

# Enhancing Object Recognition for Humanoid Robots through Time-Awareness

Andreas Holzbach<sup>1</sup> and Gordon Cheng<sup>2</sup>

**Abstract**—In this paper, we present a biologically-inspired object recognition system for humanoid robots. Our approach is based on a hierarchical model of the visual cortex for feature extraction and rapid scene categorization of natural images. We enhanced the model to be entropy-aware and real-time capable, to be able to realize object recognition over time. We integrate time in our system to model uncertainty in static object recognition by evaluating multiple recognition results of objects observed at different view-points over time using the camera system on a humanoid robot. The recognition responses are encoded as probability estimates over each trained object class. We apply a signal detection theory approach to describe the temporally and spatially distributed signals to gain a value of certainty about the object class. We show that our enhanced model outperforms the preceding model and that by integrating time as a variable we created a highly robust object recognition system.

## I. INTRODUCTION

Humans are capable of detecting and recognizing objects under the most complex circumstances. They can easily identify objects under most lightning conditions, orientation, color or size. Even objects in clutter pose little problems, in contrast to state-of-the-art computer-based object recognition systems, which struggle to perform adequately under varying situations. Therefore, it only makes sense – and maybe is the only successful way – to analyse how the visual system in biological systems works and use that knowledge for modelling those mechanisms to build a more likely effective and robust object recognition system. Only recently researchers began to look into possible architectures which process information similar to its biological prototype [1], [2], [3]. These models cover a sub-functionality of the vision processing performed by the brain; like visual attention, object recognition, tracking or learning.

Especially in the area of object recognition, models have been built as a proof-of-concept with little effort in situating them in the real-world, like realising it on a humanoid robot. HMAX [4] for example only focusses on reproducing a human’s performance in rapid scene classification – a test used in psychology and neurology to describe how fast and how well a human subject reacts to the task of distinguishing between categories in natural images shown for a very short period, where there’s no time for eye movement or shifts of

\*This work was supported (in part) by the DFG cluster of excellence Cognition for Technical systems CoTeSys of Germany, and also (in part) BMBF through the Bernstein Center for Computational Neuroscience Munich (BCCN-Munich).

Andreas Holzbach<sup>1</sup> and Gordon Cheng<sup>2</sup> are with the Institute for Cognitive Systems, Technische Universität München, Karlstr. 45/II, 80333 München, Germany. Email available at [www.ics.ei.tum.de](http://www.ics.ei.tum.de).

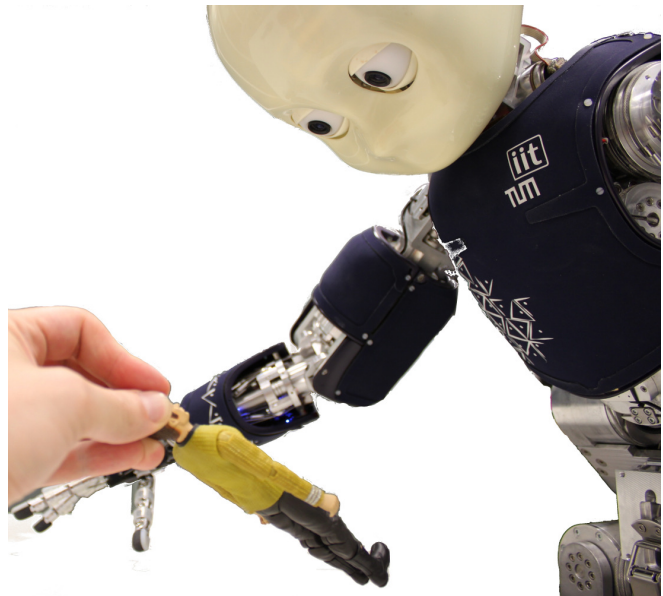


Fig. 1: Online Learning and Classification with the iCub.

attention [5], [6]. It is known that the entire process of object recognition activates much more areas in the brain than just the visual cortex, which indicates that in order to achieve similar efficiency in technical applications, simple feature generation and classification alone won’t solve the problem in the long run. We need to regard object recognition not as a distinct but as a cognitive process.

