Vision in Human and Machine

Part 6
Object Recognition Models

Heiko Wersing

Honda Research Institute Europe GmbH

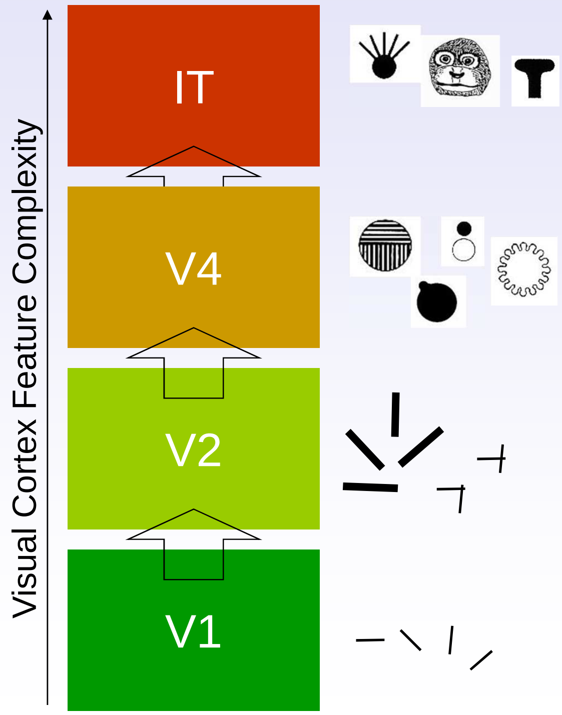
The selectivity-stability dilemma of recognition

- Stability: Invariance and robustness

- Selectivity: Classification and categorization

Biological background: The ventral path in the visual cortex

Visual cortex area | Represented Features



Faces, pre-object patterns, objects

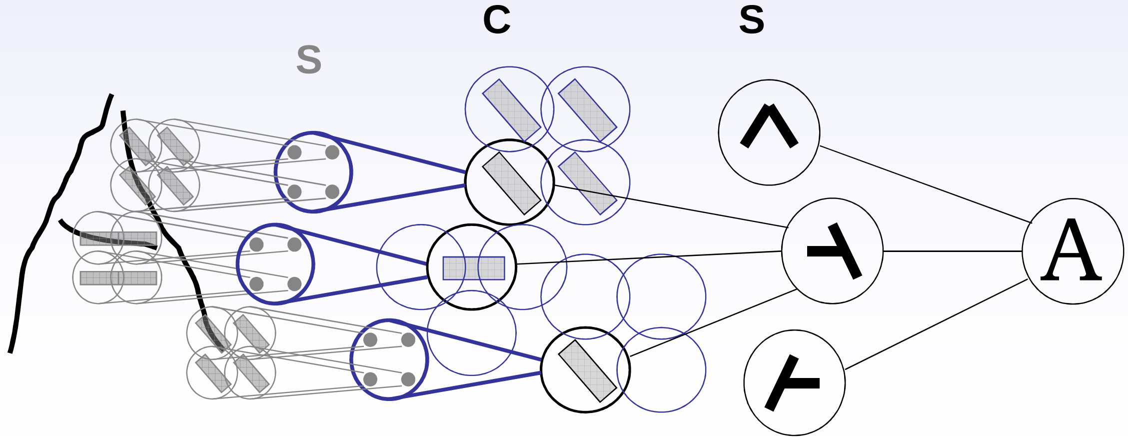
„elaborate cells“ (Tanaka)
more complex patterns

Simple features with larger
receptive fields, End-stopped cells

Simple Features: local, oriented,
Bandpass in spatial frequency,

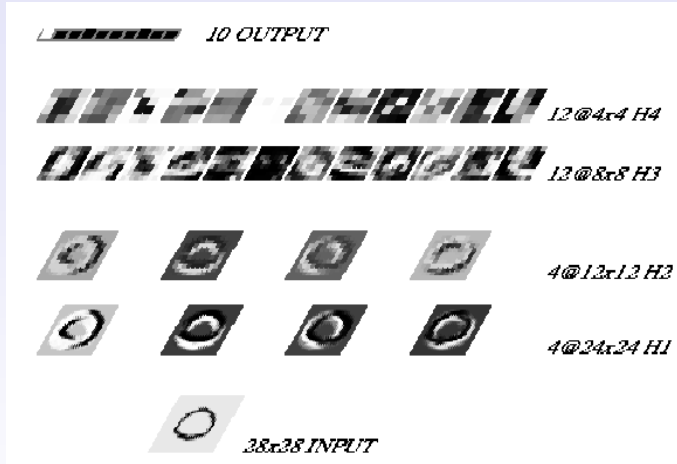
Hierarchical Recognition Models

- Fukushima 1980
  - The Neocognitron: Competetive feature learning
  - Feature detection (S) and pooling (C) stages



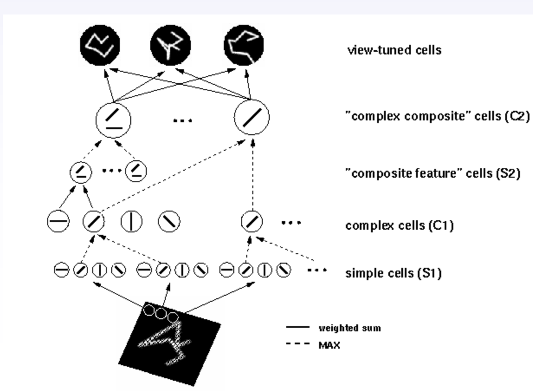
Hierarchical Recognition Models

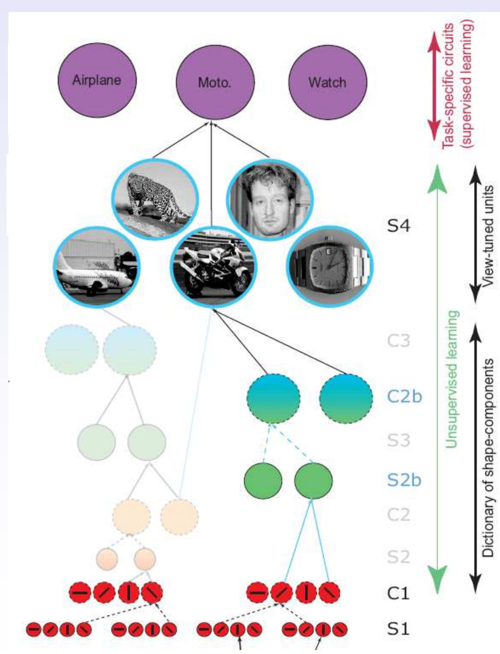
- LeCun et al. 1989
  - Backpropagation learning for handwritten Zip code recognition
  - Features are trained fully supervised
  - Multilayer perceptron



