

Biologically motivated visual behaviors for humanoids: Learning to interact and learning in interaction

Christian Goerick, Inna Mikhailova, Heiko Wersing, and Stephan Kirstein

Honda Research Institute Europe GmbH

Carl-Legien-Strasse 30

63073 Offenbach / Main

Germany

Email: christian.goerick@honda-ri.de

Abstract—In this paper we present the current improvements of our biologically motivated interacting and learning vision system for humanoids. Building on the work presented in [1] the system features now a very natural gaze selection and interaction for learning freely presented complex objects in real-time. The new features are facilitated by two major contributions. First, by the introduction of an internal needs dynamics based on unspecific and specific rewards governing and exploring the parameterization of the basic behaviors. And second, by extending the object recognition pathway by sensory and object memory pathways as well as speech input / output for interactive confirmation and object labeling.

I. INTRODUCTION

The long term goal of this work is the creation of a humanoid robot that is equipped with mechanisms for learning and development. The concrete goal here is to present an interactively behaving vision system that comprises already both kind of mechanisms: autonomous developmental mechanisms influencing the behavior generation and selection of the system and interactive learning mechanisms allowing for teaching the system new objects to be recognized online. Related work exists for both kind of mechanisms separately, see for example [2]–[5] for developmental mechanisms and [6]–[10] for related work on online learning of visually defined objects. Both mechanisms regarded separately already represent a valuable step towards autonomous adaptive systems. Since we are pursuing a systems oriented approach it is important for us to show how these mechanisms can principally be combined.

In [1] we have presented a biologically motivated interactive vision system for humanoids with preset basic behaviors being able to recognize objects and gathering training data for offline learning of new objects. The basic architecture presented at this time was already designed in a scalable and flexible way in order to allow for further research towards more intelligence and autonomy in interaction and learning. The basic behavior of the system was already defined by a small set of parameters allowing for some kind of adaptation, but the parameters were fixed by design and not learned by the system. In this work we will lift this constraint and present a mechanism building on an internal needs dynamics based on

unspecific and specific rewards governing and exploring the parameterization of the basic behaviors. We will show that based on this mechanism the system can explore to interact in real-time, paving the way towards autonomous learning of behaviors. Those behaviors can be task-unspecific as in the case of general type of interaction, or more task-specific as in the case of learning and recognizing objects. We will show that the same basic mechanism can be employed governing both areas. The current major task of the system is the interactive learning of freely presented objects. Based on the work as presented in [1] we have extended the basic architecture by a sensory and an object memory pathway as well as speech input / output for interactive confirmation and and object labeling. Together with the visual interaction behaviors described above the system reaches a new quality in learning and recognizing new objects in real-time. Based on the memory and interaction concepts there is no artificial distinction between learning and recognizing anymore, and due to the achieved speed it is possible to correct misclassifications immediately during interaction.

The body of the paper is structured as follows. First we will present the overall architecture of the system, then we will focus on the extensions for online learning and behavior exploration. We will show the achieved performance by a series of experiments and conclude with a discussion and summary.

II. OVERALL SYSTEM ARCHITECTURE

The experimental setting is a humanoid stereo camera head with pan and tilt degree of freedoms. The overall systems architecture is as depicted in figure 1. The major functional blocks can be identified as follows: The basic gaze selection loop comprises image acquisition, visual saliency, motion- and disparity-saliency selection, summation and visuomotor mapping, gaze selection and head motor control. The visual gazing behavior is determined by this loop and mainly parameterized by the weights of the different saliency maps w_D , w_M and w_V for disparity, motion and visual saliency, respectively. For details see [1]. The relative weighting of the parameters