Object recognition in technical systems still is very limited in performance and confined to specific scenarios. Especially in humanoid robotics, it is essential to model uncertainty and make use of the robot’s abilities to act on it. With a humanoid robot and its active vision system and manipulators we have the tools to model the uncertainty by including the temporal aspect. Accounting for time could push current models from static single image recognition to a higher level of object consciousness.

Rapid scene classification as applied in studies is usually not the normal way how humans perceive their environment [7]. To identify objects we usually move our eyes to different salient areas to gain some kind of certainty about our belief what the object might be [8]. Depending on the visibility of the scene, this procedure might vary in time until some certainty is gained [9]. Our system reproduces this behaviour by applying a biologically-inspired object recognition model to a time-aware architecture.

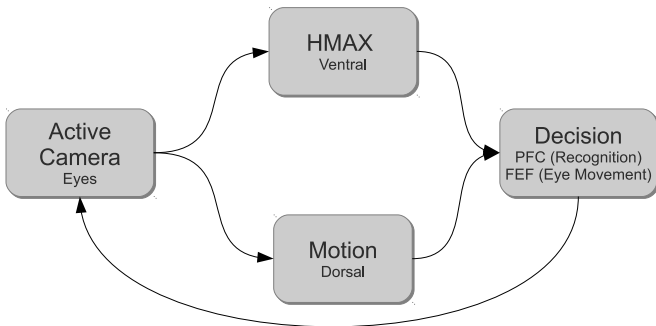


Fig. 2: System

## II. RELATED WORK

Recently, due to a deeper understanding of information processing in the brain and due to more powerful computational resources, the vision and robotics community started building more and more systems which gain their inspiration and functionality from biological models. Be it for building robots itself by studying the human corpus [10], to build robotic insects [11] or to integrate intelligence into humanoid robots [12]. Or to enhance common techniques like face recognition by using biologically-inspired features [13]. The widely applied SIFT features [14] are also inspired by neurons in the inferior temporal cortex. Some research draw more attention to active-vision systems, which have been used to solve different vision problems like: object recognition [15], [16], [17], [18]; visual search [19], [20]; visual attention [21]; or visual tracking [22]. It has also been investigated how to integrate object recognition [23], [24] and visual attention also with a focus on the aspect of computational complexity [25]. In this paper we specifically focus on the temporal aspect of object recognition, which has only received little attention so far. With this work, we wish to contribute further in this development, by enabling a robot to reason over time.

## III. SYSTEM OVERVIEW

Our system consists of a cluster of multiple PCs and GPUs to realize the hierarchical and parallel processing in the brain [26]. For communication between the different areas we employed the robot operating system (ROS), which supports modularity and makes it easy to extend our system both in hard- and software. The images are processed in two parallel nodes, which handle object recognition and motion sensitivity - similar to the ventral and dorsal stream in the visual cortex. The responses of both stream are integrated into the decision making node - which is slightly based on the functionality of the prefrontal cortex (PFC) and the frontal eye fields (FEF) - an area located in the PFC, which is responsible for guiding eye movement and saccades. The decision node integrates the classification response probabilities over time and the external motion in the visual field to calculates a certainty measure over the present object. If after several trails the robot is still uncertain about the object, it could move its eyes or torso to a different position to have a better view point, or move or turn the object itself.

### A. Object Recognition

The object recognition module presented in this paper is built on Serre *et al.*'s HMAX [4], which presents a feed-forward model of the visual cortex described by Riesenhuber and Poggio [27]. An overview is given in Figure 3. Each layer in the classical model consists of four alternating layers of simple cells (S1, S2) and complex cells (C1, C2) [28].

**S1 Layer:** The first layer is based on a representation of simple cells which react to oriented edges and bars in the receptive field. The response of these cells are quite similar to Gabor filters. The Gabor filters are created using the function

$$G_{\lambda, \theta, \psi, \sigma, \gamma}(x', y') = \exp\left(-\frac{x'^2 + y'^2 \gamma^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \quad (1)$$

with

$$x' = x \cos \theta + y \sin \theta \quad (2)$$

and

$$y' = -x \sin \theta + y \cos \theta \quad (3)$$

where  $\theta$  controls the orientation of the filter,  $\psi$  the phase offset,  $\sigma$  the variance of the Gaussian,  $\gamma$  the spatial aspect ratio and  $\lambda$  represents the wavelength of the sine function. The edge-sensitive cells contribute to the rotation invariance of the recognition system by the sensitivity to edges and bars of different orientations.