Hierarchical Recognition Models

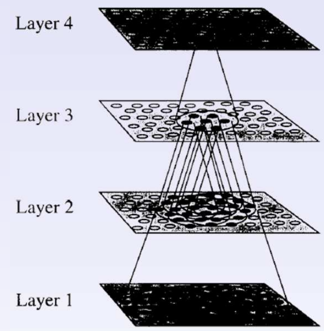
- Poggio & Edelman 1990
  - View-based representations using RBF nets
- Riesenhuber & Poggio 1999
  - Hierarchy with MAX pooling
- Serre & Poggio et al. 2003
  - Extension, comparison to biology, Exemplar-based feature learning



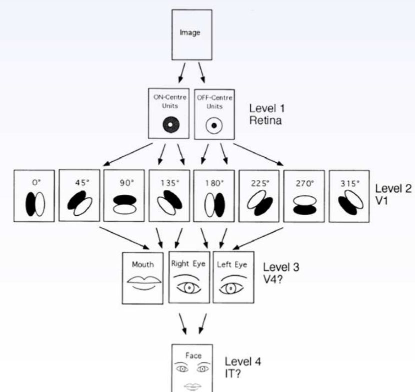


Hierarchical Recognition Models

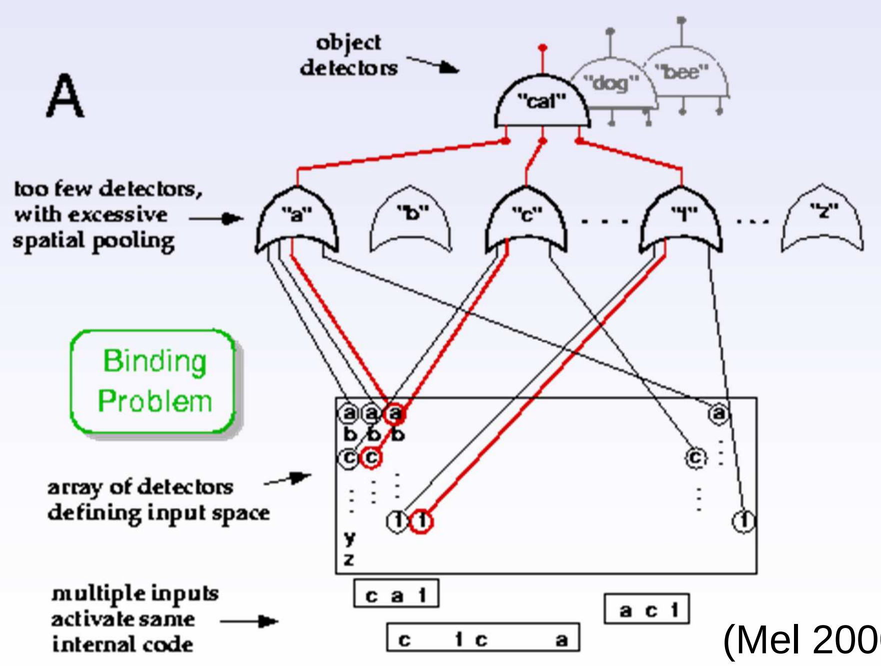
- Wallis & Rolls 1997
  - VISNET, Trace rules for feature learning (Földiak 1991)
  - Temporal unsupervised feature extraction
  - Minimize temporal variation of the feature detector



- Thorpe et al. 1998
  - SpikeNET: Temporal rank-order coding in a hierarchy
  - Strongest activated features fire first
  - Feature reponse depends on the order if incoming spikes

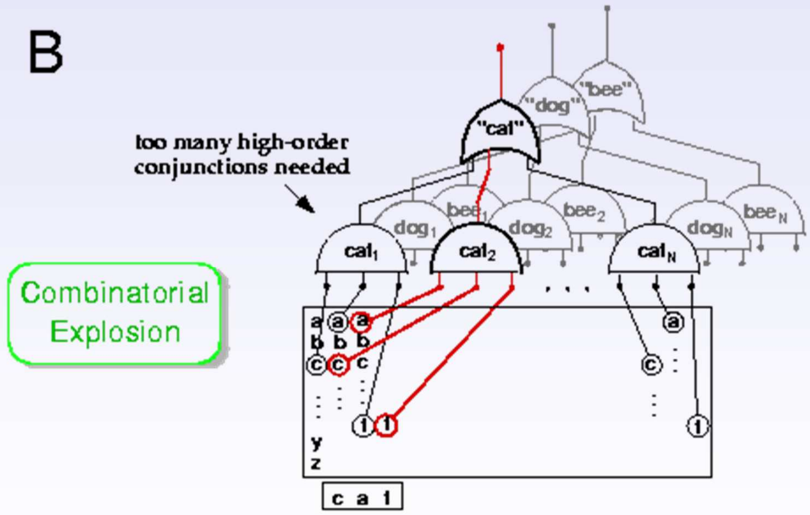


Feedforward Processing Architectures



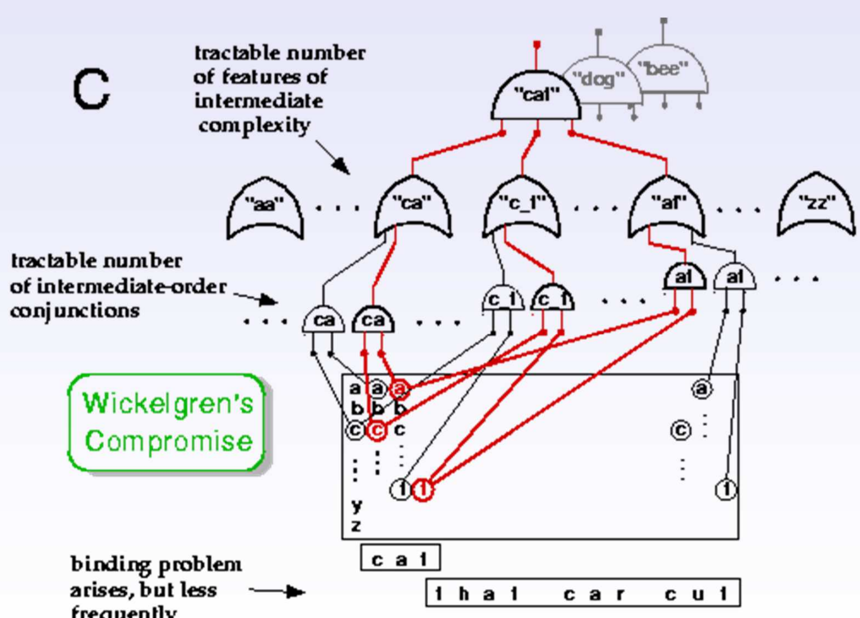
(Mel 2000)

