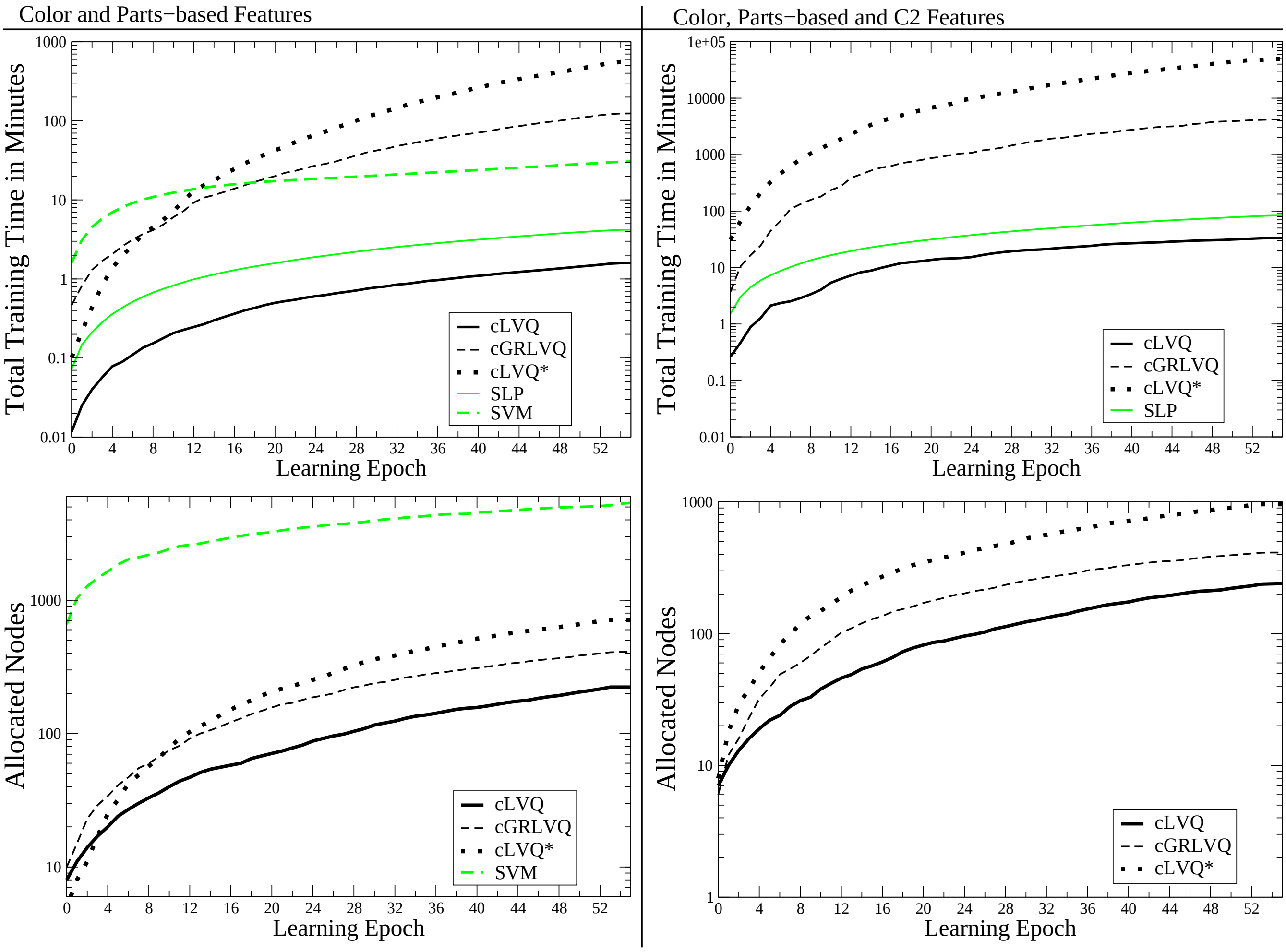

Figure 8: **Comparison of Network Resources.** The total training time of the cLVQ, cGRLVQ, cLVQ*, SVM and SLP is most crucial with respect to interactive learning shown at the top of this figure for both feature sets. It can be seen that especially in later learning epochs the simpler vector quantization methods cGRLVQ and cLVQ* require more than two orders of magnitudes more training time compared to SLP and cLVQ. The SVM is faster than cGRLVQ and cLVQ* for the experiments with color and parts-based features, but it is intractable for the addition of high-dimensional C2 features. Finally the cLVQ is even two times faster than the simple SLP. This computational efficiency of the cLVQ method is caused by the small number of selected features, but also by the smallest amount of allocated representatives as shown at the bottom of this figure. On the contrary the SVM approach has the highest memory requirement that is undesirable for life-long learning tasks.

small amount of selected category-specific and stable features. The smaller number of nodes again enhances the learning speed but also is beneficial with respect to the representational capacity for storing many different categories.

