Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

A Biologically Inspired Approach for Interactive Learning of Categories

Stephan Kirstein, Heiko Wersing

2011

Preprint:

This is an accepted article published in International Conference on Development and Learning. The final authenticated version is available online at: https://doi.org/[DOI not available]

A Biologically Inspired Approach for Interactive Learning of Categories

Stephan Kirstein
Honda Research Institute Europe GmbH
Carl-Legien-Strasse 30
63073 Offenbach/Main, Germany
Email: stephan.kirstein@honda-ri.de

Heiko Wersing
Honda Research Institute Europe GmbH
Carl-Legien-Strasse 30
63073 Offenbach/Main, Germany
Email: heiko.wersing@honda-ri.de

***Abstract*—An amazing capability of the human visual system is the ability to learn a large repertoire of visual categories. We propose an architecture for learning visual categories in an interactive and life-long fashion based on complex-shaped objects, which typically belong to several different categories. The fundamental problem of life-long learning with artificial neural networks is the so-called "stability-plasticity dilemma". This dilemma refers to the incremental incorporation of newly acquired knowledge, while also the earlier learned information should be preserved. To achieve this learning ability we propose biologically inspired modifications to the established learning vector quantization (LVQ) approach and combine it with a category-specific forward feature selection to decouple co-occurring categories. Both parts are optimized together to ensure a compact and efficient category representation, which is necessary for fast and interactive learning.**

I. Introduction

Humans are able to acquire and maintain knowledge during their complete lifetime, which commonly is termed life-long learning. In contrast to this, artificial neural networks typically acquire knowledge only during their learning phase and are fixed afterwards, so that the learned representation can not be extended afterwards. Although this strict separation in training and test phase can be powerful in constrained and stationary environmental settings but may not be suitable for applications like assistive robots or interactive agents. This is because these systems require a continuous error correction and need to enlarge their knowledge base to operate in changing and unpredictable environments. The capability of interactive learning is also important, making newly trained knowledge quickly available for the operation of the agent.

The fundamental problem of life-long and open-ended learning with artificial neural networks is the so-called "stability-plasticity dilemma". Here plasticity refers to the ability of a learning system to incorporate new acquired knowledge into its internal representation, while the learned knowledge should be preserved to guarantee the stability of previously learned information. This challenge occurs if a network model is trained with a limited and changing training ensemble that can be associated with the short-term memory (STM) concept in biology. Such limitation of the training ensemble is necessary, because it is infeasible to store all experiences that are encountered during the complete operation time of the system. Therefore we propose to incrementally learn representations under the condition, where a particular training vector can only be accessed for a limited time period. As a consequence the training with such a changing data ensemble typically causes the well-known "catastrophic forgetting effect" (French, 1999). This means that with the incorporation of newly acquired knowledge, parts of the previously learned knowledge is quickly fading out.

To enable life-long learning of arbitrary categories an exemplar-based neural network is combined with a category-specific feed-forward feature selection method, where the life-long learning of both parts is the major novelty of our proposed method. In this context a category is defined as a group of physical objects that shares visual properties. The target of the learning approach is to autonomously determine those properties based on the set of extracted features, where we assume independence of each single feature and that the category properties are reflected in the overall feature set. Beside life-long learning we are also targeting for fast interactive learning that allows learning in direct interaction with a human tutor. In the following we discuss related work addressing the biological motivation and neural networks approaching the stability-plasticity dilemma.

A. Inspiration from Psychology and Neuroscience

In the following we discuss major findings from psychology and neuroscience that influenced the development of our proposed interactive and life-long learning architecture.

1) Memory and Learning in the Human Brain: The separation into short-term memory (STM) and long-term memory (LTM) is an established biological concept (Izquierdo et al., 1999). The major distinction between both memory systems is the storage capacity, where in contrast to the LTM only very few items can be memorized in the STM. One opinion regarding the usefulness of this separation comes from the research on artificial neural networks. Here the combination of a fast learning STM and a slow learning LTM is considered to be one solution to avoid the "catastrophic forgetting effect" (French, 1999). We consider this memory separation important for our approach, therefore in the following the memory consolidation process for the information transfer between these different memory systems is discussed in more detail.

