Learning hierarchical representations of object categories for robot vision

Aleš Leonardis and Sanja Fidler

Faculty of Computer and Information Science
University of Ljubljana, Slovenia
{ales.leonardis,sanja.fidler}@fri.uni-lj.si

Summary. This paper presents our recently developed approach to constructing a hierarchical representation of visual input that aims to enable recognition and detection of a large number of object categories. Inspired by the principles of efficient indexing, robust matching, and ideas of compositionality, our approach *learns* a hierarchy of spatially flexible compositions, i.e. parts, in an unsupervised, statistics-driven manner. Starting with simple, frequent features, we learn the statistically most significant compositions (*parts composed of parts*), which consequently define the next layer. Parts are learned sequentially, layer after layer, optimally adjusting to the visual data. Lower layers are learned in a category-independent way to obtain complex, yet sharable visual building blocks, which is a crucial step towards a scalable representation. Higher layers of the hierarchy, on the other hand, are constructed by using specific categories, achieving a category representation with a small number of highly generalizable parts that gained their structural flexibility through composition within the hierarchy. Built in this way, new categories can be efficiently and continuously added to the system by adding a small number of parts only in the higher layers. The approach is demonstrated on a large collection of images and a variety of object categories.

1 Introduction

For an efficient interaction with the real-world environment it is of central importance to equip cognitive robots with a general vision system capable of recognizing a large number of objects (scenes, actions) and their generic classes [1, 17, 29]. The success of such a general system relies critically on building efficient, compact (sharable), generalizable and robust representations that can be *learned* quickly and incrementally within a continuous interaction with the world.

Visual categorization and recognition of objects have been a subject of extensive research in computer vision over the last decades. Many approaches have been developed that perform well on this challenging task [22, 24, 20, 21, 16, 15, 7, 28, 2, 18, 9]. However, most of them operate in a limited domain - the number of classes that can be recognized with the available computational resources is generally relatively low.

It is thus crucial to devote significant efforts to develop an appropriate representation of visual data (and other modalities) that would enable categoriza-

tion/recognition/detection on a larger scale. We envision the following principles to guide the design of a general vision system:

Computational plausibility. To overcome the curse of large-scale recognition, a need for *hierarchical compositional representations* has emerged [6, 3, 19]. This is also consistent with the findings on biological systems [27, 5]. A hierarchy of parts composed of parts that could limit the computationally prohibitive visual search by means of indexing and matching in each individual layer would enable an efficient way to store and retrieve information.

Statistics-driven learning. Parts and their higher level combinations should be learned in an unsupervised manner (at least in the lower layers of the hierarchy) in order to avoid hand-labeling of a large number of images as well as to capture the regularities within the visual data as effectively and compactly as possible [4, 8]. Moreover, there are strong implications that the human visual system is driven by these principles as well [11].

Robust recognition. To achieve robustness against noise and clutter, the parts comprising the individual hierarchical layers should be manifested as models to enable a robust verification of the presence of their underlying components [3]. Models should incorporate loose geometric relations to achieve the spatial binding of features [8, 19], yet at the same time encode enough flexibility to gain discrimination gradually - through composition within the hierarchy.

Accurate localization. The representation should be able to provide accurate localization of objects to enable a robot to successfully interact with the environment. This represents a problem for hierarchical systems (such as neural networks), which usually grow exponentially with the number of layers and thus necessitate a progressive reduction in resolution in order to cope with the computational load. However, this problem can be alleviated within a compositional system, where each hierarchical unit can be accurately traced back to the image by following the subcomponents from the layers below [10].

Fast, incremental learning. Learning novel categories should be fast with its efficiency increasing with the amount of visual data already "seen" by the system. To achieve such a learning capacity, the design of lower levels within the hierarchy is crucial in order to obtain the features optimally shared by various object categories. Once the *visual building blocks* are learned, learning of novel objects can proceed mainly in the higher layers and can thus continue fast and with no or minimal human supervision. In addition, the system has to be capable of online learning without the inefficient restructuring of the complete hierarchy.

The current state-of-the-art categorization methods predominantly build their representations on image patches [20, 28] or other highly discriminative features such as SIFT [26]. Since the probability of occurrence of such features is very small, masses of them need to be extracted to represent objects reasonably well. This results in computationally highly inefficient recognition, which demands matching of a large number of image features to an enormous number of prototypical ones. This drawback has been alleviated within the most recent methods that employ hierarchical clustering in a high dimensional feature space, yet the resulting

representations still demand at least a linear search through the library of stored objects/categories [20, 26].