**C1 Layer:** Complex cells have a larger receptive field than simple cells and add some degree of spatial invariance and shift tolerance to the system. S1 cells of same scale band, same orientation and adjacent filter size are connected to a complex cell. The functionality can be described as a kind of max pooling operation; The maximum value of two adjacent filters of different sizes is calculated by using a sliding window approach.

**S2 Layer:** In the third layer small patches are chosen from random positions in the receptive field of C1. Each patch set consists of 4 patches, assembled by taking each patch in the set from a C1 response of different orientation ( $0^\circ, 45^\circ, 90^\circ, 135^\circ$ ) but same position and same scale band. Serre *et al.* use different sizes of patch sets: patch sets which contain patches of size 4; patch sets with patches of size 8; with size 12 and with size 16. These patch sets are then used for two different cases

Before the training or classification case, a dictionary of patch sets needs to be built. In the standard HMAX system these patch sets are chosen randomly over multiple images.

The S2 cell response is similar to a gaussian radial basis function and can be calculated as follows

$$r_{i,k} = \exp(-\beta \|X_i - P_k\|^2) \quad (4)$$

where  $\beta$  is the sharpness of the tuning.  $X_i$  is one of the patch sets created in the S2 layer and  $P_k$  is one of the "memorized" patch set in the earlier created dictionary. The