2) Memory Consolidation: There is evidence that the medial temporal lobe (MTL) is involved in the transfer of information from STM to LTM (Scoville & Milner, 1957). However, based on studies of patients with temporally-graded retrograde amnesia the role of the MTL, including the hippocampus, is only temporary. After the successful storage of contents in the neocortex, the LTM becomes gradually independent of the medial temporal lobe structures (Squire & Zola-Morgan, 1991). This means that the human brain is able to consolidate relevant information into its LTM representation based on a limited and continuously changing STM.

3) The Cholinergic System: The cholinergic system is one of the phylogenetically old modulatory systems that influences the brain by a diffuse projection of acetylcholine (ACh) into the extracellular space of many brain regions. In general it is known that the concentration of ACh increases with attention and novelty of sensory stimuli (Hasselmo & McGaughy, 2004). Of particular interest is its effect to the memory formation and its modulatory influence on synapses.

The level of ACh is crucial for the memory formation by switching between modes of acquisition (high ACh level) and consolidation (low ACh level) (Hasselmo & McGaughy, 2004). Furthermore this modulatory influence is known to reduce the interference of new representations with already stored knowledge by stronger suppressing previously potentiated synapses compared to naive synapses (Linster, Maloney, Patil, & Hasselmo 2003). In technical terms this effect can be interpreted as a modulation of the learning rate for each synapse with the effect that the stronger a synapse was potentiated in the past the smaller the learning rate gets. Therefore this modulatory effect of the cholinergic system supports life-long learning by reducing the "catastrophic forgetting effect" in the neocortex.

B. Life-Long Learning Architectures

Life-long learning architectures typically utilize exemplar-based learning techniques like learning vector quantization (LVQ) (Kohonen, 1989). Such neural architectures are beneficial for life-long learning, because for a specific input vector the learning methods modify only small portions of the overall network. Thus stability can be better achieved compared to the multi-layer perceptron (MLP), where all weights are modified at each learning step. Additionally for life-long learning architectures often a node specific learning rate combined with an incremental node insertion rule (Hamker, 2001; Furao & Hasegawa, 2006; Kirstein et al., 2008) is used to approach the "stability plasticity dilemma". The major drawback of those identification architectures is that they inefficiently separate co-occurring classes. For the categorization of natural objects we consider this co-occurrence of categories (e.g. red-white car) as common. Therefore we propose a category-specific forward feature selection method to overcome this limitation.

With respect to online and life-long learning of categories the work of Skočaj et al. (2008) is of particular interest. It enables learning of color and shape categories by selecting a single feature that describes the particular category most consistently. The corresponding category is then represented by an incremental kernel density estimation using mixtures of Gaussians (Skočaj et al., 2008). Although this architecture shares some common targets with our learning method, the restriction to a single feature only allows the representation of categories with little appearance changes. This is basically because more complex categories typically require several features to adequately represent all category instances.

II. Category Learning Vector Quantization

We combine an incremental exemplar-based network with a forward feature selection method to enable interactive and life-long learning of arbitrary categories. Both parts are optimized together to find a balance between the selection of features and allocation of representation nodes. We refer to this architecture as category learning vector quantization (cLVQ).

To achieve the interactive and incremental learning capability the exemplar-based network part of the cLVQ method is used to approach the "stability-plasticity dilemma" of life-long learning problems. Thus we define a node insertion rule that similar to the human brain determines when and where the representation has to be enhanced. The final number of allocated nodes $\mathbf{w}^k$ and the assigned category labels $\mathbf{u}^k$ corresponds to the difficulty of the different categories itself but also to the within-category variance. Finally inspired by the cholinergic system the long-term stability of these incrementally learned nodes is considered based on an individual node learning rate Θ^k as proposed in Kirstein et al. (2009).