However, a majority of hierarchical methods perform matching of *all* prototypical units against *all* features found in an image. Mutch et al [21] (and their predecessor [25]) employ matching of all $4,000$ higher-layer templates against features extracted in each pixel and scale of the resampled pyramid. This is also a drawback in layers of clustered histograms [2] and the hierarchical convolutional network [16].

On the other hand, the success of hierarchical methods that do employ the principles of indexing and matching has been hindered by the use of hand-coded information. In [3], the authors use hand-crafted local edge features and only learn their global arrangements pertaining to specific object categories. The authors of [23] use predesigned filters and process the visual information in the feed-forward manner, while their recent version [25] replaced the intermediate layer with random combinations of local edge arrangements rather than choosing the features in accordance with the natural statistics.

Approaches that do build the layers by learning and are able to make a sufficient number of them (by starting with simple features) mostly design the parts by histogramming the local neighborhoods of parts of the previous layers [2] or by learning the neural weights based on the responses on the previous layers [16, 13]. Besides lacking the means of indexing, additional inherent limitation of such methods is the inefficiency in performing incremental learning; as novel categories arrive, the whole hierarchy has to be re-adapted. Moreover, histograms do not enable robust top-down matching, while convolutional networks would have problems with the objects or features that are supersets/subsets of other features.

While the concepts of hierarchical representations, indexing and matching, statistical learning and incrementality have already been explored in the literature, to the best of our knowledge, they have not been part of a unifying framework. This paper presents our recently developed approach to building a hierarchical representation that aims to enable recognition and detection of a large number of object categories. Inspired by the principles of efficient indexing, robust matching, and ideas of *compositionality*, our approach *learns* a hierarchy of spatially flexible compositions, i.e. parts, in a completely unsupervised, statistics-driven manner. As the proposed architecture does not yet perform large-scale recognition, it makes important steps towards scalable representations of visual categories.

The learning algorithm proposed in [12], which acquires a hierarchy of local edge arrangements by correlation, is in concept similar to our learning method. However, the approach demands registered training images, employs the use of a fixed grid, and is more concerned with the coarse-to-fine search of a particular category (i.e. faces) rather than finding features shared by many object classes.

The paper is organized as follows: in Section 2 we provide the motivation and the general theory behind the approach. The results obtained on various image data sets are shown in Section 3. The paper concludes with a summary in Section 4.

2 Designing a Hierarchical Compositional Representation

The design of a hierarchical representation presented in this paper aims to support *all* the requirements set in the Introduction. We start with a careful definition of parts (hierarchical units) in Subsec. 2.1 that enable an efficient indexing and robust matching in an interplay of layered information. This can be attained by the principles of composition [14], i.e. *parts composed of parts*, allowing for a representation that is dense (highly sharable) in the lower layers and gets significantly sparser (category specific) in the higher layers of the hierarchy. In agreement with such a definition of parts, principles of indexing and matching are roughly illustrated in Subsec. 2.2.

However, the main issue in compositional/hierarchical systems is how to recover the "building blocks" automatically by means of *learning*. Subsec. 2.3 presents our recently developed learning algorithm, which extracts the parts in an unsupervised way in the lower layers and with minimal supervision in the higher layers of the hierarchy.

The complete hierarchical learning architecture is illustrated in Figure 1.

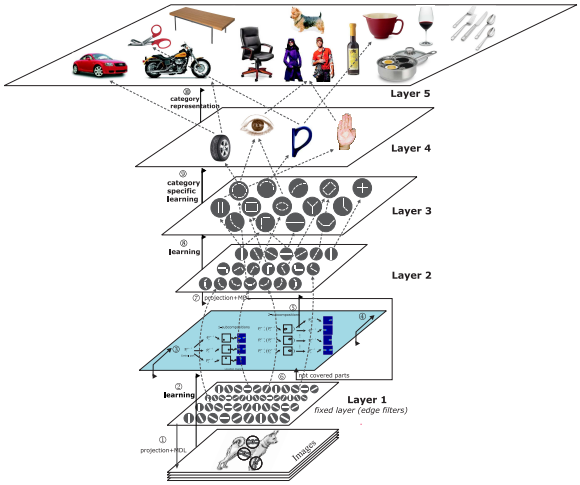

Fig. 1. Hierarchical compositional representation of object categories

2.1 Definition of parts

In accordance with the postulates given in the Introduction, each unit in each hierarchical layer is envisioned as a *composition* defined in terms of spatially flexible local arrangements of units from the previous layers. We shall refer to such composite models as *parts*. However, a clear distinction must be made between the parts within the hierarchy that will serve as a *library* of stored prototypes and the parts found in a particular image being processed. This Subsection gives the definition of the library parts, while the part *realizations* in images are explained in Subsec. 2.2.