4.5. Qualitative Evaluation of the cLVQ Feature Selection Method

Apart from the categorization performance and network resources we are also interested in how good the feature selection method of our proposed cLVQ learning algorithm is able to find correct category-specific features. Therefore ten different training runs of the cLVQ method were performed and all selected features for each category are saved together with the corresponding feature scoring values. The selected features for each category c are sorted based on the total number of occurrence in these ten runs, where frequent features are most probably critical for the representation of this particular category. Additionally each feature is visualized with a small patch, to allow a visual inspection of its usefulness for the corresponding category. We use the RGB value of the histogram bin center for color features, while for the parts-based features the grey-value patch corresponding to the highest detector activity is chosen. This highest detector activity was calculated based on the training images used for the selection of the part-based features that do not correspond to the training and test set used for the category learning shown in Fig. 5. We also consider the final scoring value h_{cf} of each selected feature. This value is identical for all learning runs and provides information about the category specificity of this feature. The results of this investigation are shown in Fig. 9 for three representative color categories and in Fig. 10 for three shape categories.

Due to the fact that the training objects are presented iteratively to the cLVQ, its wrapper feature selection method can never be perfect. A certain feature at a particular learning state might be useful, but with more experience it can become obsolete. This especially occurs for the first object presentation of a shape category, where often a color feature is selected, because due to the object rotation it is more stable than all shape features. As a consequence features that are selected only once in Fig. 9 and Fig. 10 are most probably not category-specific and in many cases unrelated to the most exemplars of the category. But such erroneous features often also have low scoring values, so that the impact of these features for the category representation is minimized. Interestingly, the number of features selected once and also their total number positively correlates with the categorization performance. Therefore both numbers indicate the difficulty of each category. Furthermore the categorization performance over different runs is more stable if the set of different selected features is small. In contrast to this a larger number of selected features which occurred 3-6 times during the different runs, indicate that several redundant feature sets with roughly the same representational power exist.

It is somehow surprising with respect to the difficulty of categories that the color categories are not in general easier compared to shape categories. This is especially visible for the category "white" shown in Fig. 9 and the category "cup" illustrated in Fig. 10. Although in all runs the correct histogram bin for white was selected, the corresponding scoring value of this feature is quite low. This small scoring value is most probably caused by reflections on glossy objects, because such spots typically cause activations of this histogram bin that are independent of the actual color of the object. Additionally "white" is the only color category for which only few training objects are completely white but many of them contain smaller fractions of white. Therefore for this category the separation from other co-occurring shape and color categories becomes more difficult. Finally it should be mentioned that among the most frequently reoccurring features a considerable amount have relatively small scoring values, even if some features with higher scoring values are available. This effect is best visible for the category "animal" in Fig. 10. This can occur if features with higher scoring values are rarely activated and thus are rejected because the measured performance gain is below the feature insertion threshold. Additionally it is probable that at least for the shape categories the combination of several features is important, so that a single feature might be