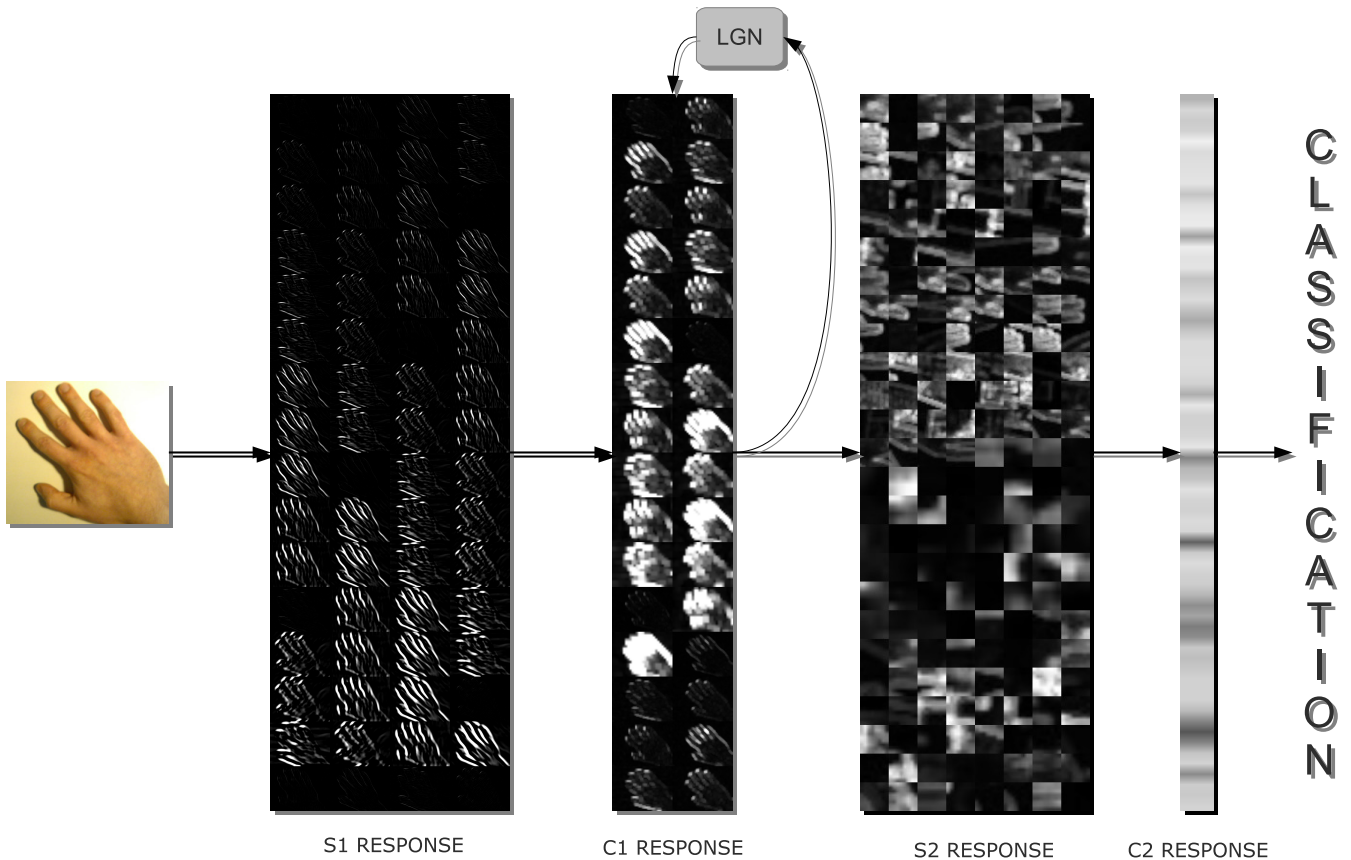


Fig. 3: Functional Overview of the architecture.

radial basis function is calculated for all patches  $i$  in the set of patch sets of S2 and for all patch sets  $k$  in the dictionary.

**C2 Layer:** Like in C1, the complex cells in the C2 layer now again perform a max operation over all the responses. For each element in the dictionary the maximum response for equation 4 is calculated using all the rbf responses of the patch sets of equal size. Using equation 4 this leads to

$$f_k = \max(\exp(-\beta\|X_i - P_k\|^2)); \forall i \quad (5)$$

which builds the feature vector  $F = \{f_0, f_1, \dots, f_d\}$  for all  $k$  in the dictionary, with  $d$  being the length of the dictionary. The feature vector can now be further used for training a classifier. For comparison reasons we used a SVM classifier as Serre *et al.* with a radial basis function kernel [4].

1) *Entropy Adaptiveness:* We enhanced this model by adding an information theoretic aspect of neural processing - the maximization of information along the pathway [29], [30]. Our system incorporates the information entropy in the S2 layer of the system. It is sensible in regard to the information a single patch carries and adaptively rejects patches which don't account for the overall information gain. We calculate the entropy of each patch by applying:

$$H(X) = - \sum_{m=1}^M p_m \log p_m \quad (6)$$

with  $p_m$  being the relative frequency of brightness value  $m$  within the patch. This approach filters out patches that show an almost plain distribution of intensities.

2) *Dictionary Creation:* In the standard HMAX implementation, the dictionary is created by randomly selecting patches as artificial neurons from a set of responses in C1. This approach bears the risk to select a non-optimal set with over-represented and redundant features. To avoid this disadvantage our method is derived by the functionality of lateral-inhibition in neurons. Only one artificial neuron per set is allowed to react above a threshold to a certain stimuli; this way there are no neurons which react too similar to the same input. Until the dictionary is not fully populated, we calculate the radial basis function response (see Eq. (4)) between each new neuron candidate and the current dictionary. If the response for all existing neurons in the dictionary is below a certain threshold the candidate is chosen as a new neuron in the dictionary. Mathematically, we can describe our dictionary as a set

$$D = \{x | \forall x, y \in D, x \neq y : \nexists y : rbf(x, y) > t\} \quad (7)$$

with  $D$  being the dictionary,  $rbf$  the radial basis function of Equation (4) and  $t$  as the threshold we wish to apply. Pseudo-Algorithm 1 displays how the dictionary is created. To make sure that the algorithm halts, we adaptively reduce the threshold  $t$  after several iterations.

