




Article

Head Pose Detection for a Wearable Parrot-Inspired Robot Based on Deep Learning

Jaishankar Bharatharaj ^{1,*} , Loulin Huang ¹, Rajesh Elara Mohan ², Thejus Pathmakumar ² , Chris Krägeloh ³  and Ahmed Al-Jumaily ¹

¹ Institute of Biomedical Technologies, Auckland University of Technology, Auckland 1010, New Zealand; loulin.huang@aut.ac.nz (L.H.); ahmed.aljumaily@aut.ac.nz (A.A.-J.)

² Engineering Product Development Pillar, Singapore University of Technology and Design, Singapore 487372, Singapore; rajeshelara@sutd.edu.sg (R.E.M.); thejus08@gmail.com (T.P.)

³ School of Public Health and Psychosocial Studies, Auckland University of Technology, Auckland 1010, New Zealand; chris.krageloh@aut.ac.nz

* Correspondence: jaishankar.bharatharaj@aut.ac.nz; Tel.: +64-09-921-9999 (ext. 8932)

Received: 8 May 2018; Accepted: 1 July 2018; Published: 3 July 2018



Abstract: Extensive research has been conducted in human head pose detection systems and several applications have been identified to deploy such systems. Deep learning based head pose detection is one such method which has been studied for several decades and reports high success rates during implementation. Across several pet robots designed and developed for various needs, there is a complete absence of wearable pet robots and head pose detection models in wearable pet robots. Designing a wearable pet robot capable of head pose detection can provide more opportunities for research and development of such systems. In this paper, we present a novel head pose detection system for a wearable parrot-inspired pet robot using images taken from the wearer's shoulder. This is the first time head pose detection has been studied in wearable robots and using images from a side angle. In this study, we used AlexNet convolutional neural network architecture trained on the images from the database for the head pose detection system. The system was tested with 250 images and resulted in an accuracy of 94.4% across five head poses, namely left, left intermediate, straight, right, and right intermediate.

Keywords: head pose detection; bio inspired robot; parrot-inspired robot; wearable pet robot; deep learning

1. Introduction

The field of robotics has found numerous applications in the recent years. Earlier, the predominant use of robots was mainly in performing tedious and repetitive tasks, such as manufacturing and transporting. But, a new generation of robots with more intelligence has been shown to benefit several other industries, including service, medical, and entertainment. Particularly, robots capable of interacting with humans with natural behavior has begun to emerge extensively due to their closeness with humans.

Several methods have been explored to establish human-machine interaction (HMI) in the literature. Gesture recognition pertains to recognition of human expressions through their hands, head, and/or body movements. In the recent years, gesture recognition has been one of the most focused research areas, where new methods and applications of interacting with medical, service, and entertainment devices have been studied. For example, Greatzel et al. [1] developed a system to replace standard computer mouse operation with hand gestures using a computer vision algorithm. The system is designed to establish non-contact human-computer interaction (HCI), which helps

surgeons to use computers during surgery. With this method, a surgeon can use the non-contact mouse by placing his/her hand stationary in the workspace for a moment and moving it to the desired position. The system uses a Kalman filter to estimate hand velocity and predict the next hand position during its movement. In another study, an intelligent wheelchair which can be controlled by the user's hand gestures was presented by Yoshinori et al. [2]. This wheelchair can detect the owner's face and follow their hand gestures to perform actions. The system allows the user to command the robotic wheelchair through hand gestures to approach or move away from them when they are not riding it. The Eigenface-based face recognition method is used in this wheelchair to detect its owner and receive hand gesture commands through the spotting recognition method. Experiments were conducted to confirm the working and usefulness of the system in several settings, and success was reported in detecting the owner's face and following their commands through hand gestures.

As with the medical field, service robotics is another field where gestures are used to control and command robots to achieve various tasks. For example, David et al. [3] proposed an approach to using showing and pointing hand gestures with a domestic service robot using a time-of-flight camera. The study evaluated the showing and pointing gestures through a set of different experiments and reported higher accuracy than existing stereo-based systems. The authors tested the system real-time with their domestic service robot which competed in a robotics competition. During this event, the robot successfully identified an object through the user's pointing gestures, and approached and delivered the object to the user. In another setting, the robot recognized the object shown by the user and delivered a similar object located in its environment. In a laboratory based testing, 16 participants showed 24 different objects through pointing gestures to the robot 192 times, and the robot correctly identified the gestures 187 times with an accuracy of 97%. Yin and Xie [4] proposed a set of hand gesture-based commands to a humanoid service robot, HARO-1, for controlling its arm movements and turn-taking. Six hand gestures for controlling the robot's six-axis arms and two hand gestures for turning the robot clockwise and anti-clockwise were used in this method. Neural network was used to segment captured hand images and the hand posture was recognized using a topological feature extraction method. The system indicated success when integrated with the service robot, HARO-1, and demonstrated the effectiveness and robustness of the approach. In another study, Lee [5] presented a new method for recognizing the gesture of a whole human body, which is used to operate the service robot. Unlike other gesture-controlled robots which use a part of a human body, this method spots and recognizes whole human body key gestures to command the robot. The system was integrated into T-Rot, a personal service robot, to evaluate real-time. With an accuracy of 97.4%, the system reported success in recognizing whole body gestures, such as 'sitting on the floor' and 'getting down on the floor'.