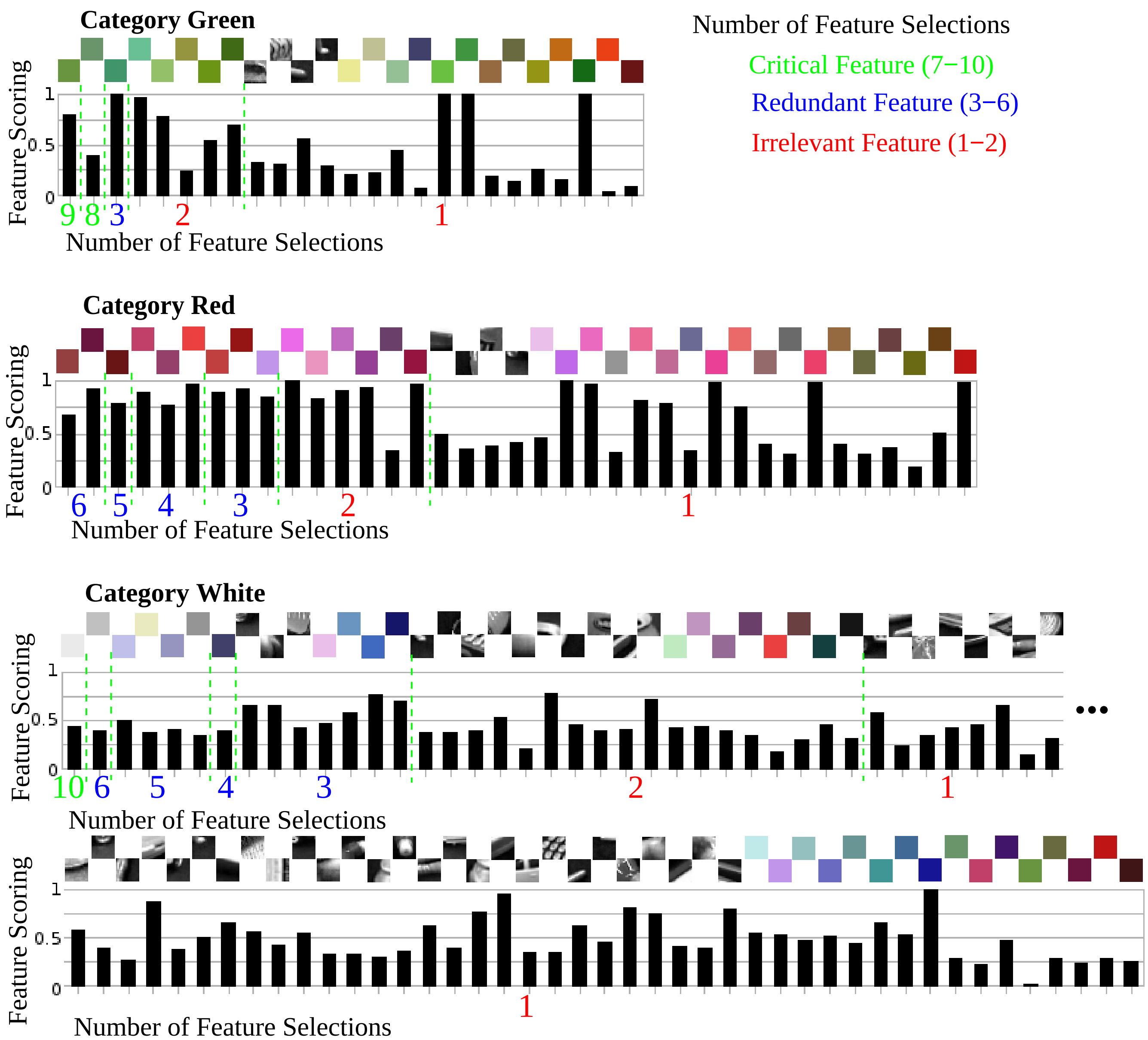

Figure 9: **Evaluation of the Feature Selection Method for Color Categories.** Illustration of the selected features of three representative color categories, where one easy, one average and one difficult category was selected. For this visualization ten different cLVQ networks are trained and the selected features of each category together with the scoring values are saved. The selected features of each category are sorted and color-coded based on the total number of occurrences in these ten runs, while the bar height correspond to the feature score of these selected features. All features that occurred at least 7 times (green) are considered as critical for the representation of this category, while feature occurrence of less than 3 time (red) are probably irrelevant or even wrong. Finally features that are selected 3-6 times (blue) indicate redundant feature sets, with similar representational capacity. Beside the occurrence of each feature the total number of selected features indicate the difficulty of the category. This is especially visible for the worst color category "white". Nevertheless even for this category the correct color feature is selected in all test runs.