---

**Algorithm 1:** Create Dictionary

---

**Data:** Dictionary  $D$ ; Dictionary size  $s$ ; Set of candidates  $C$ , Threshold  $t$   
**Result:**  $D$  with  $\text{size}(D) = s$   
**while**  $\text{size}(D) \neq s$  **do**  
  Create New Set Of Candidates( $C$ );  
  **forall**  $k$  **do**  
    **forall**  $i$  **do**  
      **if**  $\text{rbf}(D_i, C_k) > t$  **then**  
        break;  
      **end**  
    **end**  
  **end**  
   $D_{i+1} = C_k$ ;  
**end**  
**end**

---

### B. Recognition over time

Our approach uses a classifier which is able to output estimates of the membership probability of each class. Our evaluation (in Section IV-B) shows, that these probability estimates can be used as a certainty measure of the right classification. In terms of a time series of probabilities this allows us to make further assumptions: 1) If we see a high probability, we can assume less risk of false classification. 2) If we see a low probability, we're less certain, because we face a higher risk of a false negative test.

Therefore we assume, that the higher the response to a certain object and the more often this signal appears, the more likely it is that the response really represents the object we see. To model this behaviour, we make use of psychophysics – a method used for example in digital signal processing and cognitive neuroscience to describe the likeliness of a perceptual system's response to a frequent stimulus. We apply probability summation over time [31] - a method used in signal detection theory - which models the probability  $P$  that a signal is detected accounting for all  $P_i$ , with  $P_i$  being the probability that a temporal stimulus threshold is exceeded at time  $i$ .

$$P = 1 - \prod_i (1 - P_i) \quad (8)$$

Equation (8) is the probability for one channel. We are interested in  $n$  channels, or in our case  $n$  classes, which compete to reach the threshold. Therefore we apply a maximum function over the set of classes  $k$ :

$$P = \max_k (1 - \prod_i (1 - P_{i,k})) \quad (9)$$

In our case we chose the threshold to be at least  $\frac{1}{n} * 100\%$  - with  $n$  being the number of possible classes - to have a probability above 0% of getting detected, because if the probability is below the threshold, there exists at least one signal which has a higher value. We fit this constraint to an exponential distribution

$$f(x) = \alpha(\exp(x) - \beta) \quad (10)$$

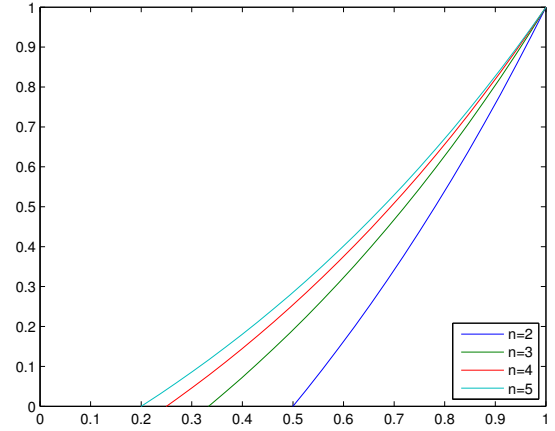


Fig. 4: Stimuli functions from Equation (13) for different number of classes  $n$ .

with

$$f(1) = 1; \quad f\left(\frac{1}{n}\right) = 0 \quad (11)$$

we get

$$f(x, n) = \frac{1}{\exp(1) - \exp\left(\frac{1}{n}\right)} (\exp(x) - \exp\left(\frac{1}{n}\right)) \quad (12)$$

Because Equation (12) is a continuous probability distribution defined for  $[1/n; 1]$ , we set  $P_{i,k,n} = f(R_{i,k}, n)$  from Equation (8) with  $R_{i,k}$  being the probability response for class  $k$  at time  $i$  from the classifier:

$$P = \max_k (1 - \prod_i (1 - \alpha(\exp(R_{i,k}) - \beta))) \quad (13)$$

with

$$\alpha = \frac{1}{\exp(1) - \exp\left(\frac{1}{n}\right)}, \quad \beta = \exp\left(\frac{1}{n}\right) \quad (14)$$

Figure 4 visualizes the resulting graphs of equation 12 for  $n = 2, 3, 4, 5$ .