Let $\mathcal{L}_n$ denote the n-th Layer. We define the parts recursively in the following way. Each part in $\mathcal{L}_n$ is characterized by the identity, $\mathcal{P}_i^n$ (which is an internal index/label within the library of parts), the center of mass, orientation, and a list of subparts (parts of the previous layer) with their respective orientations and positions relative to the center and orientation of $\mathcal{P}_i^n$. One subpart is the so-called *central part* that indexes into $\mathcal{P}_i^n$ from the lower, $(n-1)$th layer. Specifically, a $\mathcal{P}_i^n$, which is normalized to the orientation of 0 degrees and has a center in $(0,0)$, encompasses a list $\{(\mathcal{P}_j^{n-1}, \alpha_j, (x_j, y_j), (\boldsymbol{\sigma}_{1j}, \boldsymbol{\sigma}_{2j}))\}_j$, where α_j and (x_j, y_j) denote the relative orientation and position of $\mathcal{P}_j^{n-1}$, respectively, while $\boldsymbol{\sigma}_{1j}$ and $\boldsymbol{\sigma}_{2j}$ denote the principal axes of an elliptical Gaussian encoding the variance of its position around (x_j, y_j).

The hierarchy starts with a fixed $\mathcal{L}_1$ composed of local oriented filters that are simple, fire densely on objects, and can thus be efficiently combined into larger units. The employed filter bank comprises eight odd Gabor filters whose orientations are spaced apart by 45°. We must emphasize, however, that all the properties of parts comprising layers higher than 1 will be *learned*.

2.2 Detection of parts in images

For any given image, the process starts by describing the image in terms of local oriented edges. This is done *on every scale* – each rescaled version of the original image (a Gaussian pyramid with two scales per octave) is processed separately. First, each image in the pyramid is filtered by 11×11 Gabor filters. By extracting local maxima of the Gabor energy function that are above a low threshold, an image (on each scale) is transformed into a list of $\mathcal{L}_1$ parts. Each higher level interpretation is then found by an interplay of indexing (evoking part hypotheses) and matching (verifying parts). Performed in this way, the top-down mechanism is extremely robust to noise and clutter.

The indexing and matching procedure is illustrated in Figure 2.a), while detection (localization) is outlined in Figure 2.b).

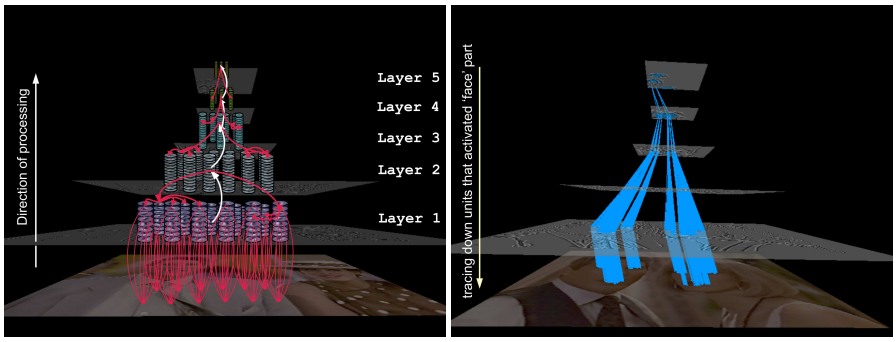

Fig. 2. a) Hierarchical indexing (evoking part hypotheses) and matching (verifying parts), b) Tracing parts back to the image by following activated subcomponents from the layers below.

2.3 Learning the hierarchy of parts

This Subsection briefly presents an approach to learning parts in successive layers in the hierarchy. The details of the learning algorithm can be found in [10]. We first introduce the necessary steps that need to be performed prior to learning and propose an algorithm that learns the higher layer compositions of parts taking into account their spatial relations.