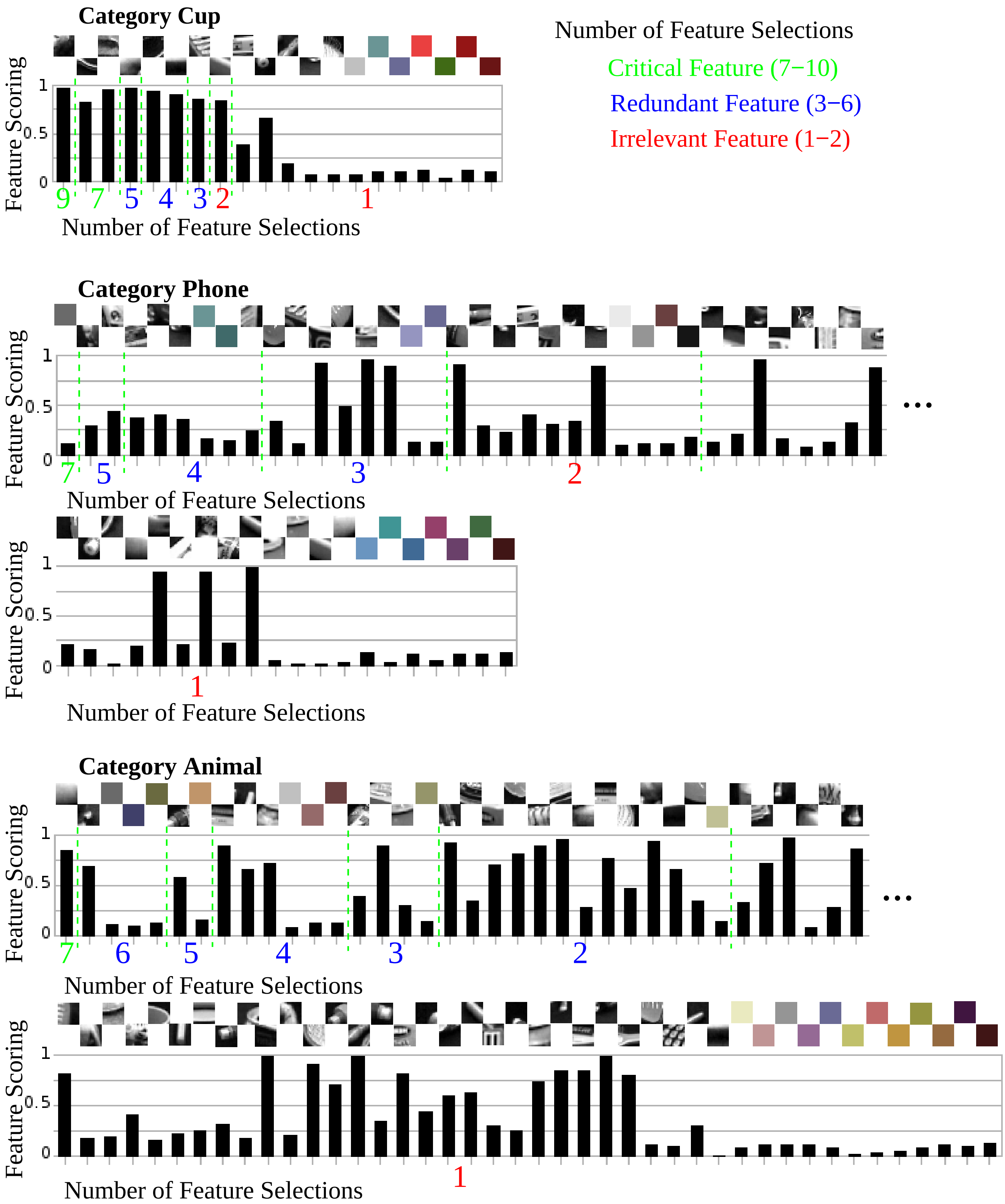

Figure 10: **Evaluation of the Feature Selection Method for Shape Categories**. Illustration of the selected features of three representative shape categories, analogously to Fig. 9. Surprisingly not all shape categories are more difficult than the color categories. This becomes clear if one compares the category "cup" with "white" depicted in Fig. 9. Furthermore many stable selected features have low scoring values, which indicate only little category specificity. This indicates that for shape categories only the combination of several features allow a stable category representation.

general and less category-specific, but in combination with other features allows a robust category detection.

