations, highlighting correspondences to poses obtained by the pose estimation algorithm on a RA patient.*

Cycle annotation

- Each clip is composed by multiple movements of flexion/abduction.
- The proposed approach aims to identify each movement inside the clips, identifying the average minimum and maximum values for each angle.
- For the analyses, the cycles were manually extracted from the sequences, using a visual tool to mark the frames from beginning and ending.

Validation

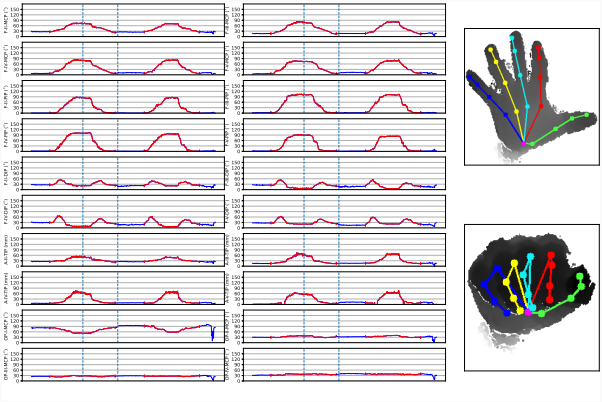


Figure 31: *Manual annotation of movement intervals of a patient acquisition. Extracted clips are marked in red.*

Synchronization and superposition of movements

- Synchronization is made by resampling the angle signals with a standard range.
- For this, we perform an interpolation in each angle signal, such that the length of each clip is set as 50 frames.
- After that, we are able to compute the average value and the standard deviation for one specific patient and considering all processed clips for both patients and control set.
- Individual measurements allow the construction of a feedback table.

Measurements example

	Finger	2		3		4		5	
		min	max	min	max	min	max	min	max
P1 - L	MTC (°)	0	80	-8	96	0	94	-6	92
	IFP (°)	-14	72	-18	88	-36	96	-48	96
	IFD (°)	0	40	0	50	0	28	-12	44
	abd (cm)	11.3		8		3.6		3.4	
P1 - R	MTC (°)	0	82	0	102	-8	70	-12	92
	IFP (°)	-24	72	-36	86	-32	76	-48	94
	IFD (°)	0	30	0	44	8	28	0	42
	abd (cm)	10.5		4		4.3		3.5	

Table 5: *Measurements extracted from one of the patients during the data acquisition session.*

Discrimination between patient and control

- Application: differentiation of sequences between patients and control sets.
- For this, we use the cycles obtained in previous steps and propose the use of Fourier descriptors in order to represent the multidimensional signal.
- This classification experiment is important to validate whether the current angle extraction pipeline is able to characterize the effect of Rheumatoid Arthritis in the flexion movement pattern.

Experiment 2: Classification

Split (80-20)

- **10 instances of classification with a random split of 80% training to 20% testing.**
- **Metric: accuracy mean and standard deviation.**

Results (Split)

Experiment (%)	Control (%)	Patient (%)	General (%)
Fourier Linear SVM	**96.31** ± 3.07	**91.97** ± 6.55	**94.14** ± 5.56
Baseline QDA	96.66 ± 3.81	89.11 ± 6.82	92.88 ± 6.69
Fourier Nearest Neighbors	97.08 ± 3.42	86.31 ± 9.39	91.69 ± 8.88
Baseline AdaBoost	94.99 ± 4.29	83.08 ± 12.56	89.04 ± 11.11
Baseline Neural Net	88.87 ± 8.93	88.65 ± 8.03	88.76 ± 8.49

Table 6: *Best performance classifiers on the Split experiment (in % accuracy).*

Experiment 2: Classification

Leave-one-person-out (LOO)

- **We choose one person and use all clips from that person as the test set.**
- **Training is done with all other sequences.**
- **Tests the generalization capacity for unseen subjects.**
- **We grouped the results in control and patient groups, showing mean and standard deviation accuracy for both.**

Analysis

- The best combination of classifier and descriptor in both experiments was the Linear SVM with the Fourier descriptor.
- Slight differences in the performance of classifiers.
- Interpretation: the proposed hand tracking and angle measurements successfully capture the differences between control and patient movements in a robust way.
- Therefore, the classification task itself does not critically depend neither on the features nor on the classifier, which is a good advantage of the proposed framework.

Analysis

- For control subjects, the accuracy reached in the majority of methods is high, surpassing 90% with low standard deviation in most cases.
- Higher error and variability on patient set (between 65% and 91% in the split experiment, and between 49% and 81% in the leave-one-person-out experiment).
- The Fourier descriptor was consistently better than the baseline descriptor.

Sample Synthesis

- The dataset is composed by 581 control clips and 310 patient clips.
- This poses the dataset as a slightly imbalanced dataset, which is usually biased towards the majority class (Burnaev et al., 2015).
- Common strategies to deal with this issue are undersampling of the majority class, oversampling of the minority class and data augmentation / sample synthesis techniques.
- For the third experiment we performed sample synthesis (SS) (Dougherty et al., 2002) to address the imbalance between the amount of samples from patients and control.