Feedforward Processing Architectures



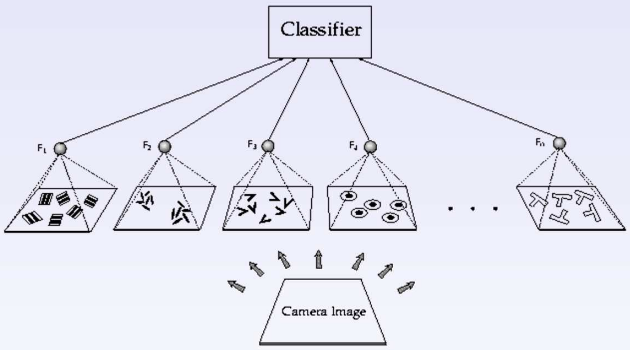
(Mel 2000)

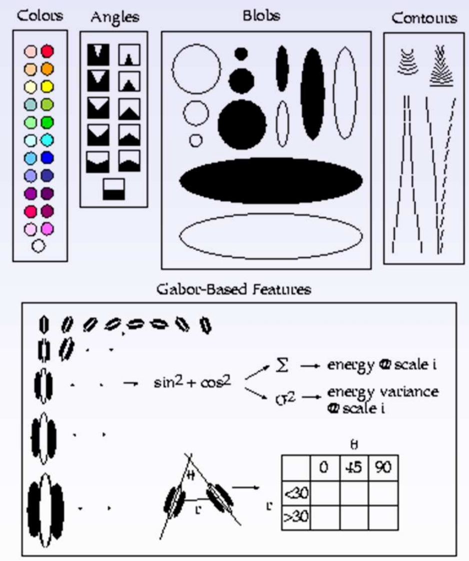
Feedforward Processing Architectures



(Mel 2000)

