Preliminary investigation on visual finger-counting with the icub robot cameras and hands

LUCAS, Alexandr <http://orcid.org/0000-0002-3165-3923>, RICOLFE-VIALA, C and DI NUOVO, Alessandro <http://orcid.org/0000-0003-2677-2650>

Preliminary investigation on visual finger-counting with the iCub robot cameras and hands

Alexandr Lucas[1(✉)], Carlos Ricolfe-Viala[2], Alessandro Di Nuovo[1]

[1] Sheffield Hallam University, Sheffield S1 1WB, United Kingdom
[2] Universitat Politecnica de Valencia, 46022 – Valencia – Spain
A.Lucas@shu.ac.uk

Abstract. This short paper describes an approach for collecting a dataset of hand's pictures and training a Deep Learning network that could enable the iCub robot to count on its fingers using solely its own cameras. Such a skill, mimicking children's habits, can support arithmetic learning in a baby robot, an important step in creating artificial intelligence for robots that could learn like children in the context of cognitive developmental robotics. Preliminary results show the approach is promising in terms of accuracy.

Keywords. Developmental robotics, finger-counting, Faster R-CNN, iCub.

1 Introduction

This paper presents preliminary results of an ongoing investigation to identify fingers using computer vision with a humanoid robotics platform in the context of cognitive developmental robotics. Even if it is one of the first ability that children learn, visually counting the fingers is not an easy task in computer vision. In one of the few previous works in robot hand detection using Machine Learning methods [7], Cartesian Genetic Programming for Image Processing algorithm was trained to detect the iCub hands and fingertips. Although the results were announced as competitive, no comparison was given due to the lack of suitable datasets at that time. A combined approach of estimating the iCub hand pose using proprioception, stereo vision and 3D model of the robot was presented in [10], with a comparison of silhouette segmentation vs edge extraction techniques. However, to the best of our knowledge, the problem of reliably identifying the number of fingers being shown in the finger-counting scenario has not been addressed yet. The importance of acquiring the skill of visually counting fingers is crucial for supporting cognitive developmental robotics models of number cognition and basic arithmetic [6, 8], which are envisioned by the EPSRC project NUMBERS. Indeed, fingers have a significant influence on the development of our counting system, with some suggesting that we likely use a base 10 system due to the number of fingers on our hands [3]. Moreover, recent studies suggest that fingers, being natural tools, play a fundamental role at many stages of maths learning, from developing a sense of numbers to acquiring proficiency in the basic arithmetic processing [4, 5, 9]. Application of these concepts to robotics formed what is now known as Cognitive Developmental Robotics (CDR), defined as the “interdisciplinary approach to the autonomous design of behavioural and cognitive capabilities in artificial agents (robots) that takes direct inspiration from the developmental principles and mechanisms observed in natural cognitive systems (humans)” [1, 2].

2 Materials and methods

2.1 Setup

As our research platform, we used the iCub humanoid robot provided by Sheffield Robotics (Fig.1). iCub is one of the most advanced child-like robotics platforms available, featuring 41 actuated DOF and tendon-driven joints. Particularly important for this work are fully functional five-fingered hands, closely emulating the human hand. The setup comprised a legless iCub fixed on a platform, with a greenscreen board positioned in the background, in such a way that the board fully covered the field of view of iCub’s eye cameras looking directly forward. Gaze direction and head pose were fixed to minimise background variance. For image segmentation algorithm to produce a hand silhouette of good quality, illumination control, and uniform background are necessary. Green was chosen as the background colour because it is the most distinct from the colours of the robot's hand and its covers.

Fig. 1. Sheffield Robotics’ iCub

Fig. 2. iCub counting against a green background (top row). These images were used to produce the DNN model dataset (bottom row)

iCub was programmed to count from one to five using fingers on the left hand. From a range of different finger-counting styles representing numbers from one to five, we chose the American Sign Language, as being most convenient given specific hardware implementation of iCub’s hands (shared actuator for the ring and little fingers).

2.2 Image data collection

In the initial data collection, for each number shown the hand cycled through slightly different orientations using the wrist pro/sup, pitch, and yaw joints, sequentially, with all other hand joints being fixed. The angle (min, max, step) values of the wrist pro/sup joint were set to be (-55, 55, 15), for the wrist pitch joint to be (-50, 25, 15) and for the wrist yaw joint to be (-20, 25, 10). For each hand pose, the robot would take a picture using both its left and right camera with 640x480 pixels resolution. This resulted in 200 pictures for each number of fingers shown from 1 to 5, totalling 1000 images for each of the two cameras. Settings for both cameras were adjusted manually at the beginning to find the optimum values of gain, shutter speed, exposure, brightness, and white balance. The top row of Fig. 2 contains a sample image from each class. However, some poses where found to be looking ambiguous, with fingers occluding each other and not clearly visible. This effect increased with the number of fingers shown. Removing ambiguous images from the database reduced its size to 619 images for the left camera and 562 for the right.