### C. Motion Detection

Objects in the visual field normally don't suddenly disappear or change its structure. So according to the principle of temporal contiguity an association is made between objects seen in rapid succession [32]. Any difference would be interpreted as a displacement or masking of the object. Without the consideration of external motion or ego-motion, it would not be possible to build a believe system over an object in the visual field.

Therefore, we model motion as a trigger for resetting the classification believe certainty. If there is unsuspected motion from an external force e.g. something moves in the field of view or the object is taken away or replaced, we reset the believe probability in Equation (13) back to 0%. We model the motion detection as a stream separate from object recognition, which responses also end in the decision node. We detect the motion by reacting to a certain threshold to account for noise in the image data.

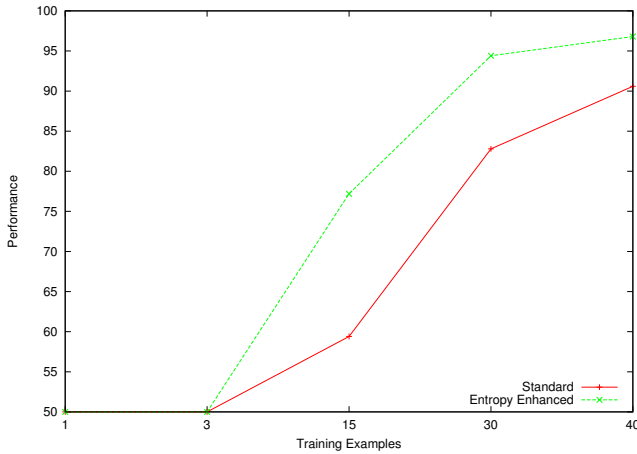


Fig. 5: Comparison of classification results for faces of the Caltech image database between the standard HMAX and our approach.

#### IV. EVALUATION

In Section IV-A we evaluate the object recognition module separately from the whole system to measure the performance of our object recognition approach over the standard HMAX. Then we test our assumptions over the probability estimates of the system (Section IV-B) and evaluate our approach in a real world scenario using the iCub humanoid robot (see Figure 1). A video contribution of our system connected with the paper is attached.

##### A. Single Object Recognition

We tested our object recognition module on the caltech-101 database; it contains 101 image categories and a background image set. Each category contains about 30-800 images. We chose airplanes, faces and cars image sets to test our approach against the standard HMAX model.

For each run, we created a dictionary and randomly picked 100 (where possible) images for training and for classification. Like in [33] we computed results with a number of positive training examples of 1, 3, 15, 30 and 40 images and negative training examples of 50 images. For each configuration the results were obtained and averaged after 10 runs. Each configuration consists of a different number of patches to pick during training and classification. Throughout the tests we chose a dictionary size of 200 (50 per patch size), which is one fifth the size compared to Serre *et al.* [33].

Our approach outperforms the original system in regard to the classification accuracy (see Figure 5), e.g. for the airplanes test set 93% compared to 86%, for the faces test set 96% compared to 90%, or cars 98% compared to 96%.

##### B. Propability Estimates over Time

The support vector machine classifier supports probabilistic outputs of the membership likeliness for each trained class [34], [35]. If a feature vector is more likely to represent an object of e.g. class 1, the probability for that class is higher than for the other classes. We tested how representative and

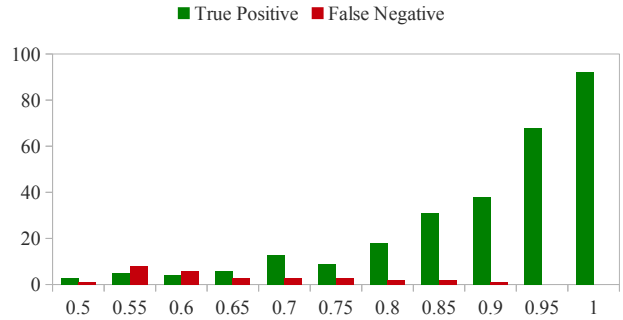


Fig. 6: Probabilities' frequency distribution of a two-class classification benchmark.