Additionally a category-specific forward feature selection method is used to efficiently separate co-occurring categories, because it defines category-specific metrical "views" on the representation nodes of the exemplar-based network. During the learning process it autonomously selects an individual subset of features for each category. Note that only these selected category-specific features are used to decide whether a category is present or not. For guiding this incremental selection process a feature scoring value h_{cf} is calculated for each category c and feature f. This scoring value is only based on previously seen exemplars of a certain category, which can strongly change if further information is encountered. Therefore the h_{cf} values are continuously updated.

A. Distance Computation and Learning Rule

The learning in the cLVQ architecture is based on a set of high-dimensional and sparse feature vectors $\mathbf{x}^i = (x_1^i, \ldots, x_F^i)$ with $f = 1, \ldots, F$. Each $\mathbf{x}^i$ is assigned to a list of category labels $\mathbf{t}^i = (t_1^i, \ldots, t_C^i)$ with $c = 1, \ldots, C$. Each $t_c^i \in \{-1, 0, +1\}$ labels an $\mathbf{x}^i$ as positive or negative example of category c. The third state $t_c = 0$ is interpreted as unknown category membership, so that all $\mathbf{x}^i$ with $t_c^i = 0$ do not influence on representation of category c.

The cLVQ representative nodes $\mathbf{w}^k$ with $k = 1, \ldots, K$ are built up incrementally. Each $\mathbf{w}^k$ is attached to a label vector $\mathbf{u}^k$ where $u_c^k \in \{-1, 0, +1\}$ is the model target output for category c, representing positive, negative, and missing label output, respectively. The winning nodes $\mathbf{w}^{k_{\min}(c)}(\mathbf{x}^i)$ are

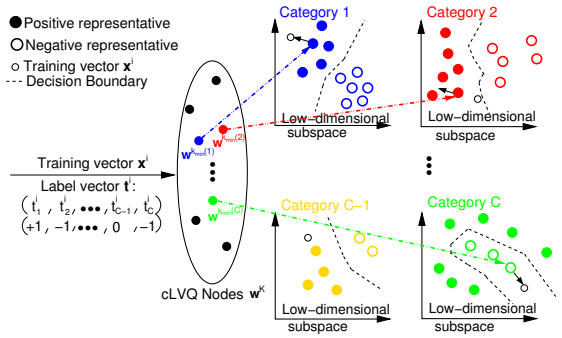

Fig. 1. **Illustration of the cLVQ Learning Rule.** Based on a training vector $\mathbf{x}^i$ and target vector $\mathbf{t}^i$ the winning nodes $\mathbf{w}^{k_{\min}(c)}$ are calculated for each category c independently. The distance computation is based on the selected category-specific features $f \in S_c$. If the categorization decision was correct the winning node $\mathbf{w}^{k_{\min}(c)}$ is shifted into the direction of the training vector. Otherwise $\mathbf{w}^{k_{\min}(c)}$ is moved in the opposite direction. If for an $\mathbf{x}^i$ the membership of a category is unknown ($t_c^i = 0$) no adaptation of the prototype node $\mathbf{w}^{k_{\min}(c)}$ is performed.

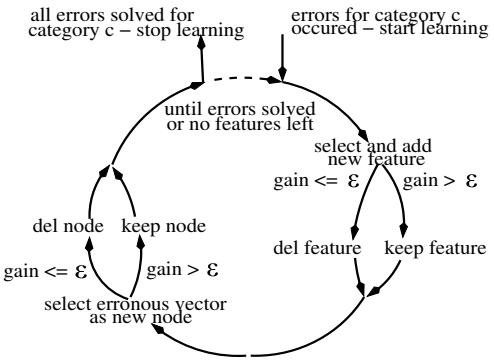

Fig. 2. **Illustration of the cLVQ Optimization Loop.** The basic idea of this optimization loop is to make small modifications to the representation of categories where categorization errors on the limited and changing set of training vectors occur. If the gain in categorization performance is above the insertion threshold the modification is kept and otherwise it is retracted.

calculated independently for each category c, where $k_{\min}(c)$ is determined in the following way:

$$k_{\min}(c) = \arg\min_k \sum_{f=1}^{F} \lambda_{cf}(x_f^i - w_f^k)^2, \quad \forall k \text{ with } u_c^k \neq 0. \tag{1}$$