2.3 Extracting the hand silhouettes

For extraction of regions of interest containing the hand with fingers silhouettes from each image, we utilised TensorFlow Object Detection API library, which in turn makes use of OpenCV routines. The labelling process was as follows:

1. The images are converted to HSV space to highlight hand pixels. Detection of contours in RGB space was found to be ineffective, due to noise and a bright light source directly over the robot's head causing glare.
2. One image with just the green background without the hand is used to detect the threshold values for pixel segmentation.
3. Noise is removed by applying a series of dilations and erosions.
4. Detection of contours of the biggest object (hand), which is then drawn separately with a polygonal approximation (see Fig. 2, bottom row).
5. ROI coordinates for each image are saved in a .xml file.

2.4 A deep neural network model for fingers detection

For validating the collected image database, we trained a DNN model to identify the number of fingers being shown. Due to a large amount of time it takes to train and tune a model from scratch, we chose to utilise the Faster R-CNN Inception Resnet image classification model pre-trained with COCO dataset from the TensorFlow Model Zoo[1] as a base model for further re-training. The out-of-the-box model parameters were used with no changes during the training. With the primary goal to test the feasibility of the approach, Faster R-CNN model was chosen as being more accurate, albeit slower than SSD models, for example. Training and testing datasets (1000 and 200 images, correspondingly) were compiled by composing random office backgrounds with the hand pixels extracted from the original images using corresponding silhouettes. 100K training steps were performed, which amounts to 100 epochs given batch size of 1 image. A checkpoint was saved every hour (approx. 3K steps). Since the training was done for fixed number of steps and no parameter-tuning was performed, the testing set also acted as the validation set.

---

[1] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md.

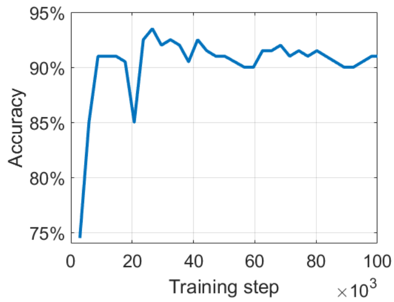

Fig. 3. The progress of classification accuracy with the test dataset

Table 1. Confusion matrix for the model with the best accuracy on the test dataset. Rows — actual values, columns — predicted values.

%	ONE	TWO	THREE	FOUR	FIVE
ONE	100.0	0.0	0.0	0.0	0.0
TWO	0.0	100.0	0.0	0.0	0.0
THREE	0.0	5.0	95.0	0.0	0.0
FOUR	2.5	0.0	0.0	75.0	22.5
FIVE	0.0	0.0	0.0	2.5	97.5
Sensitivity	100.0	100.0	95.0	75.0	97.5
Precision	97.6	95.2	100.0	96.8	81.2

3 Results

The model was evaluated at different checkpoints on a validation set of 200 images (40 for each class). Classification accuracy peaks at about 26K steps and then drops because of overfitting (Fig. 3). Confusion matrix for this checkpoint (step 26581) is given in Table 1, with most errors due to 'fours' classified as 'fives'.

Further work will focus on expanding the hand image dataset, improving the silhouette extraction and preventing miss-classification for images showing occlusion and finger overlap (such as 4-as-5).

4 Acknowledgments

This work has been supported by the EPSRC through the grant no. EP/P030033/1 (NUMBERS). Authors are grateful to the NVIDIA Corporation for donating GeForce GTX TITAN X that has been used to accelerate the computation.

5 References

1. Asada, M. et al.: Cognitive developmental robotics as a new paradigm for the design of humanoid robots. Rob. Auton. Syst. 37, 2–3, 185–193 (2001).
2. Cangelosi, A., Schlesinger, M.: Developmental robotics: from babies to robots. MIT Press (2015).
3. Dantzig, T.: Number–The Language of Science. Free Press (1954).
4. Fischer, M.H. et al.: Finger Counting and Numerical Cognition. Front. Psychol. 3, APR, 108 (2012).
5. Goldin-Meadow, S. et al.: Gesture's role in learning arithmetic. In: Edwards, L.D. et al. (eds.) Emerging Perspectives on Gesture and Embodiment in Mathematics. pp. 51–72 Information Age Publishing (2014).
6. De La Cruz, V.M. et al.: Making fingers and words count in a cognitive robot. Front. Behav. Neurosci. 8, 0–12 (2014).
7. Leitner, J. et al.: Humanoid learns to detect its own hands. 2013 IEEE Congr. Evol. Comput. CEC 2013. 1411–1418 (2013).
8. Di Nuovo, A., Jay, T.: The development of numerical cognition in children and artificial systems: a review of the current knowledge and proposals for multi-disciplinary research. IET Cogn. Comput. Syst. (2019).
9. Soylu, F. et al.: You Can Count on Your Fingers: The Role of Fingers in Early Mathematical Development. J. Numer. Cogn. 4, 1, 2363–8761 (2018).
10. Vicente, P. et al.: Robotic Hand Pose Estimation Based on Stereo Vision and GPU-enabled Internal Graphical Simulation. J. Intell. Robot. Syst. 83, 3–4, 339–358 (2016).