In order to exploit the statistical redundancy present in the visual data as effectively as possible, layers are built sequentially; only after $\mathcal{L}_{n-1}$ has been obtained, learning of the $\mathcal{L}_n$ can proceed. We must emphasize that parts can, however, be added to each of the layers later on.

Learning starts on the basis of a fixed $\mathcal{L}_1$ composed of oriented Gabor filters. Each image is thus transformed into a set of parts, encoding their location, orientation, and identity. From here on, the algorithm is general, thus we describe how learning of $\mathcal{L}_n$ is performed once $\mathcal{L}_{n-1}$ has already been obtained.

For clarity, let us denote the already learned parts (parts already added to the hierarchy — starting with a set of oriented filters) with $\mathcal{P}^{n-1}$, and the set of possible compositions with $\mathcal{C}^n$.

We propose the following steps for learning the parts: **1.)** reduction in spatial resolution to alleviate the processing time and to avoid over-learning of local neighborhoods already inspected within $\mathcal{L}_{n-1}$, **2.)** an automatic selection of the optimal neighborhood size within which compositions will proceed to be learned, **3.)** learning of compositions by sequential increase in complexity by keeping statistics of combinations with the so-called *spatial maps*, **4.)** selection of most frequent and stable compositions by keeping the number of indexing links from the previous layer low, and **5.)** grouping perceptually similar parts by projection and statistics in the original, *image layer*. The learning process is incremental: the compositions obtained are projected on images and the steps $3-5$ repeated on image parts that are not described by the selected compositions. This is done until either no more significant compositions are found or the number of indexing links reaches the computationally set maximum.

Learning of compositions

Due to the prohibitive number of possible local configurations, learning proceeds by determining statistically significant *subcompositions* with an increasing number of the subparts contained.

To achieve shift invariance of parts, we choose a part-centered coordinate system, that is, each local neighborhood is defined relatively to a certain part (which is hereafter referred to as the *central part*). Define the $s-$subcomposition as a composition modeling s subparts in addition to the central one. Learning proceeds by increasing s, starting with $s = 1$.

$1-$**subcompositions.** For the $1-$subcompositions, spatial configurations of one part conditioned on the identity of the central one are sought out. A list of possible compositions with the so-called *spatial maps* are formed: $\{\mathcal{C}^n_{s=1}\} = \{\mathcal{P}^{n-1}_i, \{\mathcal{P}^{n-1}_j, map_j\}\}$, where $\mathcal{P}^{n-1}_i$ denotes the central part and $\mathcal{P}^{n-1}_j$ the additional subpart, whose spatial position relative to the central one will be stored in map_j.

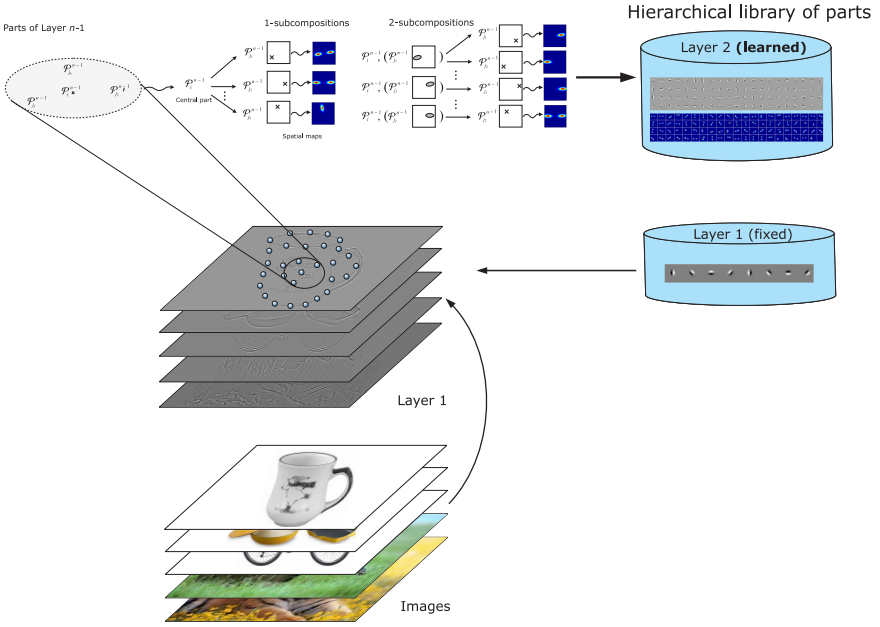

Fig. 3. Learning of compositions by sequentially increasing the number of subparts.