Sample Synthesis

Gaussian noise application

We generate synthetic data from the samples, enabling us not only to deal with data imbalance but also to evaluate the results of of our analysis method in the presence of hand pose estimation noise. For that, we applied Gaussian noise for each joint position in the skeleton: for a sequence

$$\vec{S}(t) = \{x_i(t), y_i(t), z_i(t)\}$$

for $i = 1, \cdots, N_j$ and $t = 1, \cdots, T$, we generate the augmented sequence

$$\vec{S'}(t) = \{x_i(t) + \mathcal{N}(0, \sigma), y_i(t) + \mathcal{N}(0, \sigma), z_i(t) + \mathcal{N}(0, \sigma)\}\ ,$$

where $\mathcal{N}(\mu, \sigma)$ represents a Gaussian function with μ mean and σ standard deviation, measured in millimeters.

Sample Synthesis

σ	ts=100		ts=400	
	Control	Patients	Control	Patients
Baseline, $\sigma = 1$	79.53% ± 24.86%	68.95% ± 34.86%	87.52% ± 18.67%	73.17% ± 31.37%
Baseline, $\sigma = 2$	81.07% ± 21.73%	73.78% ± 30.38%	84.88% ± 17.79%	72.56% ± 32.60%
Baseline, $\sigma = 4$	78.71% ± 19.64%	71.56% ± 27.10%	82.92% ± 20.25%	73.82% ± 28.68%
Fourier, $\sigma = 1$	87.16% ± 20.28%	71.76% ± 37.63%	89.67% ± 17.42%	73.86% ± 35.05%
Fourier, $\sigma = 2$	87.94% ± 18.80%	74.18% ± 35.93%	89.80% ± 16.73%	74.27% ± 35.86%
Fourier, $\sigma = 4$	84.42% ± 17.69%	75.79% ± 33.31%	84.82% ± 16.86%	73.37% ± 31.96%

Table 7: *Average and standard deviation SVM precision values for different train sizes and noise amounts.*

Experiment Remarks

- **The hand analysis pipeline yields coherent results for hand angles, for which the highest and lowest measurements are associated with open and closed hand patterns.**
- **The pipeline shows sufficient generalization capacity in terms of estimating hand poses in unseen scenarios.**

Experiment Remarks

- The use of simple descriptors and classifiers is enough to differentiate movement patterns from control and patient subjects.
- The high accuracy yielded from the leave-one-person-out experiment also indicates that the movement patterns are indeed separable as two distinct classes.
- Further exploration of the characteristics of such patterns can provide new findings about rheumatoid arthritis.

Experiment Remarks

- We compared ground-truth minimum and maximum values obtained from a goniometer with sensor measurements.
- This experiment shows that the range of motion intervals generated by the sensor and the goniometer have a low correlation, despite the efforts on evaluating different strategies.
- More studies and enhancements on hand pose estimation are needed in order to use the framework in practical range of motion acquisition scenarios.

Experiment Remarks

- **The lack of annotated data from patients sensibly limits the ROM measurement accuracy.**
- **With bigger and specific purpose datasets, the increase of the generalization capacity of hand pose estimation methods can help this pipeline to achieve more reliable results.**

Conclusion - Impact for computer vision i

- This thesis proposes a challenging application for hand pose estimation with a baseline solution.
- The pipeline built for estimation of hand angles can be used with different hand pose estimation methods and sensor configurations.
- The work of Ng et al. (2021) uses the flexion and abduction angles formulae to apply the self-attention hand pose estimation method in a setup with two sensors, computing the average angle value in both sensors in order to reach more robust results.

Conclusion - Impact for computer vision ii

- We believe that with the generalization capacity of current pose estimation methods makes possible the application in other knowledge areas, especially in assessments for medicine and occupational therapy.

Conclusion - Impact for occupational therapy i

- The thesis proposes a framework to analyse flexion and abduction angles as time signals.
- Compared to current movement analysis that uses maximum and minimum values, the analysis of a signal that encodes the complete movement pattern can help future characterization of movement patterns from patients in different stages of the disease.

Conclusion - Impact for occupational therapy ii

- **Objective feedback: the range of motion comparison experiment resulted in high error values, with low correlation between sensor data and goniometer data. This limits the use of our framework to provide objective feedback for the patients. However, new methods and new datasets can enhance this result.**
- **Applicability: the acquisition protocol is simple and requires a single depth sensor RealSense SR300.**
- **With further progress of the area, 2D hand pose estimation solutions can be feasible, making the setup much cheaper.**

Conclusion - Dataset i

- The created dataset is important in the sense of providing depth images of patients with rheumatoid arthritis in contrast with control images.
- The dataset has the limitation of not providing the ground-truth values for each frame, due to acquisition setup limitations.
- Experiments show that the dataset is able to provide valuable information in form of movement description for occupational therapists.

Conclusion - Dataset ii

- **The production of a dataset with ground-truth joint values for rheumatoid arthritis and other disabilities would improve the model, but was unfeasible in the current project.**

Future works

- **The pipeline is a baseline for angle estimative of patients, built such that new methods can be tested in the hand pose estimation step (*e.g. Ng et al. (2021)*).**
- **Construction of purpose-specific ground-truth datasets with patients with hand disabilities would enhance the generalization capacity.**
- **Enhancements on the generalization capacity of 2D hand pose estimation methods can provide cheaper setups.**

Future works

- Algorithm for automatic clip detection for flexion and abduction.
- Computation of angle formulae for other hand models, such as the MANO model.
- Associate the characteristics of rheumatoid arthritis with the observed behavior. Further analysis can also associate curve descriptors with the state of the disease in each patient.