The category-specific weights λ_{cf} are updated continuously and are inspired by the GRLVQ proposed by Hammer & Villmann (2002). We denote the set of selected features for category $c \in C$ as S_c. We choose $\lambda_{cf} = 0$ for all $f \notin S_c$, and otherwise adjust it according to a scoring procedure explained later. Each $\mathbf{w}^{k_{\min}(c)}(\mathbf{x}^i)$ is updated based on the standard LVQ learning rule (Kohonen, 1989), but is restricted to feature dimensions $f \in S_c$:

$$w_f^{k_{\min}(c)} := w_f^{k_{\min}(c)} + \mu\, \Theta^{k_{\min}(c)}(x_f^i - w_f^{k_{\min}(c)}) \quad \forall f \in S_c, \tag{2}$$

where $\mu = 1$ if the categorization decision for $\mathbf{x}^i$ was correct, otherwise $\mu = -1$ and the winning node $\mathbf{w}^{k_{\min}(c)}$ will be shifted away from $\mathbf{x}^i$. This node adaptation is illustrated in Fig. 1. Additionally $\Theta^{k_{\min}(c)}$ is the node-dependent learning rate as proposed in Kirstein et al. (2009).

B. Feature Scoring and Category Initialization

The learning dynamics of the cLVQ learning approach is organized in training epochs, where at each epoch only a limited amount of objects are visible to the learning method. After each epoch some of the training vectors $\mathbf{x}^i$ are removed and replaced by vectors of a new object. Therefore for each training epoch the scoring values $h_{cf} = \frac{H_{cf}}{H_{cf}+\bar{H}_{cf}}$, used for guiding the feature selection process, is updated. The variables H_{cf} and $\bar{H}_{cf}$ are the number of previously seen positive and negative training examples of category c, where the corresponding feature f was active ($x_f > 0$). The score h_{cf} defines the metrical weighting in the cLVQ representation space, with $\lambda_{cf} = h_{cf}$ for all $f \in S_c$ and $\lambda_{cf} = 0$ otherwise.

We assume that not all categories are known from the beginning, so that new categories can occur in each training epoch. Therefore if category c with the category label $t_c^i = +1$ occurred for the first time in the current training epoch, we initialize this category c with a single feature and one cLVQ node. We select the feature $v_c = \arg\max_f(h_{cf})$ with the largest scoring value and initialize $S_c = \{v_c\}$. The training vector $\mathbf{x}^i$ is selected as the initial cLVQ node, where the selected feature v_c has the highest activation, i.e. $\mathbf{w}^{K+1} = \mathbf{x}^q$ with $x_{v_c}^q \geq x_{v_c}^i$ for all i. The attached label vector $\mathbf{u}^{K+1}$ is chosen as $u_c^{K+1} = +1$ and zero for all other categories.

C. Learning Dynamics

Similar to the STM concept in biology all changes of the cLVQ network are only based on the limited and changing set of training vectors $\mathbf{x}^i$. During a single learning epoch of the cLVQ method an optimization loop (see Fig. 2) is performed iteratively that is inspired by the memory consolidation process. The basic concept behind this loop is to apply small changes to the representation of erroneous categories by testing new features v_c and representation nodes $\mathbf{w}^k$ that may lead to a performance increase. This optimization loop is composed of the following processing steps:

Step 1: Feature Testing. For each category c with remaining errors a new feature is temporally added and tested. If a category c is not present in the current training set or is error free then no modification to its representation is applied. This error-based learning is again motivated by the memory consolidation process so that representational changes are only applied if necessary. For interactive learning this is an efficient mechanism to optimally use the limited resources for learning. The feature selection itself is based on the observable training vectors $\mathbf{x}^i$, the feature scoring values h_{cf} and the e_{cf}^+ values. The e_{cf}^+ is defined as the ratio of active feature entries ($x_f^i > 0.0$) for feature f among the positive training errors E_c^+ of class c (see Kirstein et al. (2009) for further details).