With the set of prepared subcompositions, learning proceeds in local neighborhoods of all $\mathcal{L}_{n-1}$ parts found in images. After performing local inhibition, spatial maps of the remaining subparts contained in $\mathcal{C}_1^n$ are updated accordingly.

Spatial maps thus model the spatial distribution of P_j^{n-1} conditioned on the presence of P_i^{n-1} in the center of the neighborhood. The sum of its elements is the number of "votes" for the combination. After all images are processed, we detect voting peaks in the learned spatial maps, and for each peak, a spatial area that captures most of the votes is formed (modeled by an elliptical gaussian having principal axes $(\boldsymbol{\sigma}_{1j}, \boldsymbol{\sigma}_{2j})$). This area consequently represents the spatial variability of the part P_j^{n-1} within the composition $\mathcal{C}_1^n$. The sum of votes in the area of variability divided by the number of all inspected neighborhoods defines the probability of occurrence of the subcomposition.

Amongst all acquired $1-$subcompositions, we employ a selection process which discards some of the learned compositions or passes them to the next stage, at which more complex compositions are formed.

$s-$**subcompositions.** For a general s-subcomposition, configurations consisting of $s+1$ subparts altogether are built on the basis of those having s subparts. When the construction of s-subcompositions commences, empty spatial maps for possible combinations $\{\mathcal{C}_s^n\} = \{\mathcal{P}_i^{n-1}, \{\mathcal{P}_{j_m}^{n-1}, (x_{j_m}, y_{j_m}), (\boldsymbol{\sigma}_{1j_m}, \boldsymbol{\sigma}_{2j_m})\}_{m=1}^{s-1}, \{\mathcal{P}_j^{n-1}, map_j\}\}$, where the first s terms denote the central part and the learned $s-1$ subparts, are prepared. As the local neighborhoods are inspected, map_j is updated whenever all subparts forming a certain composition are found in the local image

neighborhood (after performing local inhibition). The learning procedure is illustrated in Figure 3. The selection procedure is similar to the one described previously for $1-$subcompositions.

The overall number of votes for individual parts decreases as their complexity increases. When no derived composition passes the set threshold, the layer learning process concludes.

2.4 Representation of object categories

After extensive evaluation of the proposed scheme and many statistical insights gained on large collections of images (the most important results are presented in Section 3) we propose the representation of object categories be built in the following way.

Learning sharable parts in a category-independent way can only get so far - the overall statistical significance drops, while the number of parts reaches its critical value for learning. Thus, learning of higher layers proceeds only on a subset of parts - the ones that are the most repeatable in a specific category. Specifically, we build:

Category-independent lower layers. Learning the lower layers should be performed on a collection of images containing a variety of object categories in order to find the most frequent, sharable parts.

Category-specific higher layers. Learning of higher layers should proceed in images of specific categories. The final, categorical layer then combines the parts through the object center to form the representation of a category.

Since the learning process is incremental, categories can be efficiently added to the representation by adding a small number of parts only in the higher layers.

3 Results

We applied our method to a collection of $3,200$ images containing 15 diverse categories (cars, faces, mugs, dogs, etc.). The learned parts of $\mathcal{L}_2$ and $\mathcal{L}_3$ are presented in the first row of Figure 4.

As proposed in Subsect. 2.4, the learning of $\mathcal{L}_4$ was performed only on images containing faces. The obtained parts were then learned relative to centers of faces to produce $\mathcal{L}_5$ - *category layer* (parts are presented in Figure 4, rows 4 and 5). Cars and mugs were then incrementally added to our representation. The fourth and fifth row of Figure 4 show the learned Layers, while Figure 5 depicts the learned compositionality within the hierarchical library for faces, cars and mugs. The detections of some of the learned parts are presented in Figure 6.

Clearly, only a small number of $\mathcal{L}_4$ parts are needed to represent an individual category. Since the proposed hierarchical representation would computationally handle $10-20$ times the number of $\mathcal{L}_3$ parts in Layer $\mathcal{L}_4$ (in the order of $5,000-10,000$ parts), a large number of categories could potentially be represented in this way.