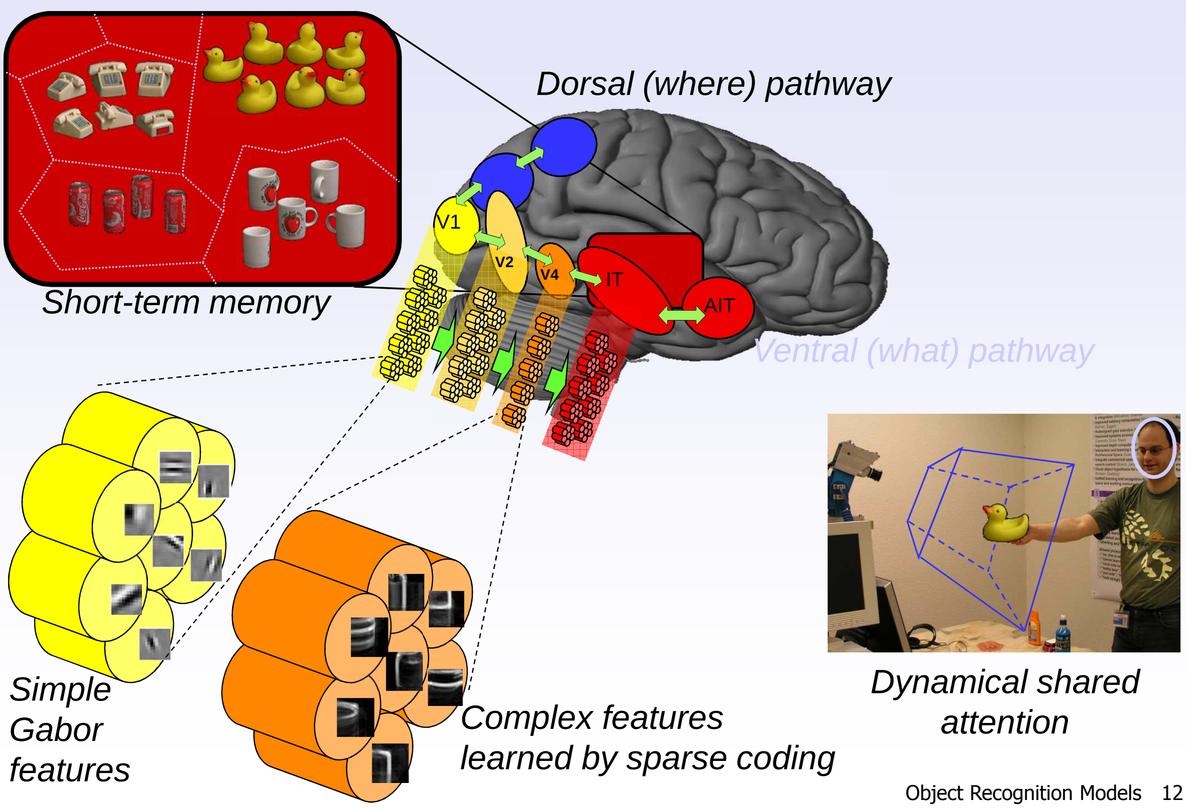
A Flat Recognition System



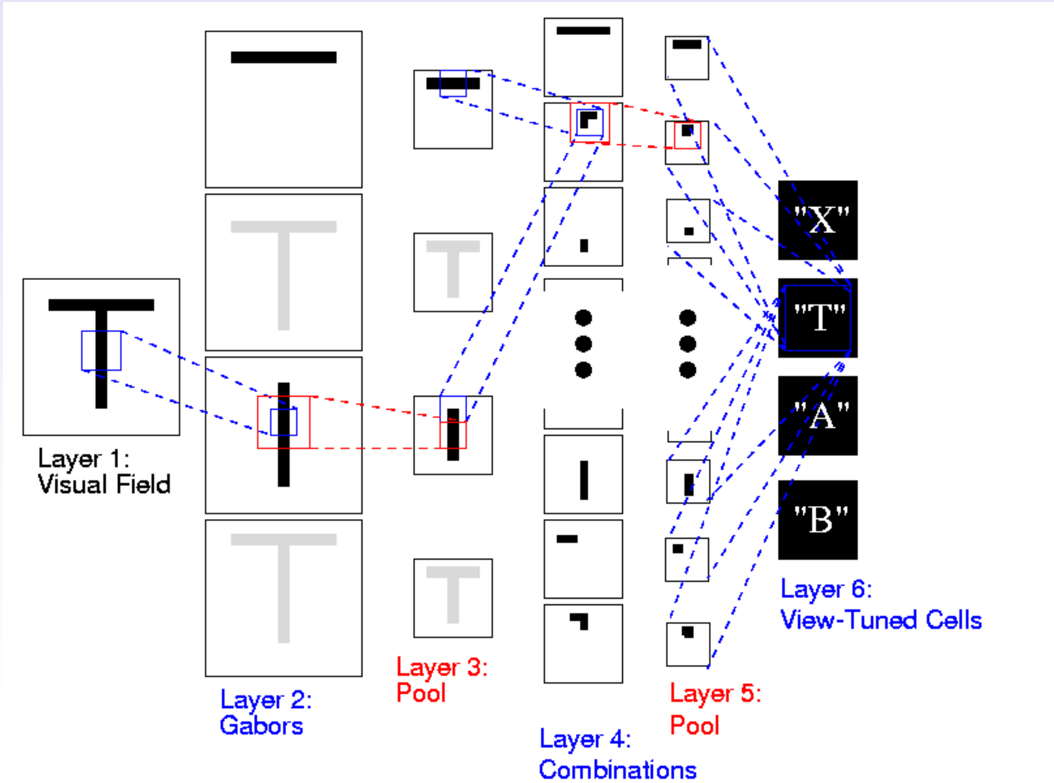


- SEEMORE (Mel 1997)
- 102 Features with full spatial pooling
- Good recognition performance
- Strong color-dependence
- Pre-segmentation is crucial

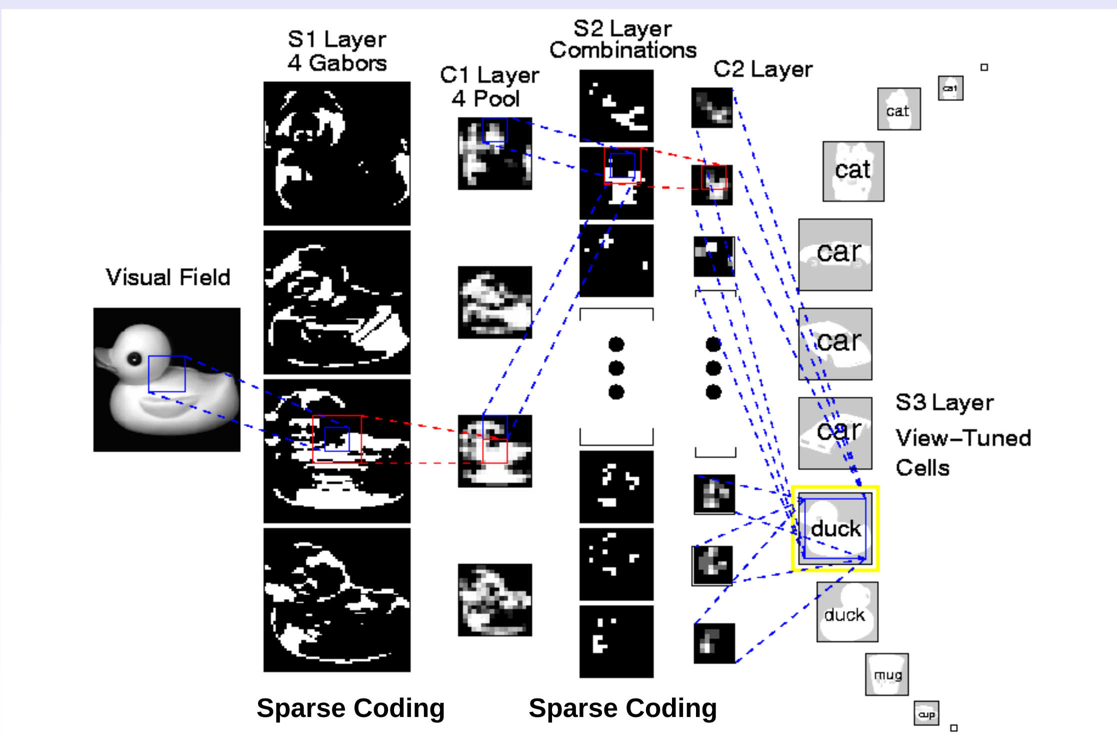
Our biologically motivated approach



Conceptual approach sparse coding hierarchy (Wersing & Koerner 2003)



A Hierarchical Sparse Coding Recognition Architecture

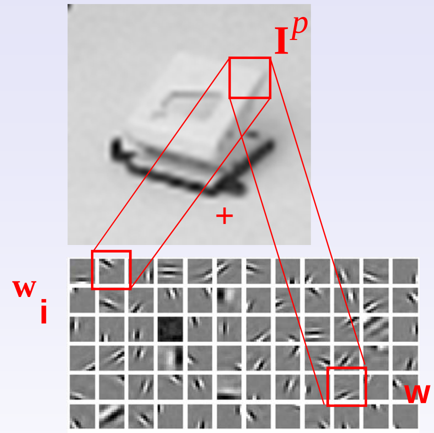


(Wersing & Körner, Neural Computation 2003)

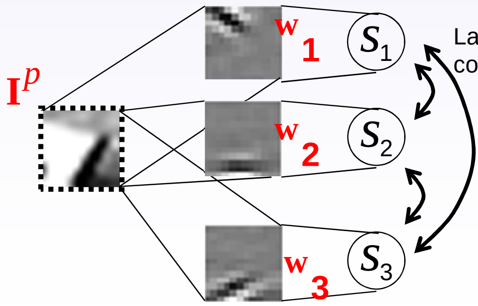
Sparse coding for feature learning

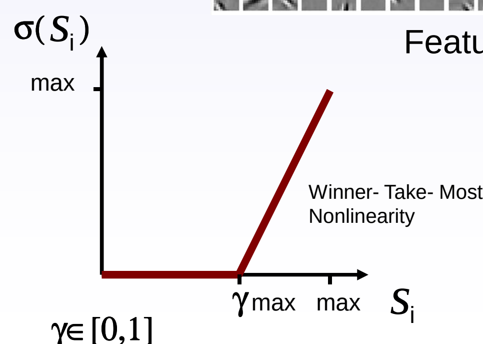
Sparse coding for learning of Wavelet-like basis similar to V1 simple cell receptive fields (Olshausen&Field 1996)

$$E_1 = \underbrace{\frac{1}{2}\sum_p ||\mathbf{I}^p - \sum_i s_i^p \mathbf{w}_i||^2}_{\text{reconstruction}} + \underbrace{\sum_p \sum_i \Phi(s_i^p)}_{\text{sparsity}},$$