Robot entertainment, a field of the entertainment industry, uses a variety of semi-autonomous and autonomous robots to entertain the users through gestures. Establishing natural interaction between humans and robots is very essential for these classes of robots due to their closeness with humans during their deployment. Vision-, voice-, and gesture-based interactions are few of the most desired interaction methods explored in entertainment robots. Particularly, gesture-based entertainment robots reported success in establishing human-robot interaction (HRI) effectively compared to other methods explored. As an illustration, Hasanuzzaman et al. [6] presented a vision-based gesture recognition system using skin color segmentation and a pattern matching technique to create HRI between the entertainment robot—AIBO—and the user. The study trained the robot to recognize eight different hand gestures from the user and perform corresponding actions, such as 'stand', 'walk forward', and 'sit'. Another study performed by Hasanuzzaman et al. [7] described a gesture-based, human-centric HRI system using Software Platform for the Agent and Knowledge management (SPAK) platform. The system uses face and gesture recognition to identify the user and corresponding actions for the gesture of respective users. With this human-centric HRI system, the robot can perform different actions for the same gesture, according to the user recognized

through the face recognition system. Several other robots, such as ROBITA, Robonaut, and Leonardo use gestures to interact with humans [8].

Head pose detection is one of the gesture recognition techniques used in various applications. It has been used in various fields such as robotics [9], computer engineering [10,11], physical science and health industry [12], natural sciences [13], and industrial academic areas [14–16]. As an illustration, Sileye and Jean-Marc [17] deployed head pose detection using the Hidden Markov Model to recognize the visual focus of attention of participants in meetings. Eric and Mohan [18] presented a vision-based head pose detection and tracking method for monitoring driver awareness. The authors propose this method to monitor driver alertness and their head pose orientation while driving. Javier and Patricio [19] proposed head pose detection between robots to decide the next action by an observing robot. In this study, the authors used two robots, one acting as a performer and the other as an observer. The performer robot changes its head pose which is processed by the observer robot to perform actions. The head pose detection and recognition system has found wider application areas such as face recognition, action recognition, gait recognition, head recognition, and hand recognition systems [20–23]. In such systems, several sensors like binary, digital, and depth cameras are used to train and detect postures [24–30]. In such systems, several machine learning feature extraction algorithms and classification methods are implemented for detection and recognition of gestures. For instance, Samina et al. [31] analyzed several feature selection and extraction methods, and presented their effectiveness in achieving high performance of learning algorithms. A real-time tracking system for human pose recognition was proposed by Jalal et al. [32] using ridge body part features, in which a support vector machine (SVM) was used to recognize different poses. In another study, a novel subspace learning algorithm, called discriminant simplex analysis (DSA), was developed by Fu et al. [33] in which the intraclass compactness and interclass separability were measured by distances.

Even though head pose detection has been studied in robotics for many years, there are applications which can be further studied to effectively use a head pose detection method. Pet robots are one of the areas where it can be very useful in establishing human–robot interactions, but was given primitive focus in the literature. The applications of pet robots are manifold ranging from medical, service, and the entertainment industry. For example, a pet robot that we have developed has been used to reduce stress levels of patients [34], improve learning abilities of children [35], and entertain participants [36]. With such multi-industry applicability, pet robots with head pose detection can further improve closeness with humans and create effective human–robot interaction models. Even though several pet robots have been designed and developed for various needs in the literature, there is a complete absence of wearable pet robots and human–robot interaction models in wearable pet robots. With known benefits of wearable pets [37,38], designing a wearable pet robot can provide extensive research and application opportunities in several sectors. In this paper, we present the design and development of a wearable parrot-inspired pet robot, KiliRo, and its human–robot interaction model using vision-based head pose detection. The novelty of this paper is threefold: First, we introduce a new design and development of a wearable parrot-inspired pet robot. Second, we provide the design of human-robot interaction model for wearable pet robots using a vision-based head pose detection method. Third, we quantitatively demonstrate the success of this system through head pose images captured from the robot wearers shoulder in five different orientations.

The remainder of this paper is organized as follows: After the presenting the system architecture of our KiliRo robot in Section 2, we outline our system consisting of methods for detecting and perceiving head pose orientation of the person wearing the robot (Section 3). In Section 4, we present the experiments involving 1380 images of head poses in five different orientations to validate our approach. Lastly, in Section 5, we conclude this study and discuss the future works.

2. Robot Architecture

The main scope of this research study is to design and develop a wearable pet robot that can mimic the head pose orientation of the wearer. In terms of morphology, the KiliRo robot can be defined as a two-legged wearable robot, having a physical appearance that resembles a parrot.