5. Discussion

We have proposed an architecture for fast interactive life-long learning of arbitrary categories that is able to perform an incremental allocation of cLVQ nodes, automatic feature selection and feature weighting. This automatic control of the architecture complexity is crucial for interactive and life-long learning, where an exhaustive parameter search is not feasible. Additionally we use the proposed wrapper method for incremental feature selection, because the representation of categories should use as few feature dimensions as possible. This can not be achieved with simple filter methods, where typically only a small amount of redundant or noisy features are eliminated. The used feature selection method enables the cLVQ to separate co-occurring categories and allows a resource efficient representation of categories, which is beneficial for fast interactive and incremental learning of categories. Recently a variant of an embedded feature selection method for LVQ networks was proposed by Kietzmann et al. (2008) based on the GRLVQ method (Hammer & Villmann, 2002) which was called iGRLVQ. This method iteratively removes features with small weighting values λ. For our categorization task this proposed backward feature selection method is not suitable because a low λ value at a certain learning epoch does not imply that this feature can not become useful at a later learning stage. Unfortunately removed features can not be readded to the corresponding iGRLVQ network at a later learning stage, especially if the reduction of computational costs is targeted. Additionally the definition of a stopping condition for the feature pruning is difficult to determine a priori, so that Kietzmann et al. (2008) prespecified the final feature dimensionality. Finally the required computational costs are considerably higher compared to forward feature selection methods. 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.

In contrast to many other categorization approaches our model is able to learn multiple categories at once, while commonly the categories are trained individually (Fritz et al., 2005; Fei-Fei et al., 2007). We applied our learning method to a challenging categorization task, where the objects are rotated around the vertical axis. This rotation causes much higher appearance changes compared to many other approaches dealing with canonical views only (Leibe et al., 2004). In contrast to this our exemplar-based method can deal with a larger within-category variation, which we consider crucial for complex categories. Furthermore we recently could show that our proposed cLVQ learning method can be integrated into a larger vision system that allows online learning of categories based on hand-held and complex-shaped objects under full rotation (Kirstein et al., 2008, 2009). This means our cLVQ approach does not only scale well to higher feature dimensionalities, but also to more complex categorization tasks in unconstrained environments.

Acknowledgment: The authors thank Stephan Hasler for providing the visualization for the parts-based features.

References

Agarwal, S., Awan, A., & Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Transaction Pattern Analysis and Machine Intelligence 26(11), 1475–1490.

Arsenio, A. M. (2004). Developmental learning on a humanoid robot. In Proc. International Joint Conference on Neuronal Networks (IJCNN), pp. 3167–3172.

Bagnall, R. G. (1990). Lifelong education: The institutionalisation of an illiberal and regressive ideology? Educational Philosophy and Theory 22(1), 1–7.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transaction on Neural Networks 3(5), 698–712.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. Machine Learning 20(3), 273–297.

Fei-Fei, L., Fergus, R., & Perona, P. (2003). A Bayesian approach to unsupervised one-shot learning of object categories. In Proc. International Conference on Computer Vision (ICCV), pp. 1134–1141.

Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. Computer Vission and Image Understanding 106(1), 59–70.

Fergus, R., Perona, P., & Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In Proc. Computer Vision and Patern Recognition (CVPR), Volume 2, pp. 264–271.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. Journal of Machine Learning Research, 3, 1289–1305.

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

Fritz, M. (2008). Modeling, Representation and Learning of Visual Categories. Ph. D. thesis, Technical University of Darmstadt.

Fritz, M., Kruijff, G.-J. M., & Schiele, B. (2007). Cross-modal learning of visual categories using different levels of supervision. In Proc. International Conference on Vision Systems (ICVS).

Fritz, M., Leibe, B., Caputo, B., & Schiele, B. (2005). Integrating representative and discriminative models for object category detection. In Proc. International Conference on Computer Vision (ICCV), Volume 2, pp. 1363–1370.

Fritzke, B. (1995). A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), Advances in Neural Information Processing Systems 7, Cambridge MA, pp. 625–632. MIT Press.

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

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16(10), 906–914.

Guyon, I. & Elissee, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182.

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.

Harris, C. & Stephens, M. (1988). A combined corner and edge detector. In Proc. Alvey Vision Conference, pp. 147–151.

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

Heisele, B., Serre, T., Pontil, M., Vetter, T., & Poggio, T. (2001). Categorization by learning and combining object parts. In Proc. Advances in Neural Information Processing Systems (NIPS), pp. 1239–1245.

Kadir, T. & Brady, M. (2001). Saliency, scale and image description. International Journal of Computer Vision 45(2), 83–105.

Kietzmann, T. C., Lange, S., & Riedmiller, M. (2008). Incremental GRLVQ: Learning relevant features for 3D object recognition. Neurocomputing 71(13–15), 2868–2879.