Features

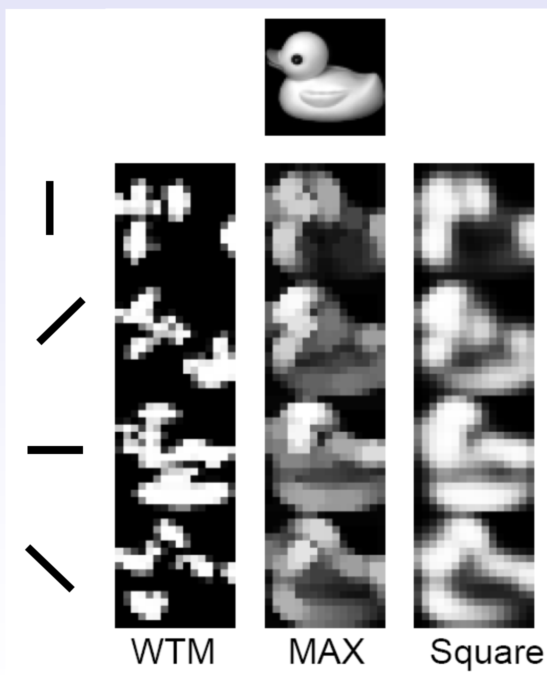




$\gamma \in [0,1]$

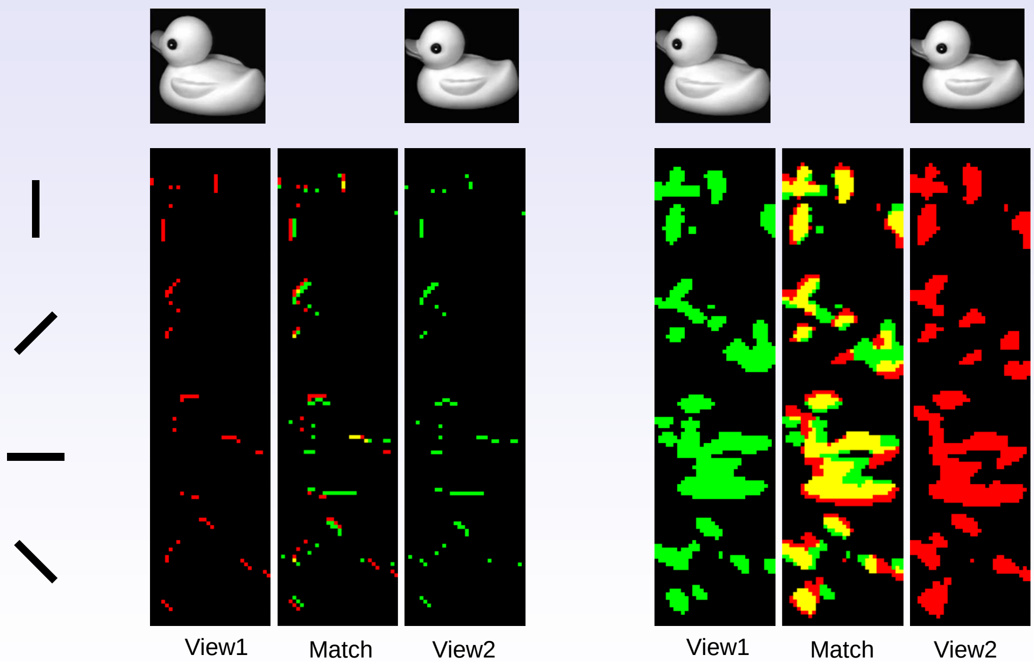
Resulting nonlinearity

- Winner-take-most local lateral competition with spatial „OR"
  - → Model of latency-based transmission of best-matching feature detections (Körner, Gewaltig, Körner, Richter, & Rodemann , 1999)



Comparison of nonlinearities for Gabor-based edge detection

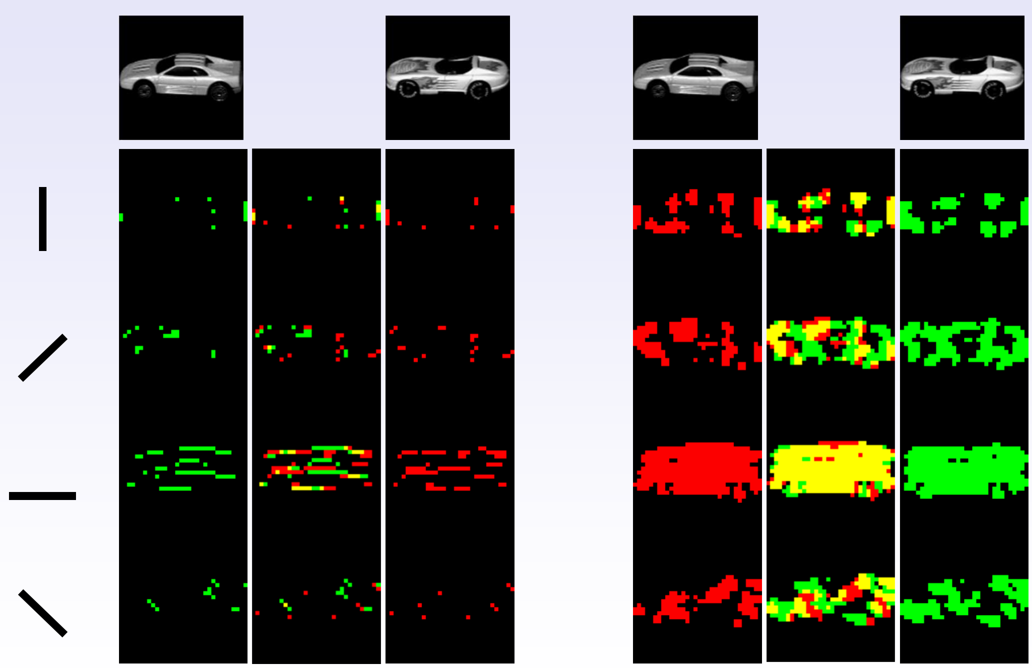
Feature Selectivity – Stability Dilemma



High Selectivity
Low Stability

Low Selectivity
High Stability

Feature Selectivity – Stability Dilemma



High Selectivity
Low Stability

Low Selectivity
High Stability

Invariant Sparse Coding

(Olshausen&Field 1996)