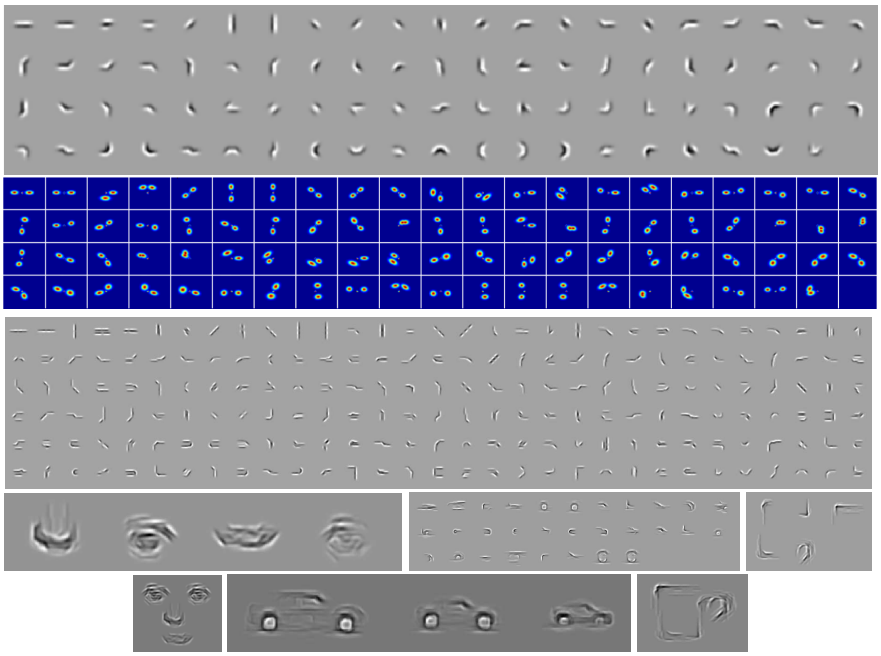

Fig. 4. Mean reconstructions of the learned parts. **1st row and 3rd row**: $\mathcal{L}_2$, $\mathcal{L}_3$, respectively (the first 186 of all 499 parts are shown), **2nd row**: Learned spatial flexibility modeled in $\mathcal{L}_2$ parts, **4th row**: $\mathcal{L}_4$ parts for faces, cars, and mugs, **5th row**: $\mathcal{L}_5$ parts for faces, cars (obtained on 3 different scales), and mugs.

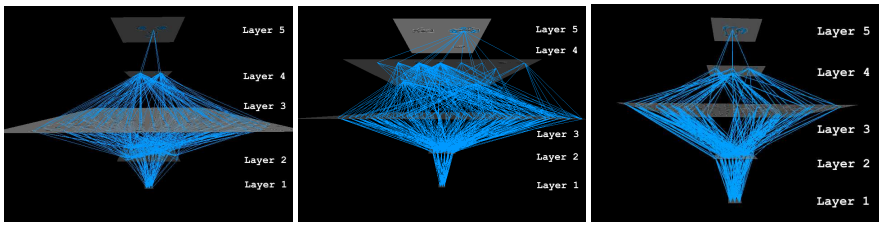

Fig. 5. Learned compositionality for faces, cars and mugs (Layers $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ are the same for all categories).

4 Summary and conclusions

This paper presents our recently developed approach to building a representation of object categories. The method learns a hierarchy of flexible compositions in an unsupervised manner in lower, category-independent layers, while requiring minimal supervision to learn higher, categorical layers.

There are numerous advantages of the proposed representation over the traditional hierarchical representations, such as neural networks. By coding hierarchical units in terms of presence/absence of a small number of components from the layer

below, the representation is realized as a very sparsely connected network and recognition can thus be implemented within an efficient and robust indexing and matching scheme. While the number of units in hierarchies usually grow exponentially with the number of layers, a properly designed indexable representation has an approximately constant computational complexity for matching all levels of the hierarchy. Moreover, by allowing for spatial flexibility of subparts based on the statistics of natural images, the otherwise prohibitive combinatorics of combination can be successfully controlled.

The recognition of compositions is additionally much less sensitive to local image variations and degradations than methods that rely on matching of high-dimensional feature vectors. Another advantage of compositionality is that while the spatial resolution has to be reduced in the higher hierarchical layers, the activated units can be traced back to the original image by following their active subcomponents. This is essential for robotic applications where accurate localization and segmentation are a crucial prerequisite to executing grasping or manipulation tasks.