We considered a set of design constraints in deciding the dimensions of the robot during the concept generation process:

- Height <250 mm;
- Weight <250 g;
- Head rotation 180 degrees;
- Operate between 10 °C and 45 °C

After a series of brainstorming sessions on concept generation and selection sessions, we developed a wearable pet robot KiliRo centered on achieving 180 degrees of rotating head design. The curvature of the robot's leg design was optimized to create a wearable robot design. The dimensions and weight of the robot played a vital role in the wearable robot design, as the wearers wore them in most cases. The dimensions of KiliRo-W and the selection of commercial devices, such as servo motors, electronic boards, etc., were opted to fit the robot design constraint on size and weight. The robot has three parts: head, body, and wings. The neck part connects the head and body. A static tail is attached at the top of the head for aesthetic appeal. The feet were designed through inspiration from parrots and were modified to adapt to the wearable design, which can fix to the body part. The robot parts were designed to be hollow to minimize the weight and optimize the three-dimensional printed materials. The specifications of the mechanical properties of the wearable parrot robot are listed in Table 1.

Table 1. Mechanical properties of wearable KiliRo robot.

Robot Full Body Material	PLA (Poly Lactic Acid)
Dimensions W × H (mm)	160.36 × 81.39
Weight (g)	140
Head rotation (degrees)	180
Head tilting (degrees)	45

The robot's head was mounted with two servo motors (SG90, manufactured by TowerPro) to provide pitch and yaw motions. The robot can turn its head 90° left and right from the center, and move up and down. The produced design can mimic 5 head positions of the wearer, namely, straight, left, right, left intermediate, and right intermediate at angles of 0°, 90°, −90°, 45°, and −45°, respectively. The exploded of KiliRo robot is presented in Figure 1 and its physical architecture is illustrated in Figure 2. The head rotation positions and the schematic of the robot are presented in Figures 3 and 4, respectively.

Initial head position of the robot is set at 0° and when it detects the wearer's head position as left, it turns 90°, and turns −90° when the wearer turns their head toward the right. Similarly, 45° and −45° are achieved when the robot detects left intermediate and right intermediate positions of the wearer's head. Even though the robot can achieve pitch motion on its head, it was not deployed during this study. The robot uses the camera mounted on its head to detect the wearer's head position and processes it using a Raspberry pi-3 small computer to detect and actuate its head accordingly. The list of hardware used in the robot is presented in Table 2. A TREK Ai-Ball portable Wi-Fi camera as the imaging sensor was used for the KILIRO robot. The Wi-Fi camera used can capture images at 30 Hz with a maximum resolution of 640 × 480. The camera possesses a range of focal length from 20 cm up to infinity with a view angle of 300°.

Table 2. Wearable KiliRo robot's list of hardware.

Hardware	Specification
Controller	Raspberry pi
Servo motor	TowerPro SG90
Servo controller	Pololu micro Maestro 18-Channer
Camera refresh rate	30 Hz
Camera resolution	640 × 480
Focal length of lens	20 cm
View angle	30 degree
Camera	WiFi Ai-Ball Camera
Battery	Li-Po 1200 mah 7.4 v

**Figure 1.** Wearable KiliRo robot—exploded view.**Figure 2.** KiliRo robot—physical architecture.



Figure 3. Wearable KiliRo robot's head pose orientations.

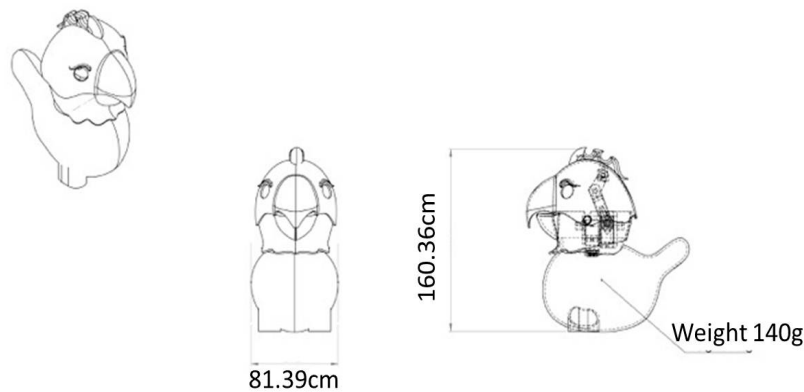


Figure 4. Wearable KiliRo robot's schematic diagram.

3. Learning Model

3.1. System Overview

Considering the constraints on power, size, and computational complexity, we explored the monocular vision embedded machine learning framework for the head pose recognition. We modeled the head pose identification as an object classification problem in the domain of computer vision. A machine learning based object classification paradigm is comprised of three steps. Feature identification and extraction is considered as the first and foremost amongst all the three, followed by the feature description which translates the extracted feature to a mathematical form where it can be used as an algebraic operand, and finally a classification model trained on numerous feature descriptors is used for the classification. An alternate approach to the above-mentioned classification scheme is the Artificial Neural Network (ANN) which works analogous to the biological brain. Besides, the main advantage of ANN over the typical classification is the absence of a dedicated feature extraction method for classification. Artificial neural network learning model consists of numerous layers stacked one over the other. Each layer consists of several nodes that encapsulate an “activation function”. The activation function decides whether the respective neurons should be “fired or not”. The information for training and testing are parsed into the ANN classification model through the “input layer”, and the input layer links to the preceding hidden layers, and all layers converge to the fully connected layer and finally to the output layer. Both classification schemes