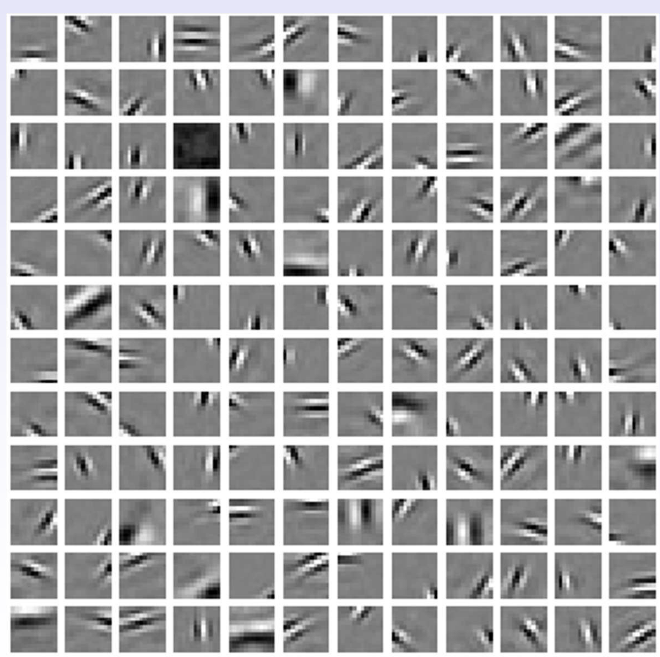
$$E_1 = \underbrace{\frac{1}{2}\sum_p ||\mathbf{I}^p - \sum_i s_i^p \mathbf{w}_i||^2}_{\text{reconstruction}} + \underbrace{\sum_p \sum_i \Phi(s_i^p)}_{\text{sparsity}},$$

Sparse Invariant coding for removing transformation redundancy
(Wersing & Körner, Neural Computation, 2003)

$$\boxed{E_2 = \frac{1}{2}\sum_p ||\mathbf{I}^p - \sum_i \sum_m s_{im}^p \mathbf{T}^m \mathbf{w}_i||^2 + \sum_p \sum_i \sum_m \Phi(s_{im}^p),}$$

Problem: Redundanz der Merkmale

Aufgrund von Invarianzen in der Eingabe erhält man transformierte Versionen „desselben“ Merkmals

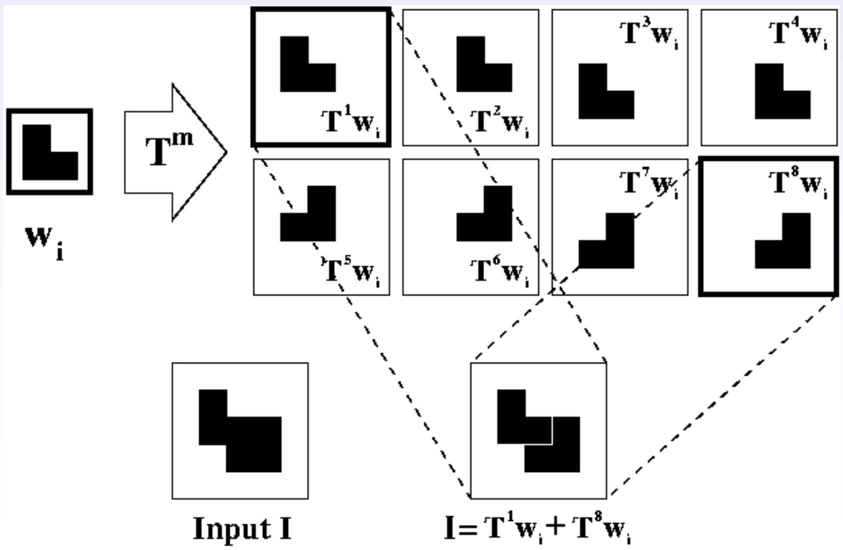


Idee: Benutze Transformationen zur Vermeidung von Redundanz

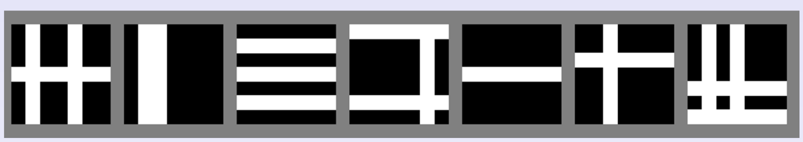
Invariant Sparse Coding (Wersing & Eggert 2002)

$$E_2 = \frac{1}{2}\sum_p ||\mathbf{I}^p - \sum_i \sum_m s^p_{im}\mathbf{T}^m\mathbf{w}_i||^2 + \sum_p \sum_i \sum_m \Phi(s^p_{im}),$$