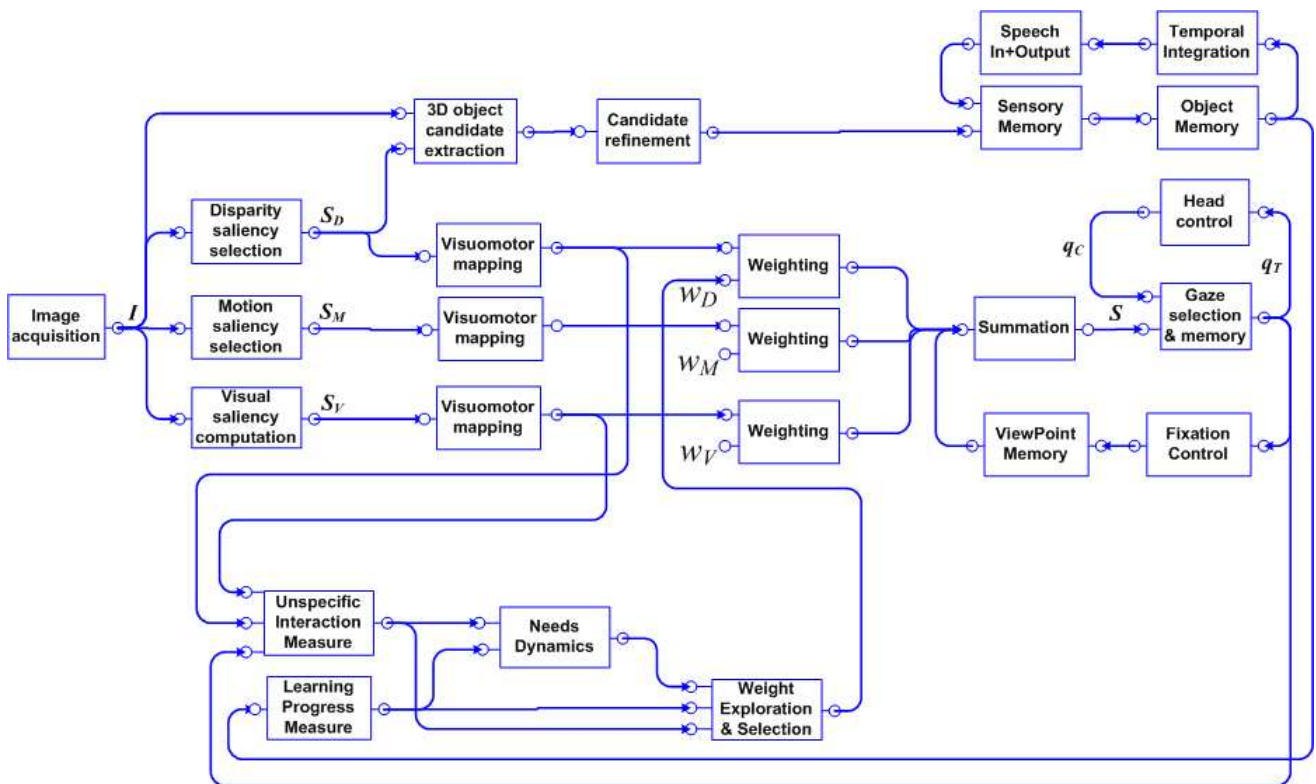


Fig. 1. Systems schematics. See text for detailed description.

determines the influence of the different saliency channels on the gaze.

Of special interest here is the disparity saliency selection and the parameter w_D . The disparity saliency selection performs a disparity computation and the selection of the closest connected region within a specific distance range and angle of view. The position of this region in image coordinates is represented as an activation blob within the map S_D . If there is no stimulus within the specified range and angle of view, the activation of the map is all zero. This simple mechanism represents a first approximation to the concept of the peripersonal space. It establishes a body-centered zone in front of the system that directly influences the behavior of the overall system. If the weight w_D is larger than zero the gaze direction is attracted to the closest object within the peripersonal space. If the weight is less than zero the system's gaze direction is repelled by the closest object within the peripersonal space. This mechanism can be used for sharing the object of interest with the system. With an appropriate set of weights the system focuses on the object a user holds into the peripersonal space of the system. In [1] the weight were heuristically fixed to be $w_V = 1.0$, $w_M = 3.0$ and $w_D = 4.0$. This results in an autonomously gazing system driven by visual saliency, that can be attracted by actions like waving with a hand at the system, and that continuously focuses on the closest object within the peripersonal space.

In section IV we will present an internal needs dynamics based on unspecific and specific rewards governing and ex-

ploring the parameterization of the basic behaviors. In other words the system will learn to interact with a user instead of having a hard coded parameter setting forcing the system to interact by design. We consider this extension as the basis of learning different kinds of behavior. The elements providing those kind of functions are the unspecific interaction measure as the unspecific reward, the learning progress measure as the specific reward, the needs dynamics as well as the weight exploration and selection.

The general kind of interaction the system can perform is the base for more task oriented performances. The first task we have chosen is the learning and recognition of freely presented complex objects in real-time. In [1] we have presented a system that could recognize those kind of objects already in real-time, but the learning and teaching did not meet real-time constraints yet. With the proposed structure for learning we build on the concept of peripersonal space for interacting. The visual region comprising the closest object within the peripersonal space is the candidate / hypothesis of the object to be learnt or recognized. In section III we present the extensions of the system leading to real-time performance of the system: the sensory memory, the working memory, the temporal integration and the speech input / output for labeling.

The resulting overall system then learns to interact and learns in interaction. In the experimental section we will show how the same needs mechanism providing the "learn to interact" part also improves the "learning in interaction" part, since the interaction can be terminated by the system if

the object learning has converged to a certain extent.

III. OBJECT LEARNING AND RECOGNITION

As described in section II, the visual region comprising the closest “object” within the peripersonal space is the candidate / hypothesis of the object to be learned or recognized. This region comprises usually the object of interest and the hand of the presenter. With a segmentation based on the adaptive scene dependent filters proposed in [11] the visual elements corresponding to the hand are removed and the segmentation of the object with respect to the background is improved. The classifier then has then to deal with the remaining visual parts not belonging to the object of interest.

Those enhanced segments are further processed by the model of the ventral visual pathway of Wersing & Körner [12] to obtain a complex feature map representation that is based on 50 shape and 3 color feature maps. The color channels are just downsampled images in the three RGB channels. The output is a high-dimensional view-based representation of the input object, that serves to classify or learn the current object. Those representations are stored within the sensory memory as long as the object did not leave the peripersonal space. This time history of sensory object hypotheses is communicated to the object memory.