mentioned require an enormous amount of training data to ensure the accuracy of the classification system. Typically, ANN requires comparatively more datasets to yield a better classification result than an SVM. However, if we consider the computational complexity induced by both the ANN and SVM classification models, ANN has an upper hand over the SVM because of the absence of complex computation induced by feature extraction. Furthermore, the ANN algorithms are friendlier for implementing in a computing device that allows parallel execution. From the perspective of real-time implementation of the classification system, computational complexity is among one of the critical challenges to address. In the scenario of a parrot inspired robot, the development of a classification system should acknowledge the following two factors: First, the system should be realizable with minimal requirements for computation. Second, the system should be realizable with minimal usage of sensors for data acquisition. Since we focus on the development of a wearable robot, the structural weight, energy consumption, size, etc. has to be minimized. Hence, the usage of a computation device that requires the development of simple and easily deployable drain surveillance and mosquito detection module. An effective way to address the above-mentioned challenge is the usage of a neural network based on an iterative learning method for classification, since it does not require dedicated feature extraction and description processes.

Extensive comparisons of different classification scenarios on SVM and artificial neural networks are reported in the literature. Support vector machines excel in performance over artificial neural networks in terms of time efficiency and accuracy in classification. However, the performance gap between neural networks and SVM in classification problems is almost negligible. Hence, in the case of developing a head pose detection module for a wearable robot, we have done a tradeoff between performance and computational complexity. We modeled the head pose detection to an image classification problem, and we are used convolutional neural networks. Supplemental research work has been reported regarding the development of head pose orientation. For instance, Rainer [39] used neural networks for head pose estimation to evaluate the participants in a workshop. Voit, Nickel, and Stiefelhagen [40] used the same approach for estimating the pan and tilt orientation on synthetic, high-resolution head images. They also used the neural network method for estimating horizontal head orientation on seminar recordings captured with multiple cameras from different viewing angles.

Even though deep learning based head pose detection is extensively studied, its application on a wearable robot is primitive. The proposed head pose identification strategy from the images taken from the shoulder of the target object emphasizes the novelty of the research work. The two major phases involved in both schemes include a head pose database generation and head pose classification. The database generation phase is done by acquiring head pose images from different persons. The images are taken using a camera mounted inside the eye socket of the robot. The user keeps the robot on his left shoulder and multiple images of their head are taken.

We remotely accessed the embedded computer using SSH secure shell protocol and the images of the head pose when the user oriented his head towards the left, right, left-intermediate, right intermediate, and straight positions. Figure 1 explains the various steps involved in both identification schemes using convolutional neural networks. We implemented the classification in a C++ platform linked with Caffe library on a Linux platform. The training of the neural network was done on a Linux PC, with the GPU and the pre-trained model stored in the embedded computer (Raspberry Pi). The learning of neural networks demands high computational requirements. However, the deployment and testing of a trained neural network does not require high computational power. Hence, we used a workstation with Nvidia GeForce GT-730 GPU to accelerate the training process and generate the pre-trained deployment file after the training process. After the training of AlexNet, the deployment file was transferred into the micro-computer of the parrot robot for real-time usage. The flow of the overall system is illustrated in Figure 5.

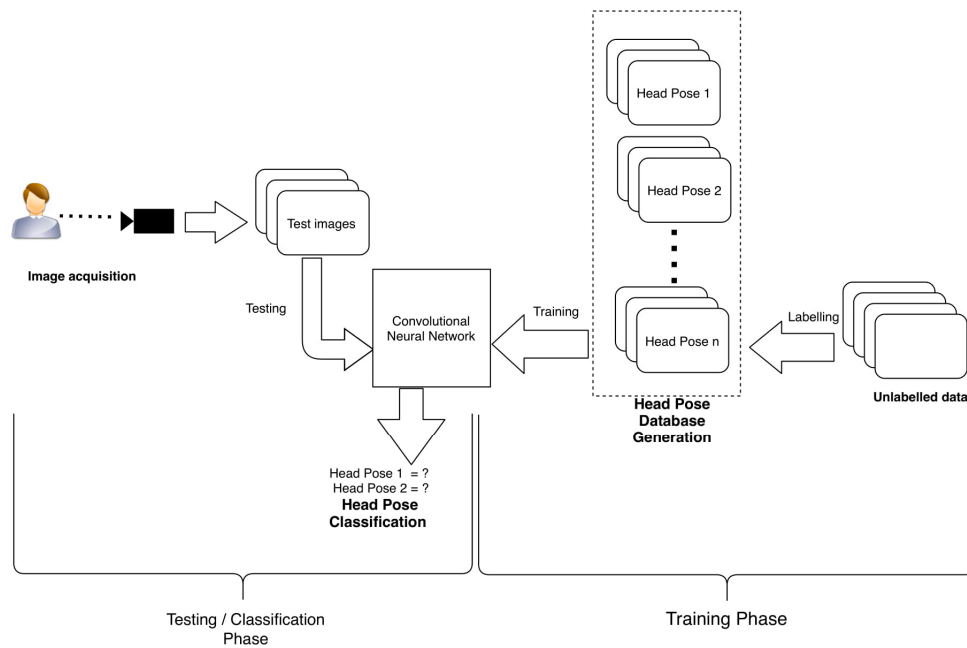


Figure 5. System flow diagram of the KiliRo robot.