For the feature testing a candidate v_c is added to the category-specific feature set S_c that potentially improves the categorization performance of category c by having a high scoring value h_{cf}. Additionally the feature candidate should

also quickly resolve remaining errors of this particular category. Therefore we choose $v_c = \arg\max_{f \notin S_c}(e^+_{cf} + h_{cf})$ and add $S_c := S_c \cup \{v_c\}$. The added feature dimension modifies the cLVQ metrics by changing the decision boundaries of all Voronoi clusters of category c, which potentially reduces the categorization errors. Thus based on all vectors $\mathbf{x}^i$ we calculate the categorization performance of the erroneous categories c. If the performance increase is larger than the prespecified threshold ϵ^1 the v_c is permanently added. Otherwise it is removed and is excluded for further iterations of this epoch.

Step 2: LVQ Node Testing. Similar to Step 1 we test new LVQ nodes only for erroneous categories. We propose to insert new LVQ nodes based on training vectors $\mathbf{x}^i$ with most categorization errors. This leads to a compact representation, because a single node typically improves the representation of several categories. In this optimization step we insert new representation nodes $\mathbf{w}^k$ until for each erroneous category c at least one new node is inserted. As categorization labels $\mathbf{u}^k$ for these nodes only the correct targets labels for the categorization errors are assigned. All other categories c are kept unchanged by setting $\mathbf{u}^k_c = 0$.

Again we calculate the performance increase based on all currently available training vectors. If this increase for category c is above the threshold ϵ^2, we make no modifications to the LVQ node labels of the newly inserted nodes. Otherwise we set the labels u^k_c of this set of newly inserted nodes $\mathbf{w}^k$ to zero. If due to this evaluation step all u^k_c become zero then we remove the corresponding $\mathbf{w}^k$.

Step 3: Stop condition. If all remaining categorization errors for the current training set are resolved or all features candidates f of the categories c are tested then the next training epoch is started. Otherwise go to **Step 1** and test further feature candidates and LVQ representation nodes.

III. Experimental Results

In the following section our proposed cLVQ architecture is compared with a single layer perceptron (SLP) and an incremental support vector machine (SVM) (Martinetz et al., 2009). The comparison with the SLP network architecture is done because this is the simplest neural network model that fulfills the requirements of the categorization task. Therefore SLPs are used to measure the baseline performance. For each category one output node is used. The output $out^{slp}_c = 1/1 + \exp(-\mathbf{w}^{slp}_c * \mathbf{x}^i - \theta_c)$, where $\mathbf{w}^{slp}_c$ is a single linearly separating weight vector with bias θ_c for each category c. The training is based on standard stochastic gradient descent in the sum of quadratic difference errors between training target and model output.

Additionally our proposed cLVQ method is compared with the support vector machine (SVM) approach. So far this established learning technique was mainly used for offline batch learning, but recently Martinetz et al. (2009) proposed the SoftDoubleMaxMinOver method enabling also an incremental training of SVMs. This incremental learning capability of the SoftDoubleMaxMinOver approach makes it interesting for our life-long learning task. We expect that the categorization performance of our cLVQ ranges in between the simple SLP and SVM. Nevertheless with the focus on interactive learning the memory consumption and learning speed are as important as the categorization performance. Furthermore we decided to use linear kernels and a hard margin criterion for all SVM experiments. In a pre-study we also tested rbf kernels, but the linear kernels achieve a comparable performance and in contrast to the rbf kernels the results are less affected by the exact choice of the parametrization.

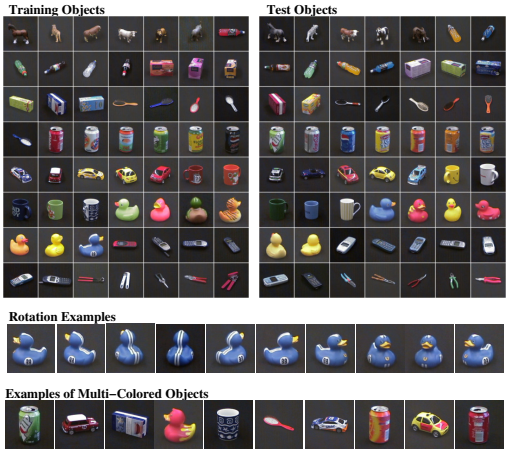

Fig. 3. **Object Ensemble.** Examples of all training (left) and test objects (right) used for our categorization task, where 15 different categories are trained. As color categories red, green, blue, yellow and white are trained. The shape categories are animal, bottle, box, brush, can, car, cup, duck, phone, tool. Each object was presented in front of a black background and is rotated around the vertical axis (bottom), resulting in 300 color images per object.