Within the object memory, a persistent representation that carries consolidated and consistently labeled object views is created. As long as an object is presented within the peripersonal space and has not been labeled or confirmed, the obtained feature map representations of views are stored incrementally within the sensory memory. At the same time, all newly appearing views are being classified using the persistent object memory. If the human teacher remains silent, then the system will either generate a class hypothesis, or reject the presented object as unknown and verbalize this using the speech output module. The human teacher can confirm the hypothesis or make a new suggestion on the correct object label. As soon as feedback by the teacher is available, the learning architecture starts the concurrent transfer from the sensory memory buffer into the consolidated object memory. This extends over the whole history of collected views during the presentation phase and also proceeds with all future views, as long as the object is still present in the peripersonal space. The labeling of the current object can be done by the teacher at any time during the dialogue and is not restricted to being a reaction on a class hypothesis of the recognition system. The concept of a context-dependent memory buffer makes a separation into training and testing phases unnecessary. The transfer from the sensory to the object memory is sufficiently fast to remain unnoticed to the human trainer and the learning success can be immediately tested, allowing for a real online learning interaction.

The mechanism facilitating the online learning is an adaptive vector quantization working on the feature representations as detailed in [13]. For each class a set of reference vectors is maintained. During learning, new reference vectors are created if the incoming patterns are sufficiently different from the

already stored reference vectors for this class. For recognition, the incoming patterns are efficiently compared to the reference vectors of all classes.

The speech input and output is very important for the intuitive training interaction with the system. We use a system with a headset, which is the current state-of-the-art for speaker-independent recognition. The vocabulary of object classes is specified beforehand, to be able to label arbitrary objects we also use wildcard labels such as “object one”, “object two” etc.

The resulting system shows a natural and smooth interaction with users. The hypothesis built into the system is that objects are presented within the hand, otherwise there are no assumptions about the objects. The properties leading to the robust recognition are distributed over the system. Translation invariance is achieved by gazing, scale invariance by normalizing the 3D object hypothesis by distance estimation delivered by the disparity computation. Rotation invariance is enhanced by normalizing the first principle axis of the object. The online learning performance is facilitated by the efficiency of the hierarchical processing and by the locality of the plasticity, i.e. learning only on the highest hierarchy level.

For more details on the vision part the reader is referred to [14]. Here we focus on the more abstract level of elements in order to present the coupling with the behavioral part. The most significant elements here are the disparity saliency selection with the corresponding weight w_D and the object memory. The modulation of w_D determines whether the attention of the system is drawn towards the object within the peripersonal space ($w_D > 0$) or whether it is repelled by the object within the peripersonal space ($w_D < 0$). The behavior learning will build on this modulation. The object memory provides the signal for the learning progress. It is based on the number of reference vectors that are transferred from the sensory memory buffer into the consolidated object memory.

IV. GROUNDED BEHAVIOR LEARNING

The previous section described gaze selection with fixed weights of the saliency channels. The experiments (section V) show that such an attentional system provides a very robust and natural means of interacting with a robot. However it has a drawback: if an object is not presented by a human, but is simply a static part of the scene close to the system, it will also be fixated. This can be interpreted as a symbol grounding problem. The mapping from the depth signal to the interaction hypothesis is created by the designer. The reality does not always correspond to this mapping but the system can not find out the discrepancy on its own. It can not detect if the depth signal comes from “background” or from the user who wants to interact.

One possible solution would be the redesign of the hypothesis about the object to learn. Additionally to the closeness we could require the presence of motion, skin color or speech. There are two reasons why we prefer another solution.

First, the abovementioned percepts are not sufficiently robust. The user can present an object for a while without

moving and saying anything. He can do it also in a way that the skin is not visible. Also vice versa, it is possible that the speech, skin color and movement are present without an intention of the user to teach the system an object. The state of an object as "learnable" or not is rather hidden than perceptible.

The second, and more important reason is our aim to equip the system with means to recognize on its own the failure of the reactive behavior and to adapt appropriately. For this purpose we introduce two measures of the quality of systems behavior: the quality of interaction with the environment and the learning progress of the object recognition.

In [5] it was proposed to use the learning progress as a measure of getting better in predicting the results of one's behavior. Here the learning progress is not a general measure for all behaviors, but specific to the object recognition. In this way we can decouple the general evaluation of a situation as favorable for learning from learning progress which can be delayed or specific to the implementation of learning algorithm. In [15] we discuss in detail the difference between usage of specific and general rewards.

Generally speaking the quality of the interaction with the environment is high if the action of the robot leads to consistent sensory observations. In our example we measure directly the correlation between the gaze direction and the position of the object in the depth map. Thus the quality of interaction is high if either the robot tracks the object, or the object follows the gaze direction of the robot. Both situations are favorable for learning. The learning progress is high if the transfer occurs from the sensory memory buffer into the consolidated object memory.