3.2. Database Generation

The process of database generation involves acquisition of images required for the classification and training of deep neural networks. Here, our intention was to classify the images captured by the robot into Left, Right, Straight, Left Intermediate, and Right intermediate classes. Our database consists of Left, Right, Straight, Left Intermediate, and Right intermediate as the five classes. We captured 250 images for each class from five subjects, and they were captured in an indoor setting at our university laboratory with sufficient lighting. It took about 15 min to capture 250 images. We mounted the camera on the parrot robot's eye sockets and captured the images of the head pose while a person was wearing the robot. In addition, we chose four different individuals to wear the parrot robot and collect the images of different head poses. The number of classes is always scalable and not limited to the number five. In this case, we chose five different head poses that were visually distinct. We captured 250 images for each class. For generating diversity in the database, for each class we acquired images from four different individuals. Likewise, we have collected a total of 1000 images (labeled as 250 images from each class). The KiliRo robot uses an Ai-Ball camera for capturing the images of the head poses. Having a resolution of 640×480 (VGA), focal length of 200 mm, frame rate of 30 fps, and a view angle of 60 degrees, this camera met the requirements for our head pose detection needs. As the Ai-Ball camera is compact (diameter of 30 mm and 35 mm in length) and lightweight (less than 100 g), it was a perfect choice for our robot. This compact design makes it more suitable for wearable robots with limited on-board computational power. Figure 6 illustrates the examples of five head poses used in our study.

Before generating the image database, all the images are resized to a gray scale rectangular matrix of 60×60 . To use during the training phase, we computed the image mean for the created dataset. To boost the performance of the neural network, the calculated mean image was subtracted from the individual images in the dataset during the training phase figure. The calculation of the mean image ensures our data would have a zero mean. For instance, a training dataset X with five images can be represented as;

$$X = \{x_0; x_1; x_2; x_3; x_4\} \quad (1)$$

where x_i (i can take values 0, 1, 2, 3, 4) represents the individual image in the database.

From (1), the mean of the data is calculated as:

$$X_{Mean} = \frac{1}{4} \sum_{i=0}^4 x_i \quad (2)$$



Figure 6. Photographs of the wearer's head poses, (a) Left, (b) Left Intermediate, (c) Right, (d) Right Intermediate, and (e) Straight.

3.3. Training and Classification

The head pose detection was performed in the classification phase. The convolutional neural networks (CNN) are made up of multiple convolutions and a pooling layer followed by one or more fully connected layers. A series of convolution and pooling processes in the hidden layer realizes extraction of translational and rotational invariant features from the input image. The advantage of CNN over other classification methods is its simplicity in training and possession of few parameters to consider while training. We used AlexNet CNN architecture trained on the images from our database for implementing the head pose identification system. The AlexNet method consists of five convolution layers, and two fully connected layers (Figure 7). In between, there are layers that do the max pooling. Table 3 shows the AlexNet architecture.

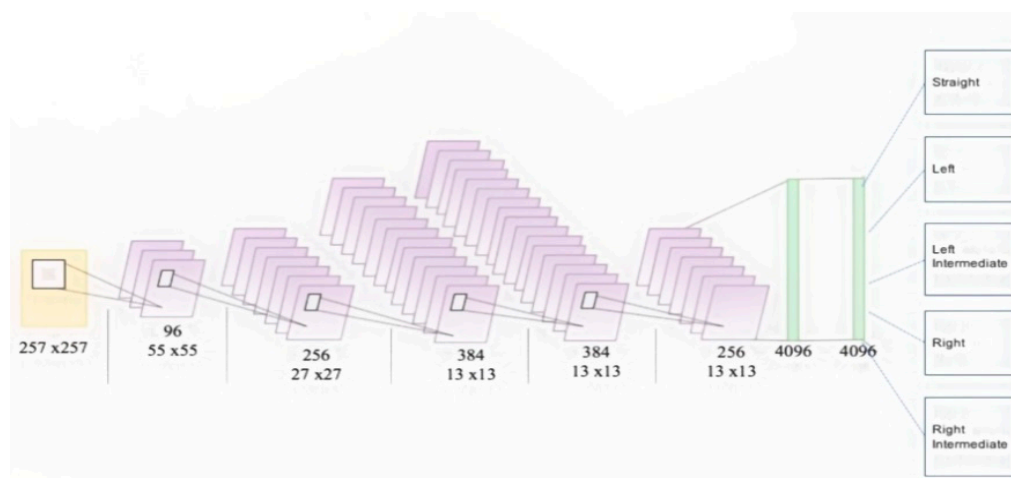


Figure 7. AlexNet method illustrations.

The activation function used was the ReLU (Rectified Linear Unit) for the convolution layers,

$$f(x) = \max(0; x) \quad (3)$$

Besides, ReLU we used a softmax function at the fully connected layers before the output. The softmax function normalizes the outputs of each unit in the fully connected unit to range between 0 and 1 such that the sum of output will be unity.

$$\sigma(z)_j = \frac{z^j}{\sum_{k=1}^K e^{z_k}} \quad (4)$$

Table 3. AlexNet architecture.