Furthermore, the design of the hierarchical units is incremental, where new categories can be continuously added to the hierarchy. The results show that only a small number of higher layer parts are needed to represent individual categories, therefore the proposed scheme would allow for an efficient representation of a large number of visual categories.

The representation, however, currently only supports 2D inference. Incorporating 3D (depth) and motion information is the topic of our ongoing work.

Acknowledgement

This research has been supported in part by: Research program Computer Vision P2-0214 (RS), EU FP6-004250-IP project CoSy, EU MRTNCT- 2004-005439 project VISIONTRAIN, and EU FP6-511051 project MOBVIS. The authors would like to thank Luka Čehovin and Miha Drenik for developing the graphical interface.

References

1. Cognitive Systems for Cognitive Assistants (CoSy). *EU FP6-004250-IP IST Cognitive Systems Integrated project*, 2004-2008.
2. A. Agarwal and B. Triggs. Hyperfeatures - multilevel local coding for visual recognition. In *ECCV (1)*, pages 30–43, 2006.
3. Y. Amit and D. Geman. A computational model for visual selection. *Neural Comp.*, 11(7):1691–1715, 1999.
4. H. B. Barlow. Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, 30:1561–1571, 1990.
5. S. Brincat and C. Connor. Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, 49(1):17–24, 2006.
6. A. Califano and R. Mohan. Multidimensional indexing for recognizing visual shapes. *PAMI*, 16(4):373–392, 1994.

7. D. J. Crandall and D. P. Huttenlocher. Weakly supervised learning of part-based spatial models for visual object recognition. In *ECCV (1)*, pages 16–29, 2006.
8. S. Edelman and N. Intrator. Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, 27:73–110, 2003.
9. R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR(2)*, pages 264–271, 2003.
10. S. Fidler and A. Leonardis. Towards scalable representations of visual categories: Learning a hierarchy of parts. In *CVPR 2007*.
11. J. Fiser and R. N. Aslin. Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci U S A*, 99(24):15822–15826, 2002.
12. F. Fleuret and D. Geman. Coarse-to-fine face detection. *IJCV*, 41(1/2):85–107, 2001.
13. K. Fukushima, S. Miyake, and T. Ito. Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE SMC*, 13(3):826–834, 1983.
14. S. Geman, D. Potter, and Z. Chi. Composition systems. *Quarterly of App. Math.*, Vol. 60(Nb. 4):707–736, 2002.
15. K. Grauman and T. Darrell. Efficient image matching with distributions of local invariant features. In *CVPR (2)*, pages 627–634, 2005.
16. F.-J. Huang and Y. LeCun. Large-scale learning with svm and convolutional nets for generic object categorization. In *CVPR*, pages 284–291, 2006.
17. L. Jamone, G. Metta, F. Nori, and G. Sandini. James: A humanoid robot acting over an unstructured world. In *6th IEEE-RAS International Conference on Humanoid Robots*, pages 143–150, 2006.
18. B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In *ECCV'04, SLCV Workshop*, 2004.
19. B. W. Mel and J. Fiser. Minimizing binding errors using learned conjunctive features. *Neural Computation*, 12(4):731–762, 2000.
20. K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. In *CVPR06*, pages 26–36.
21. J. Mutch and D. G. Lowe. Multiclass object recognition with sparse, localized features. In *CVPR06*, pages 11–18, 2006.
22. A. Opelt, A. Pinz, and A. Zisserman. A boundary-fragment-model for object detection. In *ECCV (2)*, pages 575–588, 2006.
23. M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neurosc.*, 2(11):1019–1025, Nov. 1999.
24. S. Savarese, J. Winn, and A. Criminisi. Discriminative object class models of appearance and shape by correlatons. In *CVPR (2)*, pages 2033– 2040, 2006.
25. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Object recognition with cortex-like mechanisms. *PAMI*, 29(3):411–426, 2007.
26. E. Sudderth, A. Torralba, W. Freeman, and A. Willsky. Learning hierarchical models of scenes, objects, and parts. In *ICCV*, pages 1331–1338, 2005.
27. K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, (4):832–838, 2001.
28. S. Ullman and B. Epshtein. *Visual Classification by a Hierarchy of Extended Features.* Towards Category-Level Object Recognition. Springer-Verlag, 2006.
29. K. Welke, E. Oztop, A. Ude, R. Dillmann, and G. Cheng. Learning feature representations for an object recognition system. In *6th IEEE-RAS International Conference on Humanoid Robots*, pages 290–295, 2006.