In order to monitor these signals we give them a quality of rewards and introduce corresponding needs vector with two elements $N(t) \in R^2$. The needs are satisfied if their values are close to zero. If the needs are below a chosen threshold $N^0 > 0$ they are set to this threshold. Otherwise they change according to dynamics of the Lotka-Volterra type:

$$\tau_N dN/dt = N(t) .* (R_0 - R(t) - N(t))$$

where $.*$ means component-wise multiplication, τ_N is a time constant, $R_0 \in R^2$ characterize the speed of the need growth in absence of rewards, and $R(t) \in R^2$ are the corresponding rewards for the unspecific interaction quality and the learning progress.

If the reward is absent for a long time the need exceeds the threshold for starting the exploration. The system tries out a different weight w_D^{new} of the depth map according to the following simple heuristic: $w_D^{new} = w_D^{old} + S_E * D_E$, where D_E is the direction and S_E the strength of exploration. Exploration makes N_{hystr} steps in one direction. If in all of these steps the need continues to increase, then it changes the direction and increases the strength.

As soon as the system discovers negative values of the disparity weight it starts to avoid the object in the peripersonal space. If the object is just "background", then it does not react and there is nearly no correlation between the action of the system and the sensory map. If the object is shown by

a user, then it is natural for the user to slightly follow the head movement of the robot in order to stay in interaction.

This means that an "appropriate", from human point of view, behavior would be to stop tracking the object if there is no learning progress and force through avoiding of the object to provide a new view of the object or to provide a new object. If the interaction from user's side is observed, then the system should switch again into the tracking mode.

For such behavior the systems should "know" that

- 1) the tracking provides the maximal learning progress,
- 2) it should stop the tracking if the learning progress is missing for a long time,
- 3) if interaction is observable in avoidance, then it is probable to get the learning progress by tracking.

Below we describe how we implement the system with the above stated properties. The point two is created by design of the monitoring mechanism. The point three is covered by our choice of rewards because the probability of "good" tracking can be derived from the similarity of the rewards during interactive tracking and avoidance. The first point needs the representation of possible rewards and careful vector quantization of the behavior-reward space.

The actual implementation of the vector quantization is very simple because we were interested more in the interplay of the different parts of the system than in the perfect working of one part. We record the disparity weight and corresponding reward into a table of possible constellations. The table has the following format: $[W_M(i), R_M(i), C(i)], i \in [1 \dots M]$, where M is the number of recorded constellations, W_M represents a used weight of the disparity, and R_M represents the observed rewards. The confidence C keeps track of how good the entry matches the observing data. It is initialized with 1.0 and its updating will be described later.

The situation at the time-step t is compared to the entries of the table according to following formulas inspired by [16]:

$$\begin{aligned} \lambda_W(i, t) &= \exp(-\|W_M(i) - w_D(t)\|^2 / \delta_P^2) \\ \lambda_R(i, t) &= \exp(-\|R_M(i) - R(t)\|^2 / \delta_R^2) \\ \lambda(i, t) &= \lambda_W(i, t) * \lambda_R(i, t). \end{aligned} \quad (1)$$

The similarity measures $\lambda_W(i, t)$ and $\lambda_R(i, t)$ are close to 1.0 if the actual constellation is close to the table entry with index i . The entry with the highest responsibility $\lambda(i, t)$ is most similar to the actual constellation. The parameters δ_P and δ_R define the responsibility radius of the recorded constellations and thus the sampling rate of recording.

Every time a highest responsibility of known constellations is lower than a threshold λ_T , a new entry is added to the table. The confidence of the best matching entry i_{best} (with highest responsibility) is increased if the responsibility is above a threshold C_T and decreased otherwise:

$$\begin{aligned} C(i_{best}, t) &= f(C(i_{best}, t-1) + \tau_C * \Delta C) \text{ with} \quad (2) \\ \Delta C &= \sin((\lambda(i_{best}, t) - \lambda_T - C_T) / C_T * \pi / 2) \end{aligned}$$