No.	Layer	Maps and Neurons	Kernel
0	Input	3 Maps of 227×227	-
1	Convolution	96 Maps of 55×55 neurons	11×11
2	Max Pooling	96 Maps of 27×27 neurons	3×3
3	Convolution	256 Maps of 13×13 neurons	5×5
4	Max Pooling	256 Maps of 13×13 neurons	3×3
5	Convolution	384 Maps of 13×13 neurons	3×3
6	Convolution	384 Maps of 13×13 neurons	3×3
7	Convolution	256 Maps of 13×13 neurons	3×3
8	Max Pooling	256 Maps of 6×6 neurons	3×3
9	Fully Connected	4096 neurons	1×1
10	Fully Connected	4096 neurons	1×1

The network used had five outputs that corresponded to five different classes (head poses). Hence, the output neuron corresponding to the input image will be activated by the softmax function. The neural network architecture mentioned in Table 1 is represented as protocol buffer data. The protocol buffer language-neutral platform enables serialization of structured data in a faster and simple manner. Before training, the databases that consist of head-pose image sets were converted to a lightning memory mapped database (LMDB) format. The neural network was trained in CPU without GPU support. The learning of the network is performed in batch size of 128 using stochastic gradient descent (SGD). We have trained the network for 400 epochs at a learning rate of 0.0001. Figure 8 shows the learning rate vs. each epoch. The loss in the system was calculated by forward pass of the network. The figure shows loss in the training phase during each epoch in training. The mean image generated from the dataset is shown in Figure 9.

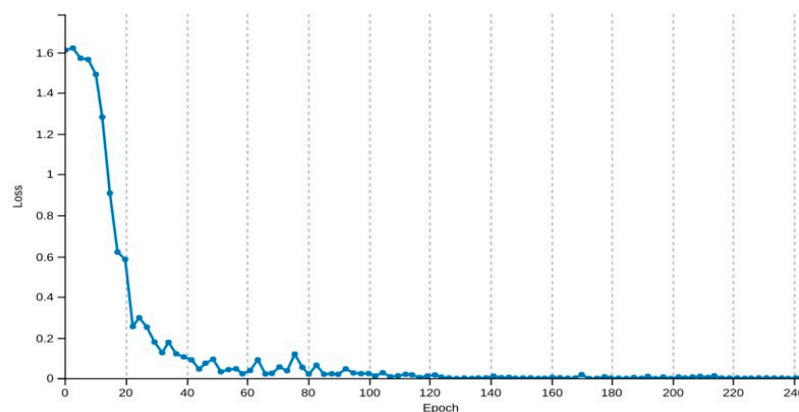


Figure 8. Graph representing loss vs. number of epoch during the training of the neural network with a prepared database.

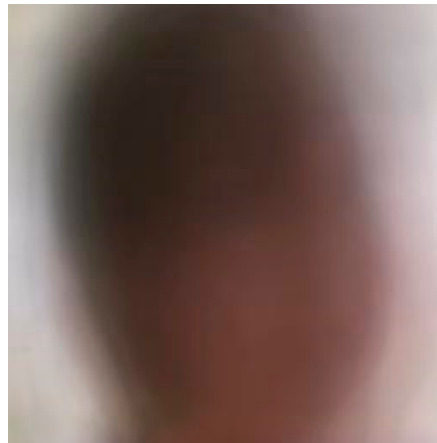


Figure 9. Mean image generated from the dataset.

4. Results

After creating the database for head pose detection, the testing phase was performed. As in the learning phase, the images required for the testing phase used the same camera and position on the wearer's shoulder. The illustrations of five head pose orientations in the experimental phase are presented in Figure 10. Table 4 presents the results of the head pose detection system tested with 250 images. Overall, the system resulted in an accuracy of 94.4%.



Figure 10. Illustrations of head pose orientation.

Table 4. Head pose detection experimental results.

	0	1	2	3	4	Accuracy (%)
0	50	0	0	0	0	100
1	0	48	0	2	0	96
2	0	0	50	0	0	100
3	0	0	0	50	0	100
4	0	11	1	0	38	76

The results were analyzed by determining the probability of obtaining the result if performance was simply random. For that purpose, data can be presented as a number of correct responses (i.e., turning the head in the same direction as the human) and number of incorrect responses (i.e., turning the head in the wrong direction). As there are five possible directions, the probability of getting the direction correct is 0.20. For three of the required responses, Class 0, 2, and 3, the robot emitted the correct responses for 50 out of the maximum number of 50 trials. According to a binomial distribution, the probability of obtaining this result is near 0. For images corresponding to Class 2, performance was not flawless, and the robot emitted the correct response on 48 of the 50 trials. For Class 4 images, responses were the worst, with responses correct for only 38 of the 50 trials. However, even in those two cases, the probability of obtaining this result simply due to chance is near 0. Performance is thus significantly better than chance ($p < 0.0001$) for all responses.