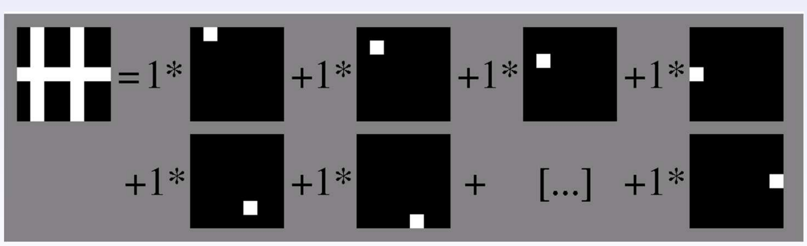
T^m: Invarianztransformation



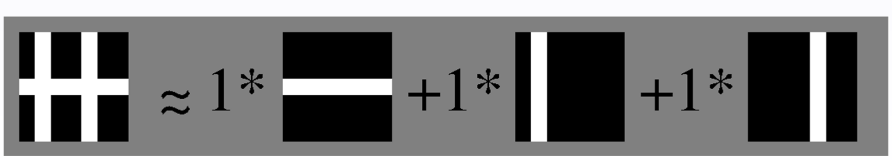
Sparse invariant coding: Toy example



Ensemble of 2000 example patterns: 1-3 Random Stripes

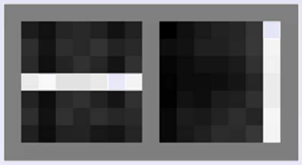


Non-Sparse Basis Representation



Sparse Basis Representation

Learned representations with translation invariance



b) 2 basis vectors

c) 4 basis vectors

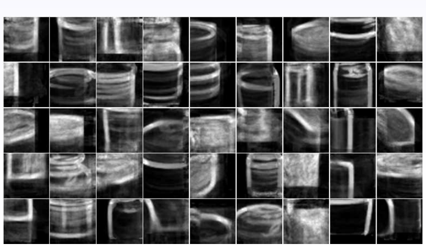
d) 8 basis vectors

Results of Sparse Invariant Coding Learning Algorithm

Result of Principal Component Analysis – 20 Basis vectors

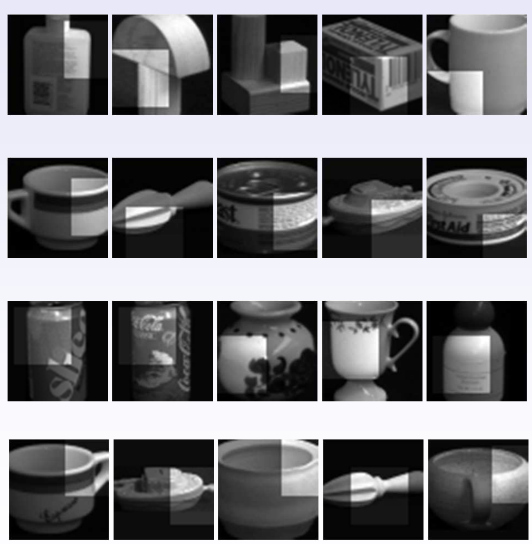
Sparse Invariant Features of Objects

COIL 100 object database

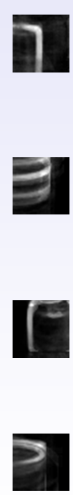


Sparse Invariant Features

Strong Feature Responses

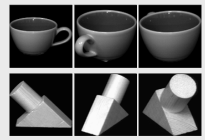


Sensitive C2 Feature

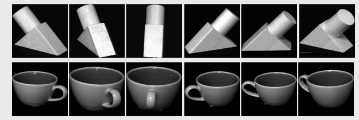


Comparison on benchmark data COIL 100

Example:
2 Objects



Training data

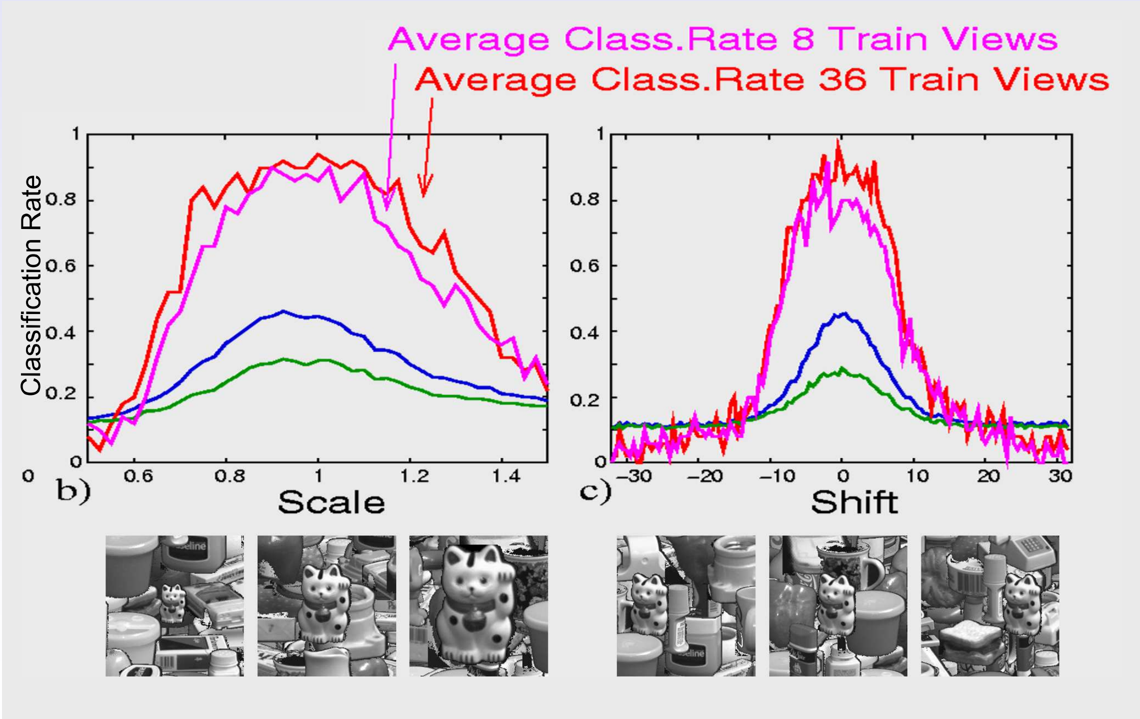


Test data (6 out of 24)

	Method	30 Objects Training views 36	8	2	4 Training views Number objects 10	30	100
Simple image match	NNC	100	96.3	70.5	86.5	81.8	70.1
Eigenobjects	Columbia	100	92.5	67.1	92.1	84.6	77.0
Support Vector Machine	SVM	100	95.6	71.0	91.0	84.9	74.6
Visual Hierarchy online learn	tpl−VTU	100	95.6	77.6	93.5	89.7	79.0
Visual Hierarchy with add. Opt.	opt−VTU	100	95.7	80.1	94.9	89.9	79.1

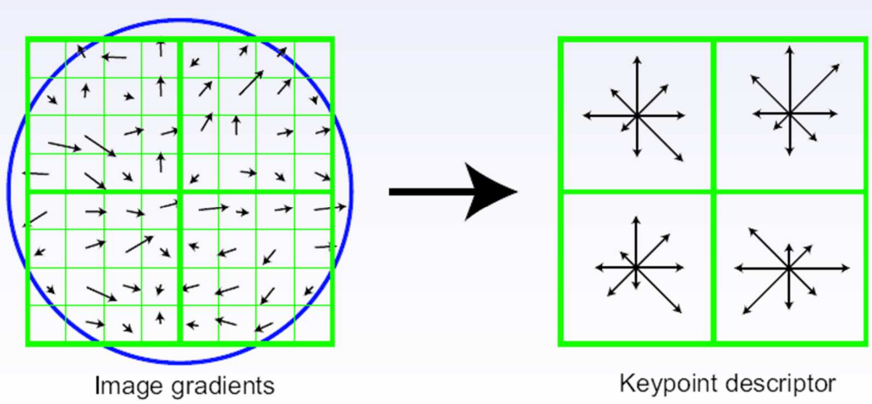
(Wersing & Körner, Neural Computation 2003)