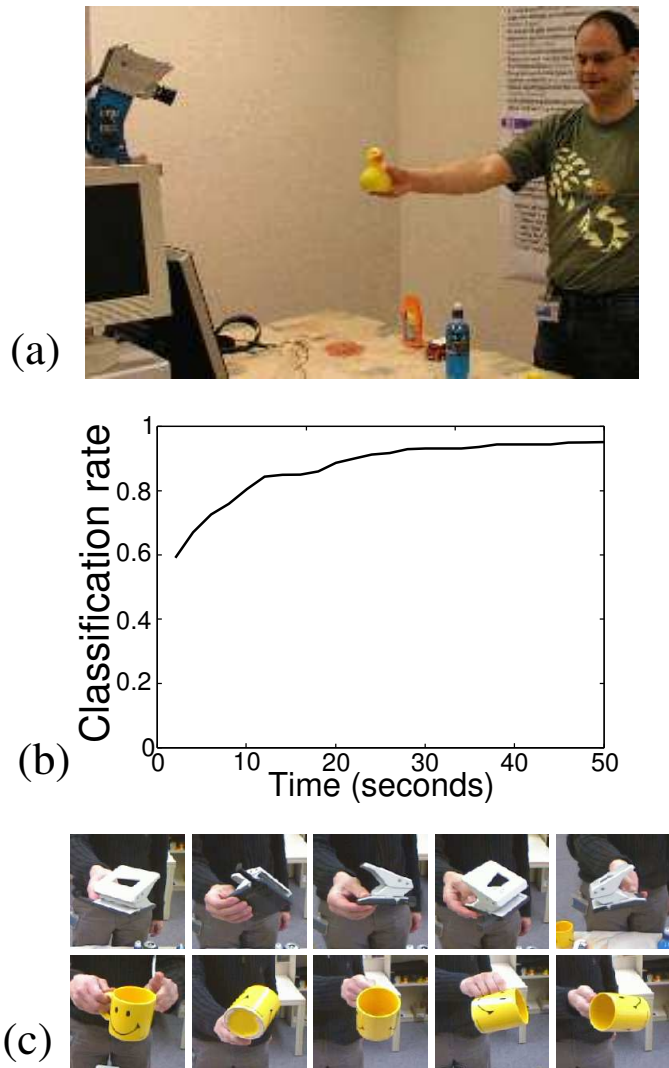


Fig. 2. Presentation scenario for our online learning architecture (a), and average recognition performance versus training time (b) for training the 10th object after 9 were already trained, with and without segmentation and temporal integration.(c) demonstrates the typical rotation variation that is applied during all experiments.

Here $f(x)$ is a step function, so that the confidence is truncated over 1.0 and below 0.0. If the confidence of the best matching entry is too low, then the entry is changed to the actual constellation.

While exploring possible constellations the system may record a very improbable situation (high learning progress while avoiding of the object). On the other side some typical situations are not always persistent. For example the tracking behavior can give a high learning progress if the user shows different object views and gives no learning progress if the user doesn't move an object. For this reason we can't use statistical learning. We have to decide whether a constellation has to be unlearned or whether it is only actually not possible for a short time period. For this purpose we introduce a temporary

measure of a reward mismatch:

$$R_{MM}(i, t) = \begin{cases} R(t) - R_M(i) + R_T, & \text{if } i = i_{best}, \\ 0.0, & \text{else.} \end{cases}$$

This value is positive if the actual reward is higher than recorded and negative if the difference to the recorded reward exceeds the tolerance margin R_T .

The reward mismatch gives an a posteriori information about how likely a reward recorded in the best matching entry is. For an a priori decision to switch the behavior we also need a priori information. We suppose that constellations with similar reward to the observed one are a priori more likely.

The similarity $\lambda_R(i, t)$ as a priori information and reward mismatch $R_{MM}(i, t)$ as a posteriori information both give us a hint if it is likely to get the reward recorded in entry i . This information is accumulated over time as likeliness $l_R(i, t)$ of getting reward recorded in entry i :

$$l_R(i, t) = \tau_l * l_R(i, t-1) + (1.0 - \tau_l) * (\lambda_R(i, t) + R_{MM}(i, t)),$$

where parameter τ_l expresses the conservaty of the system's belief about the actual context. This parameter together with the parameter of confidence decrease τ_C has to be chosen in the way that the switching of behavior occurs on the faster time-scale than switching of the entries in the recording table.

Finally the weight of the disparity map is selected in three steps:

- 1) The system monitors in which constellation it is by calculating responsibility $\lambda(i, t)$, eqn. (1).
- 2) The system calculates the likeliness of recorded constellations $l(i, t)$. The constellations are possible if their likeliness is higher than the threshold l_T .
- 3) The system chooses the weight from the possible constellations with the maximal reward for the highest need. We call the index of the chosen constellation i_{max} . If both needs are at the lowest level, the priority is given to the need of learning progress.

In the second part of the experimental section we will report on the character of the achieved behavior adaptation.

V. EXPERIMENTS

We conduct our experiments with a pan / tilt stereo camera head with humanoid dimensions. The first experiment we would like to report on is the interactive learning of freely presented complex objects. Here we assume optimal experimental conditions and cooperative users. Those experiments show primarily the performance of the online learning and recognition subsystem.

The complete system has been realized on a cluster of one dual processor PC for gaze control and image capture, one desktop PC running the speech recognition and synthesis system, and one dual processor PC performing all visual processing and online learning after the gaze selection. It is implemented within our integration framework [17]. The recognition system is running at a frame rate of roughly 6Hz, which enables interaction and online learning with direct feedback on the learning result. A generic training scenario

is shown in Fig. 2a, with typical ROI views of objects that are being processed. During all experiments the objects were freely rotated by hand to obtain a strong appearance variation.