The chosen SVM method already supports incremental SVM learning, but so far only growing training sets are considered (Martinetz et al., 2009). In contrast to this we train for each category one SVM with a limited and changing set of training vectors. This means that at the beginning of each epoch the oldest object in the training set (limited to three objects only) is replaced by a new one. Furthermore we reinitialize the SVMs at each training epoch with all support vectors from the previous epoch and update them according to the current set of training vectors. To our knowledge this is the first attempt to train SVMs in this fashion and utilize them in the context of life-long learning.

A. Image Ensemble

As experimental setup we use an image database shown in Fig. 3 that is composed of 56 training and 56 distinct test objects, containing five different color and ten shape categories. Note that some objects are assigned to multiple color categories. Each object was rotated around the vertical axis in front of a black background. For each of the training and test objects 300 views are collected. For the labeling of each color category all views that cover $> 30\%$ of the object surface are labeled as $t_c = +1$, so that some objects are assigned to multiple color categories. If the covered area is below this value we choose $t_c = 0$, while $t_c = -1$ is used if the category is not present at all. For the shape categories one

c is assigned to $t_c = +1$ and all remaining shape categories are labeled with $t_c = -1$.

B. Feature Representation

For the representation of visual categories we combine simple color histograms with a parts-based feature representation, but we do not utilize this a priori separation for our category learning approach. We are particularly interested in discovering the structure of the categories from the feature vectors by using a flexible metrical adaptation. Therefore for each object view all extracted features are concatenated into a single structureless feature vector. We use color histograms because they combine robustness against view and scale changes with computational efficiency (Swain & Ballard, 1991). The parts-based shape feature extraction (Hasler et al., 2007) is based on a learned set of category-specific feature detectors that are based on SIFT descriptors. Commonly these descriptors are only determined around some highly structured interest points, while the used feature extraction method applies them at all image positions. This especially allows the representation of less structured categories. For the final shape feature response only the maximum detector value is selected, so that all spatial information is neglected.

C. Categorization Performance

Typically for offline categorization tasks the generalization to novel objects is the most important measurement. In contrast to this for our desired interactive learning the required training time is crucial, while for the life-long learning and the scalability of the learning problem also the memory efficiency is as important as the performance.

It can be seen in Fig. 4 that the categorization performance of the SVM approach is distinctly higher compared to the simpler SLP networks. This is astonishing, because both methods learn a single linear hyperplane for each category and therefore in principle can converge to a similar solution. This large difference in categorization performance is partly due to the additional large margin criterion of the SVM that especially for our relatively small amount of high-dimensional training vectors is beneficial with respect to a good generalization performance. Furthermore the incrementally collected support vectors are a good memory of the most important training vectors, because they are the vectors closest to the decision hyperplane. Therefore the used incremental SVM approach is suited for life-long learning. In contrast to this the SLP estimates the linear hyperplane by directly adapting the weights that typically causes the "catastrophic forgetting effect" for the limited and changing set of training vectors as used in our categorization task.