Robustness to Clutter, Scaling, and Shifting

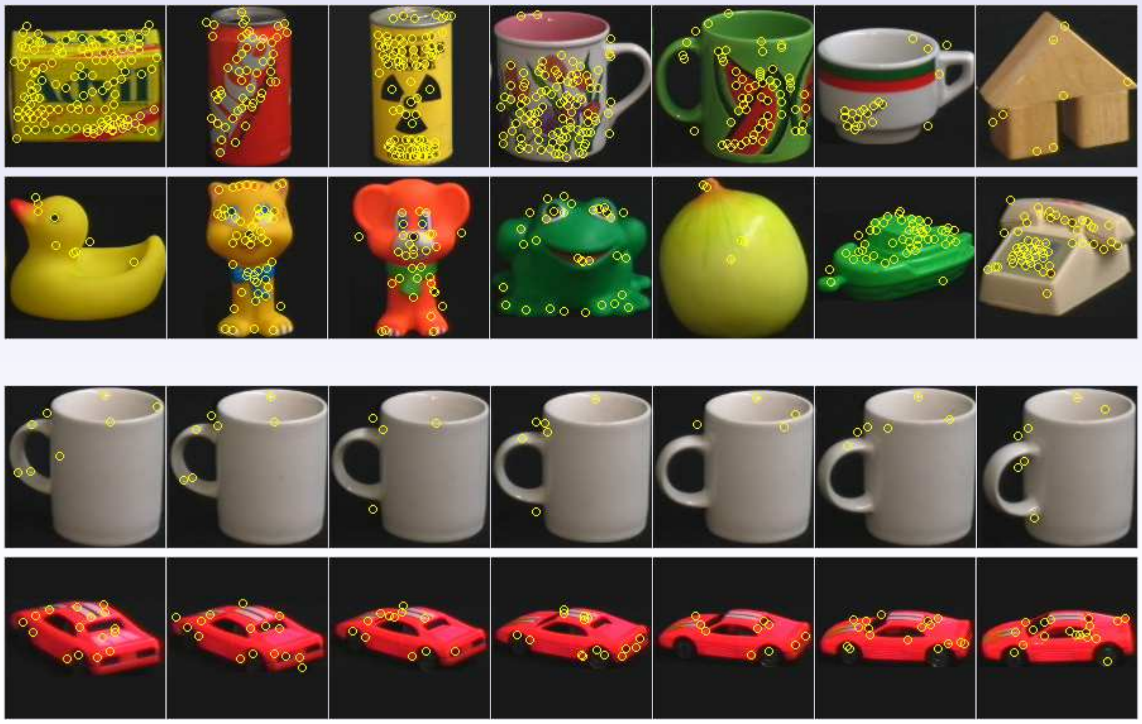


Detour – Standard computer vision recognition approach

- Very popular: SIFT feature-based recognition
- Scale-invariant feature transform (Lowe 2004)
- Interest points → Local gradient histogram



Example images with marked keypoints

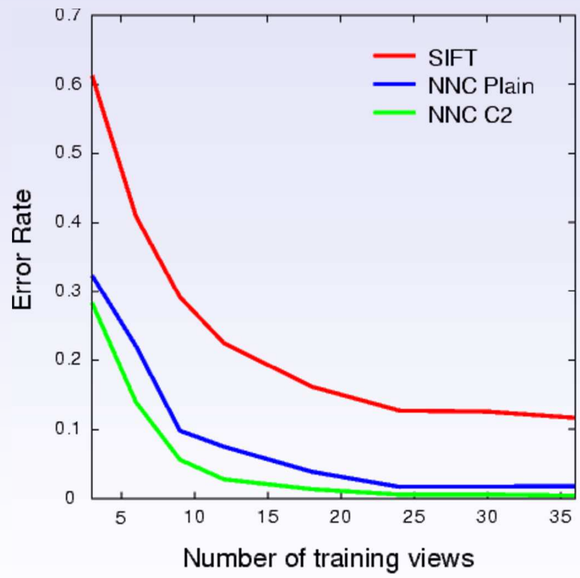


Experiments

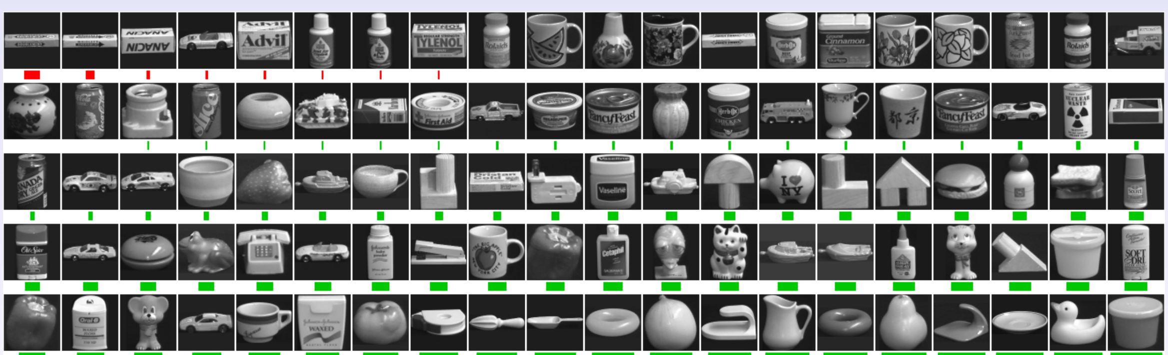
- Software: Evolution Robotics' ViPR™
  - might be different from described algorithm
  - D.G. Lowe called ViPR™ the "gold standard implementation" of the SIFT algorithm
  - AIBO was equipped with this system for landmark navigation
- Database: Coil-100 (128x128 pixel, gray-scale)

Error rates

- SIFT
  - optimal setting used
  - → most features found
  - many false negatives, although detection threshold was very low
  - → for some views no match found at all
  - → not enough matching votes in Hough step ?
- Nearest-neighbor
  - on C2 layer slightly better than on plain images



Winner takes all (SIFT vs. BIPCA)



SIFT wins   Ventral Hierarchy wins