In Fig. 2b we show a plot of the recognition performance versus training time during online learning. For this evaluation we train nine objects from a training set of 10 objects that was generated by storing 300 views per object from a typical training session. Then the tenth object is trained in steps of 10 images (1.67 sec in Fig. 2c) and a testing step is performed. The test is done by classifying a completely disjoint test set of 300 views per object that was collected using a different training person. Test performance is measured over all 300 test images of the currently trained object giving the classification rate as percentage of correctly recognized objects at this point of online learning. Then training proceeds until all 300 training images are used. The plot shown in Fig. 2b shows the resulting classification rate, averaged over an ensemble of experiments, where each of the 10 objects was one time the final object.

We visualize the actual time course of the different memory types during a training session of 18 objects in Figure 3. The plot displays the number of used representatives in the sensory and object memories together with the training dialogue (abbreviated, the actual dialogue is a little more elaborate). Starting from a completely empty object memory, we first perform a training of 10 objects. In this first phase the system first consistently matches the cola can to the previously trained “sun cream” object, and thus classifies the cola can initially as “sun cream”, which is then corrected by the teacher. Due to the similar red-white color and shape composition the “mini car” is also first confused with the cola can, and is corrected. Due to the shape similarity the green bottle is first labeled as blue bottle, which is a reasonable error, as long as no correction signal is given. After the feedback by the teacher, the system has learned to discriminate the first 10 objects after 5 minutes of training from many different viewing angles, which is evaluated directly afterwards. In the second training phase 8 objects are added. The initial confusion occurs quite reasonably between cola can and a yellow can, another red car and the mini car, a new blue mug and the first blueish patterned mug, and a new blue rubber duck and the initial yellow one. After the initial training in the second phase, the garlic press and police car object have to be additionally refined. After that second retraining phase, all 18 objects are classified from any reasonable viewing angle without further errors.

An important property of the system is that learning occurs most of the time and is not separated into artificial training and testing phases. This can be seen from the time course in Fig. 3, where during the first evaluation of the first 10 objects between 320s and 420s the object memory is still expanding, due to the confirmation signals of the human teacher on the system classifications. The same applies to the second evaluation and error correction phase between 640s and 850s. The complete duration of the session until no further recognition errors are encountered is about 12 minutes. This highlights the gain in learning speed that can be achieved due to the active error

correction process during learning. When the object memory is enlarged over time, we encounter a slight slowing down of the system frame rate from 6Hz to approximately 4Hz, since the comparison to the memory takes longer.

For the next experiment we assume that the user could be uncooperative, i.e. presenting objects without labeling or presenting objects statically without providing new views for learning.

Figure 4 shows the run of a typical experiment. The weight of the disparity channel is set initially to $w_D = 4.0$. This is a pre-designed solution as described in section II. The first entry made by the system into the record table corresponds to just looking around without interaction and learning. About the time-step 10 of the needs monitoring subsystem the user starts teaching a new object. The second entry put into the table reflects the fact that the system can receive a high interaction quality and learning progress during tracking. About time-step 20 the user introduces a static object. It can't be learned because it is not labeled. It is also not interacting, thus the system decides that it is in the situation without interaction. However the system doesn't know any other behavior than tracking yet and keeps fixating the static object.

With time the needs are growing over the threshold and the system starts exploration. During the exploration (time-steps 30 – 70) the system records 3 new constellations: that it can ignore an object ($w_D = 0.0$), avoid a static object and avoid an object that tries to stay in interaction ($w_D = -2.0$). After this exploration and learning phase the system shows a more appropriate behavior. While the user presents the object statically so that the learning progress decreases (time-steps 100 – 103) the likeliness to get learning progress from tracking decreases. At the time-steps 114, 115 the system switches to the avoidance mode ($w_D = -2.0$). But because the user follows the system switches back to tracking (time-step 116). Time-steps 180, 181 represent a similar situation. If the user doesn't follow the system remains in avoidance mode or switches to ignoring (steps 188 – 196).

The shown results are preliminary and a more careful analysis has to be done regarding the stability of results over a longer run with different user behaviors. Further we would like to investigate the sensitivity of the algorithms to the choice of parameters (thresholds, time constants, etc.). It would be also interesting to investigate more elaborative vector quantization mechanisms.

VI. CONCLUSION

In this paper we have presented the current improvements and state of our biologically motivated interacting and learning vision system for humanoids. To our knowledge it is the first system showing real online learning and recognition of several objects of arbitrary appearance in conjunction with an internal needs dynamics governing and exploring the parameterization of the basic behaviors. Both parts individually represent already major contributions to the current research landscape. It was one aim of this work to explore how these two parts can principally interact. We don't consider the proposed interaction