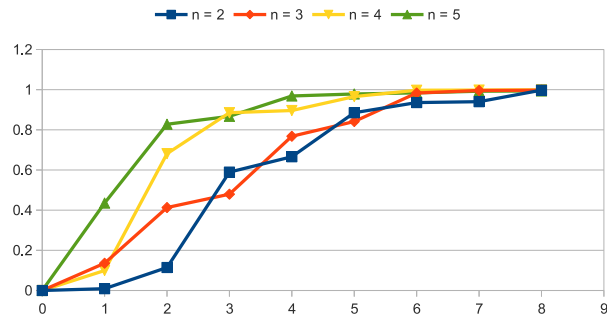


Fig. 7: Example of the believe graph development over time for different number of classes  $n$  for responses hitting the threshold.

valuable those results are for being applied to our system. Figure 6 shows the probabilities' frequency distribution of a two-class classification benchmark over multiple runs. The green bars represent the true positive, the red bars the false negative test results. It shows, that the probabilities give a good estimate of how likely the class assignment is. The false negative votes had an average probability of about 63% whereas the true positive probability was about 87% (keep in mind: these probabilities are not the classification average, but the average of the probability responses of the true positive/false negative tests). We ran multiple test to verify the correctness of our probability over time approach. Figure 7 shows the ascending graphs for different number of classes.

##### C. Real-world Scenario

We tested our approach for usability on real world scenario with live image acquisition. To be able to speed up processing we enhanced our system to be multi-core and multi-GPU capable. We used the iCub humanoid robot (see Figure 1) for image acquisition and the objects shown in Figure 8 for training and testing. We used the first three objects (hand, action figure, table tennis bat) for training. We obtained about 15 feature vectors for each object. We then tested the trained system on the two previous unseen figures (the iCub hand and a second action figure). All objects were correctly classified over time, because the system could cope with temporary misclassifications of the SVM classifier.



Fig. 8: Training and Test Objects.

1) *Motion Detection*: It is worthy to note, we experienced object mix-ups without the motion detection module. This happens if the object shown is exchanged after a certain believe value is already reached, so that the remaining time can't compensate for the new object. We were able to avoid this behaviour by activating motion detecting, which resets the probabilities and believe values. We achieve the motion detection by simple subtraction, thresholding and calculating the percentage of change pixels.

## V. CONCLUSION

In this paper we have presented a biologically-inspired object recognition framework, which applies signal detection theory to describe the temporally and spatially distributed signals over time by integrating separate probability estimates. This approach provided our robots with a certainty measure about their believe of the perceived world. We showed that our enhanced HMAX system outperforms current models in classification performance (up to  $\approx +6\%$ ) and speed, and that the integration of time as a variable helps to provide a highly robust object recognition system functioning in the real-world.