Kira, K. & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In Proc. Association for the Advancement of Artificial Intelligence (AAAI), pp. 129–134.

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., Gross, H.-M., & Körner, E. (2008). An integrated system for incremental learning of multiple visual categories. In Proc. International Conference on Neural Information Processing (ICONIP), pp. 811–818. Springer.

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

Kohavi, R. & John, G. (1997). Wrappers for feature subset selection. Artificial Intelligence, 97, 273–324.

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

Leibe, B., Leonardis, A., & Schiele, B. (2004). Combined object categorization and segmentation with an implicit shape model. In ECCV workshop on statistical learning in computer vision, pp. 17–32.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110.

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

McCloskey, M. & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. Psychology of Learning and Motivation, 24, 109–164.

Mikolajczyk, K., Leibe, B., & Schiele, B. (2006). Multiple object class detection with a generative model. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Ng, A. Y. & Jordan, M. I. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In Proc. Advances in Neural Information Processing Systems (NIPS), pp. 849–856.

Opelt, A., Fussenegger, M., Pinz, A., & Auer, P. (2004). Weak hypotheses and boosting for generic object detection and recognition. In Proc. European Conference on Computer Vision (ECCV), Volume 2, pp. 71–84.

Ozawa, S., Toh, S. L., Abe, S., Pang, S., & Kasabov, N. (2005). Incremental learning of feature space and classifier for face recognition. Neural Networks 18(5-6), 575–584.

Perkins, S., Lacker, K., & Theiler, J. (2003). Grafting: Fast, incremental feature selection by gradient descent in function space. Journal of Machine Learning Research, 3, 1333–1356.

Polikar, R., Udpa, L., Udpa, S., & Honavar, V. (2001). Learn++: An incremental learning algorithm for supervised neural networks. IEEE Transactions on System, Man and Cybernetics (C) 31(4), 497–508.

Roth, P. M., Donoser, M., & Bischof, H. (2006). On-line learning of unknown hand held objects via tracking. In Proc. Second International Cognitive Vision Workshop (ICVW).

Schapire, R. E. (1990). The strength of weak learnability. Machine Learning 5(2), 197–227.

Skočaj, D., Berginc, G., Ridge, B., Štimec, A., Jogan, M., Vanek, O., Leonardis, A., Hutter, M., & Hewes, N. (2007). A system for continuous learning of visual concepts. In Proc. International Conferance on Vision Systems (ICVS).

Skočaj, D., Kristan, M., & Leonardis, A. (2008). Continuous learning of simple visual concepts using incremental kernel density estimation. In Proc. International Conference on Computer Vision Theory and Applications (VISAPP), Funchal, Madeira, Portugal, pp. 598–604.

Steels, L. & Kaplan, F. (2001). AIBO's first words. The social learning of language and meaning. Evolution of Communication 4(1), 3–32.

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

Thomas, A., Ferrari, V., Leibe, B., Tuytelaars, T., Schiele, B., & Gool, L. V. (2006, June). Towards multi-view object class detection. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, USA.

Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Proc. IEEE Conference on Computer Vision and Pattern Recogntion (CVPR), pp. 511–518.

Wersing, H., Kirstein, S., Götting, M., Brandl, H., Dunn, M., Mikhailova, I., Goerick, C., Steil, J., Ritter, H., & Körner, E. (2007). Online learning of objects in a biologically motivated architecture. International Journal of Neural Systems, 17, 219–230.

Wersing, H. & Körner, E. (2003). Learning optimized features for hierarchical models of invariant object recognition. Neural Computation 15(7), 1559–1588.

Willamowski, J., Arregui, D., Csurka, G., Dance, C. R., & Fan, L. (2004). Categorizing nine visual classes using local appearance descriptors. In Proc. ICPR Workshop on Learning for Adaptable Visual Systems.

Yang, Y. & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In Proc. International Conference on Machine Learning, pp. 412–4120.