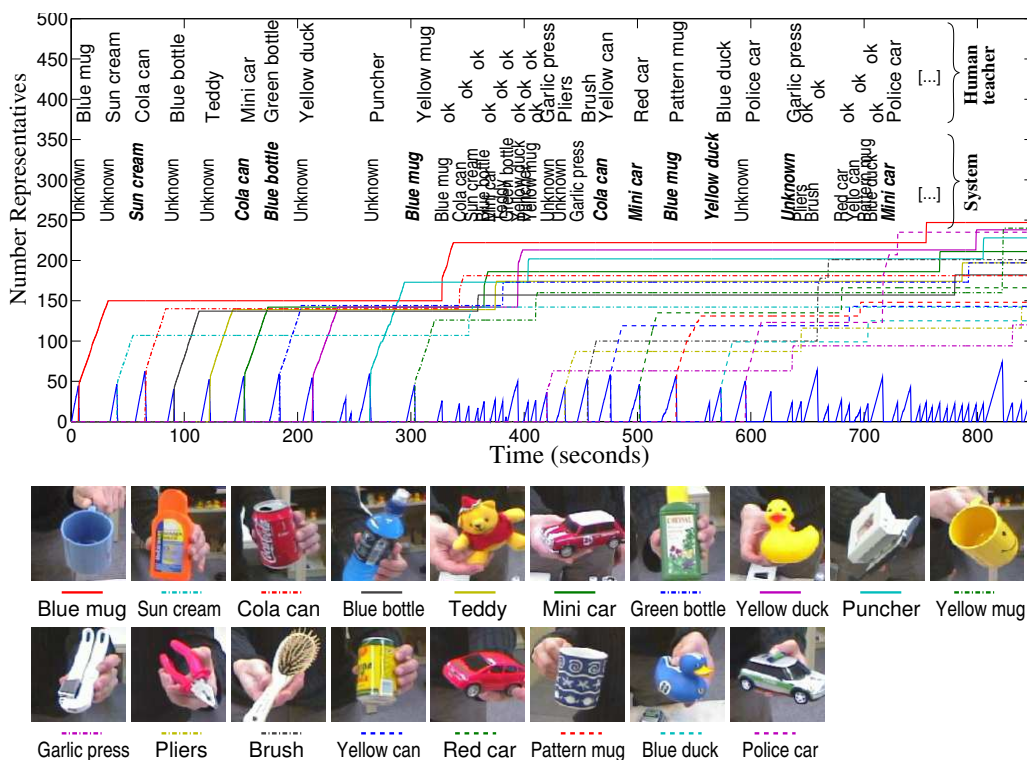


Fig. 3. Temporal learning dynamics during a training session for 18 objects. The plot shows the number of representatives for the sensory memory (“sawtooth” at bottom of plot) and representatives for each object in the object memory over time. The top row states the given labels by the human trainer, while the bottom row gives the classification results of the system, before a human labeling is given. Errors of the system are printed in bold italics. From 0 to 310s the first 10 objects are trained, the recognition of these 10 objects is evaluated from 320s to 420s without any errors. From 420s to 730s another 8 objects are added, and all 18 objects are checked after 730s without errors.

as final, but it provides an idea what kind of research questions can be addressed by such kind of integrations.

ACKNOWLEDGMENT

The authors would like to thank Mark Dunn, Jochen Steil, Bram Bolder, Antonello Ceravola, Marcus Stein, Martin Weißbach, Julian Eggert, Sven Rebhan, Herbert Janßen, Holger Brandl, Michael Götting and Edgar Körner for their discussions, contributions and advice.

REFERENCES

- [1] C. Goerick, H. Wersing, I. Mikhailova, and M. Dunn, “Peripersonal space and object recognition for humanoids,” in *Proceedings of the IEEE/RSJ International Conference on Humanoid Robots (Humanoids 2005)*, Tsukuba, Japan, 2005.
- [2] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini, “Developmental robotics: A survey,” *Connection Science*, vol. 15, no. 4, pp. 151–190, 2003.
- [3] O. Sporns and W. H. Alexander, “Neuromodulation and plasticity in an autonomous robot,” *Neural Netw.*, vol. 15, no. 4, pp. 761–774, 2002.
- [4] L. Meeden, D. Blank, and J. Kumar, D. and Marshall, “Bringing up robot: Fundamental mechanisms for creating a self-motivating, self-organizing architecture,” *Cybernetics and Systems*, vol. 36, no. 2, 2005.
- [5] P.-Y. Oudeyer and F. Kaplan, “The discovery of communication,” *Connection Science*, vol. 18, no. 2, 2006.
- [6] L.-M. Garcia, A. A. F. Oliveira, R. A. Grupen, D. S. Wheeler, and A. H. Fagg, “Tracing patterns and attention: Humanoid robot cognition,” *IEEE Intelligent Systems*, vol. 15, no. 4, pp. 70–77, 2000. [Online]. Available: <http://www.computer.org/intelligent/ex2000/x4070abs.htm>