5. Conclusions

This paper presented a novel head pose detection system for a bio-inspired wearable pet robot, KiliRo, to classify five head poses from the side angle of the wearer. The presented deep learning based head pose detection system reported an overall accuracy of 94.4%. This is the first time head pose detection was demonstrated on a wearable pet robot. Even though the proposed system works well in bright indoor settings, the system delivers a sub-optimal performance in outdoor and low-light conditions. This problem can be addressed by upgrading the current imaging sensor to an advanced imaging sensors that delivers best output in low-light condition. Besides, the motion blur induced when the person moves his/her head in a faster manner is another factor that adversely affects the accurate detection of head-pose. The future extension of this research work will be focusing on integrating the camera with faster refreshing and better resolution by compromising the additional computational complexity introduced in the system. We also aim to enhance the system with more head pose detection abilities for our parrot-inspired pet robot and conduct real-time experiments involving children and adults to evaluate the effects of such robot in providing companionship. Furthermore, we will be working on improving the existing database of five classes to an advanced database with different subclasses to detect the head orientation in more detailed manner.

Author Contributions: J.B., L.H., and R.E.M. conceived and designed the experiments; J.B. and T.P. obtained the data and performed the experiments; C.K. analyzed the data; A.A.-J. provided conceptual and technical guidance for all aspects of the study. J.B. wrote the paper draft which was then modified by other authors.

Funding: The study is supported by the Institute of Biomedical Technologies, Auckland University of Technology, New Zealand.

Acknowledgments: This study is supported by the Institute of Biomedical technologies at the Auckland University of Technology, New Zealand. The funding is gratefully acknowledged. We also acknowledge the support from Wei Tong, Singapore University of Technology and Design during the experiment phase.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Graetzel, C.; Fong, T.; Grange, S.; Baur, C. A non-contact mouse for surgeon-computer interaction. *Technol. Health Care* **2004**, *12*, 245–257.
2. Kuno, Y.; Murashina, T.; Shimada, N.; Shirai, Y. Intelligent wheelchair remotely controlled by interactive gestures. In Proceedings of the 15th International Conference on Pattern Recognition (ICPR-2000), Barcelona, Spain, 3–7 September 2000; Volume 4, pp. 672–675.
3. Droeschel, D.; Stückler, J.; Holz, D.; Behnke, S. Towards joint attention for a domestic service robot-person awareness and gesture recognition using time-of-flight cameras. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 1205–1210.
4. Yin, X.; Xie, M. Hand gesture segmentation, recognition and application. In Proceedings of the 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation (Cat. No.01EX515), Banff, AB, Canada, 29 July–1 August 2001; pp. 438–443.

5. Lee, S.W. Automatic gesture recognition for intelligent human-robot interaction. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 10–12 April 2006; pp. 645–650.
6. Hasanuzzaman, M.; Ampornaramveth, V.; Zhang, T.; Bhuiyan, M.A.; Shirai, Y.; Ueno, H. Real-time vision-based gesture recognition for human robot interaction. In Proceedings of the IEEE International Conference on Robotics and Biomimetics, Shenyang, China, 22–26 August 2004; pp. 413–418.
7. Hasanuzzaman, M.; Zhang, T.; Ampornaramveth, V.; Gotoda, H.; Shirai, Y.; Ueno, H. Adaptive visual gesture recognition for human–robot interaction using a knowledge-based software platform. *Robot. Auton. Syst.* **2007**, *55*, 643–657. [[CrossRef](#)]
8. Breazeal, C.; Takanishi, A.; Kobayashi, T. Social robots that interact with people. In *Springer Handbook of Robotics*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1349–1369.
9. Uddin, M.T.; Uddiny, M.A. Human activity recognition from wearable sensors using extremely randomized trees. In Proceedings of the International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 21–23 May 2015; pp. 1–6.
10. Asri, H.; Mousannif, H.; Al Moatassime, H.; Noel, T. Big data in healthcare: Challenges and opportunities. In Proceedings of the International Conference on Cloud Technologies and Applications (CloudTech), Marrakech, Morocco, 2–4 June 2015; pp. 1–7.
11. Jalal, A.; Kamal, S.; Kim, D.S. Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System. *KSII Trans. Int. Inf. Syst.* **2018**, *12*. [[CrossRef](#)]
12. Zhan, Y.; Kuroda, T. Wearable sensor-based human activity recognition from environmental background sounds. *J. Ambient Intell. Humaniz. Comput.* **2014**, *5*, 77–89. [[CrossRef](#)]
13. Jalal, A.; Kamal, S.; Kim, D. A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems. *Int. J. Int. Multimed. Artif. Intell.* **2017**, *4*. [[CrossRef](#)]
14. Ren, Y.; Chen, Y.; Chuah, M.C.; Yang, J. User verification leveraging gait recognition for smartphone enabled mobile healthcare systems. *IEEE Trans. Mob. Comput.* **2015**, *14*, 1961–1974. [[CrossRef](#)]
15. Jalal, A.; Sarif, N.; Kim, J.T.; Kim, T.S. Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home. *Indoor Built Environ.* **2013**, *22*, 271–279. [[CrossRef](#)]
16. Shier, W.A.; Yanushkevich, S.N. Biometrics in human-machine interaction. In Proceedings of the International Conference on Information and Digital Technologies, Zilina, Slovakia, 7–9 July 2015; pp. 305–313.
17. Ba, S.O.; Odobez, J.M. Recognizing visual focus of attention from head pose in natural meetings. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2009**, *39*, 16–33. [[CrossRef](#)] [[PubMed](#)]
18. Murphy-Chutorian, E.; Trivedi, M.M. Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. *IEEE Trans. Intell. Trans. Syst.* **2010**, *11*, 300–311. [[CrossRef](#)]
19. Ruiz-Del-Solar, J.; Loncomilla, P. Robot head pose detection and gaze direction determination using local invariant features. *Adv. Robot.* **2009**, *23*, 305–328. [[CrossRef](#)]
20. Mudjirahardjo, P.; Tan, J.K.; Kim, H.; Ishikawa, S. Comparison of feature extraction methods for head recognition. In Proceedings of the International Electronics Symposium (IES), Surabaya, Indonesia, 29–30 September 2015; pp. 118–122.
21. Farooq, A.; Jalal, A.; Kamal, S. Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map. *KSII Trans. Int. Inf. Syst.* **2015**, *9*. [[CrossRef](#)]
22. Almudhahka, N.Y.; Nixon, M.S.; Hare, J.S. Automatic semantic face recognition. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 180–185.
23. Yoshimoto, H.; Date, N.; Yonemoto, S. Vision-based real-time motion capture system using multiple cameras. In Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI2003), Tokyo, Japan, 1 August 2003; pp. 247–251.
24. Koller, D.; Klinker, G.; Rose, E.; Breen, D.; Whitaker, R.; Tuceryan, M. Real-time vision-based camera tracking for augmented reality applications. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Lausanne, Switzerland, 15–17 September 1997; pp. 87–94.