## REFERENCES

- [1] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. Rodriguez-Sanchez, and L. Wiskott, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1847–1871, 2013.
- [2] T. Poggio, J. Mutch, F. Anselmi, L. Rosasco, and Z. Joel, "The computational magic of the ventral stream : sketch of a theory ( and why some deep architectures work )." vol. 2011, 2012.
- [3] Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng, "Building high-level features using large scale unsupervised learning," *arXiv preprint arXiv:1112.6209*, 2011.
- [4] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *Computer Vision and Pattern Recognition, CVPR 2005*, vol. 2. Ieee, 2006, pp. 994–1000.
- [5] M. V. Peelen, L. Fei-Fei, and S. Kastner, "Neural mechanisms of rapid natural scene categorization in human visual cortex," *Nature*, vol. 460, no. 7251, pp. 94–97, 2009.
- [6] F. F. Li, R. VanRullen, C. Koch, and P. Perona, "Rapid natural scene categorization in the near absence of attention," *Proceedings of the National Academy of Sciences*, vol. 99, no. 14, pp. 9596–9601, 2002.
- [7] H. Deubel and W. X. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," *Vision Research*, vol. 36, no. 12, pp. 1827 – 1837, 1996.
- [8] M. J. Farah, "Is an object an object? cognitive and neuropsychological investigations of domain specificity in visual object recognition," *Current Directions in Psychological Science*, vol. 1, no. 5, pp. 164–169, 1992.
- [9] R. L. Goldstone, "Perceptual learning," *Annual review of psychology*, vol. 49, no. 1, pp. 585–612, 1998.
- [10] R. Pfeifer, M. Lungarella, and F. Iida, "Self-organization, embodiment, and biologically inspired robotics," *science*, vol. 318, no. 5853, pp. 1088–1093, 2007.
- [11] R. J. Wood, "The first takeoff of a biologically inspired at-scale robotic insect," *Robotics, IEEE Transactions on*, vol. 24, no. 2, pp. 341–347, 2008.
- [12] Y. Bar-Cohen and C. Breazeal, "Biologically inspired intelligent robots," in *Smart Structures and Materials*. International Society for Optics and Photonics, 2003, pp. 14–20.
- [13] E. Meyers and L. Wolf, "Using Biologically Inspired Features for Face Processing," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 93–104, July 2007.
- [14] D. Lowe, "Object recognition from local scale-invariant features," in *iccv*. Published by the IEEE Computer Society, 1999, p. 1150.
- [15] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: A survey of recent developments," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.
- [16] R. Bevec and A. Ude, "The acquisition of visual representations for object recognition by autonomous pushing," in *21th International Workshop on Robotics*. Esa, 2012, p. 156.
- [17] A. Andreopoulos, S. Hasler, H. Wersing, H. Janssen, J. K. Tsotsos, and E. Korner, "Active 3d object localization using a humanoid robot," *Robotics, IEEE Transactions on*, vol. 27, no. 1, pp. 47–64, 2011.
- [18] C. Goerick, H. Wersing, I. Mikhailova, and M. Dunn, "Peripersonal space and object recognition for humanoids," in *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*. IEEE, 2005, pp. 387–392.
- [19] B. Rasolzadeh, M. Björkman, K. Hübner, and D. Kragic, "An active vision system for detecting, fixating and manipulating objects in the real world," *The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 133–154, 2010.
- [20] T. Halverson and A. J. Hornof, "A computational model of active vision for visual search in human-computer interaction," *Human-Computer Interaction*, vol. 26, no. 4, pp. 285–314, 2012.
- [21] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 2, pp. 300–312, 2007.
- [22] V. Mahadevan and N. Vasconcelos, "Biologically inspired object tracking using center-surround saliency mechanisms," 2013.
- [23] A. Ude, D. Omrčen, and G. Cheng, "Making object learning and recognition an active process," *International Journal of Humanoid Robotics*, vol. 5, no. 02, pp. 267–286, 2008.
- [24] A. Ude, C. Gaskett, and G. Cheng, "Support vector machines and gabor kernels for object recognition on a humanoid with active foveated vision," in *Proceedings. IEEE/RSJ*, vol. 1. IEEE, 2004, pp. 668–673.
- [25] A. Ude, V. Wyart, L. Lin, and G. Cheng, "Distributed visual attention on a humanoid robot," in *Proceedings of*, 2005, pp. 381–386.
- [26] A. Holzbach and G. Cheng, "A neurologically motivated computational architecture for real-time object recognition," *Frontiers in Computational Neuroscience*, no. 258.
- [27] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature neuroscience*, vol. 2, no. 11, Nov. 1999.
- [28] D. Hubell and T. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, 1959.
- [29] A. Holzbach and G. Cheng, "An information theoretic approach to an entropy-adaptive neurobiologically inspired object recognition model," *Frontiers in Computational Neuroscience*, no. 135.
- [30] T. O. Sharpee, H. Sugihara, A. V. Kurgansky, S. P. Rebrik, M. P. Stryker, and K. D. Miller, "Adaptive filtering enhances information transmission in visual cortex," *Nature*, vol. 439, no. 7079, pp. 936–42, Feb. 2006.
- [31] A. B. Watson, "Probability summation over time," *Vision research*, vol. 19, no. 5, pp. 515–522, 1979.
- [32] N. Li and J. J. DiCarlo, "Unsupervised natural visual experience rapidly reshapes size invariant object representation in inferior temporal cortex," *Neuron*, vol. 67, no. 6, p. 1062, 2010.
- [33] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 3, pp. 411–26, 2007.
- [34] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *The Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.
- [35] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.