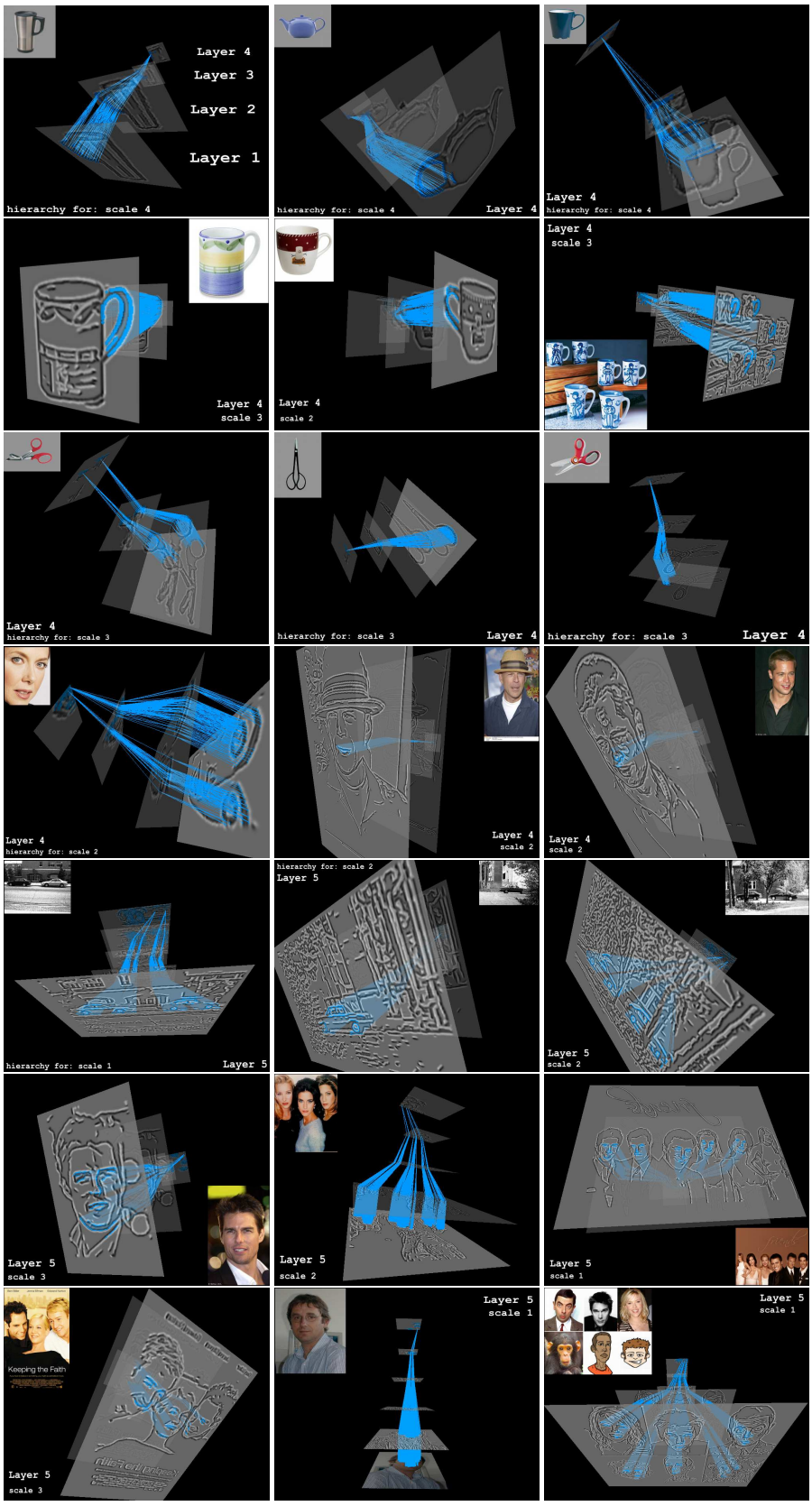

Fig. 6. Detection of some of the $\mathcal{L}_4$ and $\mathcal{L}_5$ parts. Detection proceeds bottom-up as described in Subsec. 2.2. Active parts in the top layer are traced down to the image through activated subparts.