25. Cheng, W.C. Pedestrian detection using an RGB-depth camera. In Proceedings of the International Conference on Fuzzy Theory and Its Applications (iFuzzy), Taichung, Taiwan, 9–11 November 2016; pp. 1–3.
26. Kamal, S.; Jalal, A.; Kim, D. Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM. *J. Electr. Eng. Technol.* **2016**, *11*, 1921–1926. [[CrossRef](#)]
27. Yamazoe, H.; Habe, H.; Mitsugami, I.; Yagi, Y. Easy depth sensor calibration. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 465–468.
28. Jalal, A.; Lee, S.; Kim, J.T.; Kim, T.S. Human activity recognition via the features of labeled depth body parts. In Proceedings of the International Conference on Smart Homes and Health Telematics, Artimino, Italy, 12–15 June 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 246–249.
29. Jalal, A.; Kim, J.T.; Kim, T.S. Development of a life logging system via depth imaging-based human activity recognition for smart homes. In Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea, 19 September 2012; Volume 19.
30. Ye, M.; Yang, C.; Stankovic, V.; Stankovic, L.; Kerr, A. Gait analysis using a single depth camera. In Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 285–289.
31. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the Science and Information Conference (SAI), London, UK, 27–29 August 2014; pp. 372–378.
32. Jalal, A.; Kim, Y.; Kim, D. Ridge body parts features for human pose estimation and recognition from RGB-D video data. In Proceedings of the Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Hefei, China, 11–13 July 2014; pp. 1–6.
33. Fu, Y.; Yan, S.; Huang, T.S. Classification and feature extraction by simplexization. *IEEE Trans. Inf. Forensics Secur.* **2008**, *3*, 91–100. [[CrossRef](#)]
34. Bharatharaj, J.; Huang, L.; Al-Jumaily, A.; Mohan, R.E.; Krägeloh, C. Sociopsychological and physiological effects of a robot-assisted therapy for children with autism. *Int. J. Adv. Robot. Syst.* **2017**, *14*, 1729881417736895. [[CrossRef](#)]
35. Bharatharaj, J.; Huang, L.; Al-Jumaily, A.M.; Krageloh, C.; Elara, M.R. Effects of Adapted Model-Rival Method and parrot-inspired robot in improving learning and social interaction among children with autism. In Proceedings of the International Conference on Robotics and Automation for Humanitarian Applications (RAHA), Kollam, India, 18–20 December 2016; pp. 1–5.
36. Bharatharaj, J.; Huang, L.; Mohan, R.E.; Al-Jumaily, A.; Krägeloh, C. Robot-assisted therapy for learning and social interaction of children with autism spectrum disorder. *Robotics* **2017**, *6*, 4. [[CrossRef](#)]
37. Rossignol, E.A.; Kominsky, M.A. Wearable Pet Enclosure. U.S. Patent 5,277,148, 11 January 1994.
38. Skloot, R. Creature comforts. *New York Times*, 31 December 2008, p. 31.
39. Stiefelhagen, R. Estimating head pose with neural networks—results on the pointing04 icpr workshop evaluation data. In Proceedings of the Pointing 2004 Workshop: Visual Observation of Deictic Gestures, Cambridge, UK, 22 August 2004; Volume 1.
40. Voit, M.; Nickel, K.; Stiefelhagen, R. Neural network-based head pose estimation and multi-view fusion. In Proceedings of the International Evaluation Workshop on Classification of Events, Activities and Relationships, Southampton, UK, 6–7 April 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 291–298.