A. Pseudocode Notation of the cLVQ Approach

Initialize $S_c = \emptyset$ for $\forall c$

1. Update training set $T = T \backslash T^{oldest} \cup T^{new}$
2. Update $h_{cf} = \frac{H_{cf}}{H_{cf}+\bar{H}_{cf}}$ with (Φ as Heaviside function):
   $H_{cf} := H_{cf} + \sum_{i \in T^{new}} \Phi(t_c^i) * \Phi(x_f^i)$ and
   $\bar{H}_{cf} := \bar{H}_{cf} + \sum_{i \in T^{new}} \Phi(-t_c^i) * \Phi(x_f^i)$ //(see Eq.5,6,7)
3. Initialize each new category c
   if $S_c = \emptyset$ and $\exists\, t_c^i = 1$
   - $v_c = \arg\max_f(h_{cf})$ //feature with highest scoring value
   - $j = \arg\max_{i \in T}(x_{v_c}^i)$ //$\mathbf{x}^j$, where feature has highest activity
   - $K = K + 1$ and $\mathbf{w}^K = \mathbf{x}^j$; $u_c^K = 1$ else $\mathbf{u}_{S \neq c}^K = 0$ //insert node
   - $S_c := S_c \cup \{v_c\}$ //add feature
4. Setup selection lists
   - Initialize $P_c = \{1, ..., F\} \backslash S_c$ //list of insertable features
   - Initialize $Q_c = S_c$ //list of removable features
5. cLVQ optimization loop (see Fig.2)
   - LVQ node update and error counting
     - for all $i \in T$
       * if $k_{\min}(c)$ exists, update $\mathbf{w}^{k_{min}(c)}$ //(see Eq.2,3)
       * Compute E_c^+ and E_c^- //(see Eq.8,9)
   - for all c with $\#E_c^+ \cup \#E_c^- = \emptyset$ //add or remove feature
     - if $\#E_c^+ \geq \#E_c^-$ //higher amount of detection errors
       * Compute e_{cf}^+ //(see. Eq.10)
       * Choose $v_c = \arg\max_{f \in P_c}(h_{cf} + e_{cf}^+)$
       * $S_c := S_c \cup \{v_c\}$ //add feature
       * Store $E_c^{pre} = \#E_c^+ + \#E_c^-$
       * Recompute $E_c^{post} = \#E_c^+ + \#E_c^-$ with added v_c
       * if $E_c^{pre} - E_c^{post} > \epsilon^1$ //add feature permanently
         · Keep v_c; $P_c = P_c \backslash \{v_c\}$
       * else //remove and exclude feature
         $S_c = S_c \backslash \{v_c\}$; $P_c = P_c \backslash \{v_c\}$
     - else //higher amount of rejection errors
       * Compute $e_{cf}^- = \sum_{i \in E_c^-} \Phi(x_f^i) / \sum_{i \in E_c^-} 1$
       * Choose $v_c = \arg\max_{f \in Q_c}(h_{cf} + e_{cf}^-)$
       * $S_c := S_c \backslash \{v_c\}$ //remove feature
       * Store E_c^{pre} and compute E_c^{post}
       * if $E_c^{pre} - E_c^{post} \leq \epsilon^1$ //readd and exclude feature
         $S_c := S_c \cup \{v_c\}$; $Q_c = Q_c \backslash \{v_c\}$
   - Add new nodes
     - Initialize $Z = \{c | E_c^+ \cup E_c^- \neq \emptyset\}$ //erroneous c list

- $K^0 = K$ //number of nodes before insertion step
- for all $i \in T$ //collect errors per $\mathbf{x}^i$
  - $F^i = \{c | i \in E_c^+ \cup E_c^- \neq \emptyset\}$
- while $Z \neq \emptyset$ //select vector with most errors, where at least for one category no node was inserted so far
  - $j = \arg\max_{\{i|F^i \cap Z \neq \emptyset\}} \#F^i$
  - $K = K + 1$; $\mathbf{w}^K = \mathbf{x}^j$; //insert node
  - $\mathbf{u}^K = \begin{cases} t_c^j & : \quad j \in E_c^+ \cup E_c^- \\ 0 & : \quad \text{else} \end{cases}$. //set target vector
  - $Z_j = \{c | t_c^j \neq 0\}$; $Z := Z \backslash Z_j$
- Store E_c^{pre} and compute E_c^{post}
- for all c

  if $E_c^{pre} - E_c^{post} \leq \epsilon^2$

  Set $\mathbf{u}_c^k = 0 \ \forall k > K^0$
- if $u_c^k = 0$ for $\forall c$

  Remove node k //node does not contribute to any category

- Stop condition //no errors or no insertable features left
  - if $E_c^+ \cup E_c^- = \emptyset$ or $P_c = \emptyset \, \forall c$

    goto Step 1 //start new learning epoch
  - else

    goto Step 5 //test new features and nodes