The performance of our cLVQ architecture and the SVM approach is comparable for the color categories as can be seen in Fig. 4. Thus for categories with a few stable and category-specific features our proposed feature selection method can capture the most relevant information of those categories. In contrast to this for the shape categories the SVM outperforms our proposed cLVQ learning method. The major reason for

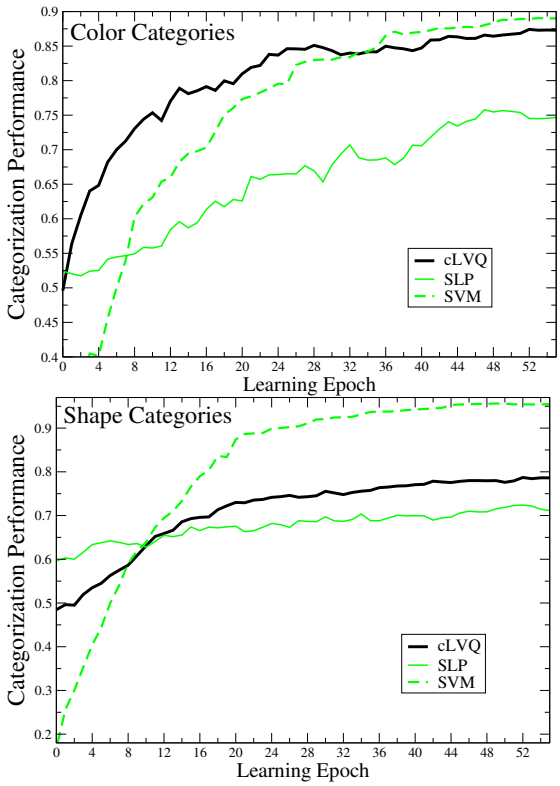

Fig. 4. **Comparison of Categorization Performance.** For this comparison of our proposed cLVQ with SVM and SLP we performed ten different runs with identical parameter set but random object order. The categorization performance is calculated after each training epoch, based on all test objects. This means that the performance is calculated based on the representation of the objects seen so far, simulating an interactive learning session. For color categories the performance of cLVQ and SVM almost identical, while the SLP performance is distinctly worse. In contrast to this for shape categories the SVM approach is superior to cLVQ and SLP. Although SVM reaches the highest performance it requires most training time and representational resources (see Fig. 5), whereas the cLVQ combines a good categorization performance with an resource efficient representation.

this lower generalization performance is because we consider only single features that are additionally less category-specific compared to the color features. Therefore the overall feature selection process becomes more difficult. This causes that for the first presented exemplar of a shape category by mistake often features of the co-occurring color category are selected, because they are more stable than the extracted shape features. Also the detection of combination features is difficult with the current cLVQ approach, so that the selection of multiple features is a promising future extension of our learning approach. Although there is still potential for improving the feature selection process, the cLVQ capture most relevant information and reaches a higher performance than the SLP networks.

With respect to the required network resources, depicted in Fig. 5, our learning approach is better suited for interactive and life-long learning. This is because cLVQ is one order of magnitude faster than the SVM approach and it is even faster than the simple SLP networks. This computational efficiency is caused by the selection of small sets of category-specific features, where for SLP and SVM all extracted features are

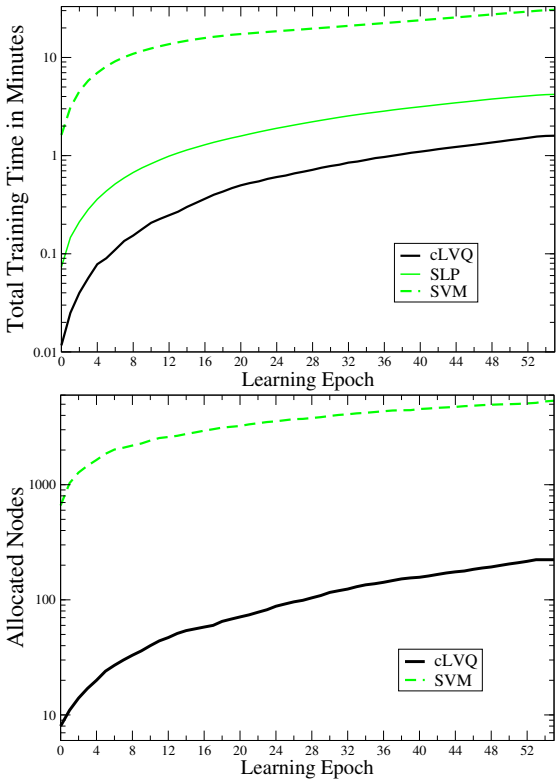

Fig. 5. **Comparison of Network Resources.** The total training time of the cLVQ, SVM and SLP is most crucial with respect to interactive learning. The cLVQ is one magnitude faster than the SVM and even two times faster than the simple SLP. This computational efficiency is caused by the forward feature selection process, where commonly only a small set of features is allocated for each category. This learning speed of the cLVQ enables fast interactive learning. Additionally the SVM approach has a memory requirement, which is undesirable for life-long learning tasks.