- [7] L. Steels and F. Kaplan, “AIBO’s first words. The social learning of language and meaning,” *Evolution of Communication*, vol. 4, no. 1, pp. 3–32, 2001. [Online]. Available: <http://arti.vub.ac.be/steels/aibo.ps>
- [8] D. Roy and A. Pentland, “Learning words from sights and sounds: a computational model,” *Cognitive Science*, vol. 26, no. 1, pp. 113–146, 2002. [Online]. Available: [http://dx.doi.org/10.1016/S0364-0213\(01\)00061-1](http://dx.doi.org/10.1016/S0364-0213(01)00061-1)
- [9] A. Arsenio, “Developmental learning on a humanoid robot,” in *Proc. Int. Joint Conf. Neur. Netw. 2004, Budapest*, 2004, pp. 3167–3172.
- [10] H. Bekel, I. Bax, G. Heidemann, and H. Ritter, “Adaptive computer vision: Online learning for object recognition,” in *German Pattern Recognition Symposium*, 2004, pp. 447–454.
- [11] M. Götting, J. Steil, H. Wersing, E. Körner, and H. Ritter, “Adaptive scene-dependent filters in online learning environments,” in *Proceedings Eur. Symp. Neur. Netw. ESANN, Bruges*, 2006.
- [12] H. Wersing and E. Körner, “Learning optimized features for hierarchical models of invariant recognition,” *Neural Computation*, vol. 15, no. 7, pp. 1559–1588, 2003.
- [13] S. KIRSTEIN, H. WERSING, and E. KÖRNER, “Rapid online learning of objects in a biologically motivated recognition architecture,” in *27th Pattern Recognition Symposium DAGM*. Springer, 2005, pp. 301–308. [Online]. Available: documents/KirsteinWersingKoerner_DAGM05.pdf
- [14] H. Wersing, S. KIRSTEIN, M. GÖTTING, H. BRANDL, M. DUNN, I. MIKHAILOVA, C. GOERICK, J. STEIL, H. RITTER, and E. KÖRNER, “A biologically motivated system for unconstrained online learning of visual objects,” in *Proc. Int. Conf. Art. Neur. Netw. ICANN*, 2006, accepted.
- [15] I. Mikhailova, W. von Seelen, and C. Goerick, “Usage of general developmental principles for adaptation of reactive behavior,” in *Proceedings of the 6th International Workshop on Epigenetic Robotics, Paris, France*, 2006, accepted.
- [16] D. M. Wolpert and M. Kawato, “Multiple paired forward and inverse models for motor control,” *Neural Networks*, vol. 11, no. 7-8, pp. 1317–1329, 1998.
- [17] A. Ceravola and C. Goerick, “An integrated approach towards re-

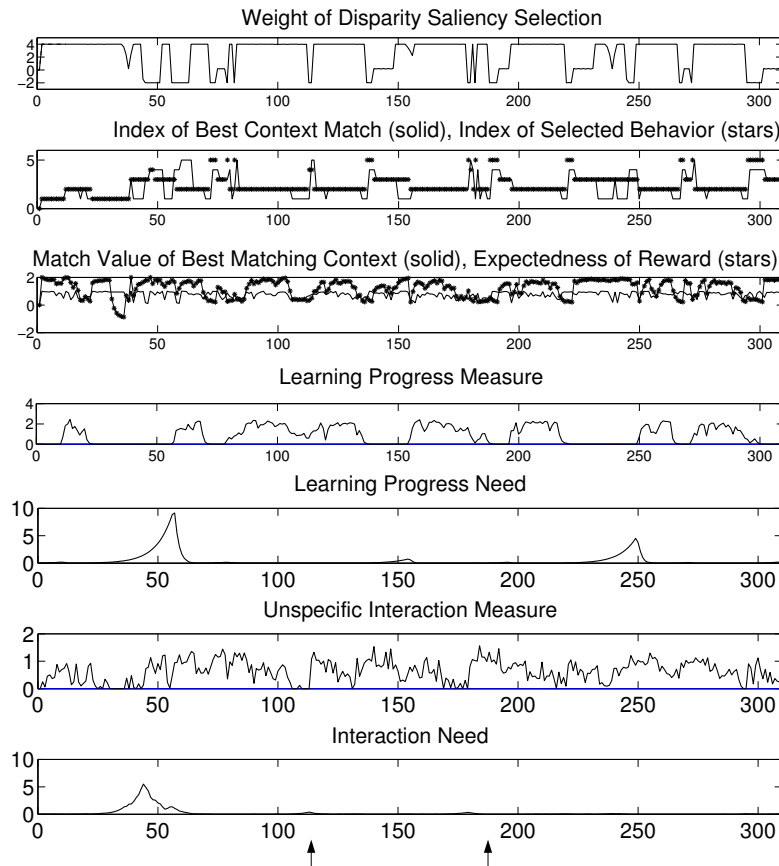


Fig. 4. Experiment run. See text for detailed description. The upper graph shows the evolution of the disparity saliency selection weight w_D . The second graph from the top shows the index i_{best} of the best matching context and the index (i_{max}) of the selected weight representing the selected behavior. The third graph from the top shows the value of the best matching context ($\lambda(i_{best})$) and the likeliness of receiving reward (li_{max}). The fourth graph from the top shows the learning progress measure as provided from the object memory. The fifth graph from the top shows the corresponding need. The last two graphs show the unspecific interaction measure and the corresponding need. The time-steps 114 – 116 shows the adaptation of the behavior in case of a short uncooperativeness of the user. The time-steps 188 – 196 show the adaptation in the case of a user who quits from interaction.

searching and designing real-time brain-like computing systems,” in *Proceedings of the First International Symposium on Nature-Inspired Systems for Parallel, Asynchronous and Decentralized Environments*, 2006.