used during the training process. Additionally our cLVQ approach allocates much less representational nodes compared to SVM, which again positively influences the learning speed.

IV. Conclusion

The proposed biologically inspired cLVQ learning approach enables fast interactive and life-long learning of complex categories. The major novelty of cLVQ compared to state-of-the-art is the resource efficient and automatic allocation of network resources by combining an incremental exemplar-based neural network with a forward feature selection. This automatic control of the architecture complexity is crucial for our desired learning task, because an exhaustive parameter search is not feasible for interactive and life-long learning. In contrast to previous approaches on life-long learning manly targeting for identification tasks, our proposed feature selection method enables the cLVQ to separate co-occurring categories. Although the cLVQ enables interactive learning compared to the SVM there is still potential for improving the categorization performance of shape categories. Therefore the incorporation of some basic ideas from SVM into our feature weighting and selection framework is a promising direction for future work.

References

French, R. M. (1999). Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences 3(4), 128–135.

Furao, S. & Hasegawa, O. (2006). An incremental network for on-line unsupervised classification and topology learning. Neural Networks 1(19), 90–106.

Hamker, F. H. (2001). Life-long learning cell structures–continously learning without catastrophic interference. Neural Networks, 14, 551–573.

Hammer, B. & Villmann, T. (2002). Generalized relevance learning vector quantization. Neural Networks 15(8-9), 1059–1068.

Hasler, S., Wersing, H., & Körner, E. (2007). A comparison of features in parts-based object recognition hierarchies. In Proc. ICANN, 210–219.

Hasselmo, M. E. & McGaughy, J. (2004). High acetylcholine sets circuit dynamics for attention and encoding; low acetylcholine sets dynamics for consolidation. Progress in Brain Research, 145, 207–231.

Izquierdo, I., Medina, J. H., Vianna, M. R., Izquierdo, L. A., & Barros, D. M. (1999). Seperate mechanisms for short- and long-term memory. Behavioral Brain Research 103(1), 1–11.

Kirstein, S., Denecke, A., Hasler, S., Wersing, H., Gross, H.-M., & Körner, E. (2009). A vision architecture for unconstrained and incremental learning of multiple categories. Memetic Computing 1(4), 291–304.

Kirstein, S., Wersing, H., & Körner, E. (2008). A biologically motivated visual memory architecture for online learning of objects. Neural Networks, 21, 65–77.

Kohonen, T. (1989). Self-Organization and Associative Memory. Springer-Verlag, third edition.

Linster, C., Maloney, M., Patil, M. M., & Hasselmo, M. E. (2003). Enhanced cholinergic suppression of previously strengthened synapses enables the formation of self-organized representations in olfactory cortex. Neurobiology of Learning and Memory 80(3), 302–314.

Martinetz, T., Labusch, K., & Schneegaß, D. (2009). Soft-DoubleMaxMinOver: Perceptron-like Training of Support Vector Machines. IEEE Transactions on Neural Networks 20(7), 1061–1072.

Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. The J. of Neuropsychiatry and Clinical Neurosciences, 20, 11–21.

Skočaj, D., Kristan, M., & Leonardis, A. (2008). Continuous learning of simple visual concepts using incremental kernel density estimation. In Proc. VISAPP, 598–604.

Squire, L. R. & Zola-Morgan, S. (1991). The medial temporal lobe memory system. Science, 253, 1380–1386.

Swain, M. J. & Ballard, D. H. (1991). Color indexing. Int. J. of Computer Vision 7(1), 11–32.