# Merging color and shape in a hierarchical pattern recognition model

SPATIAL COGNITION

A5-[ActionSpace]

Sven Eberhardt (sven2@uni-bremen.de), Christoph Zetzsche, Manfred Fahle and Kerstin Schill

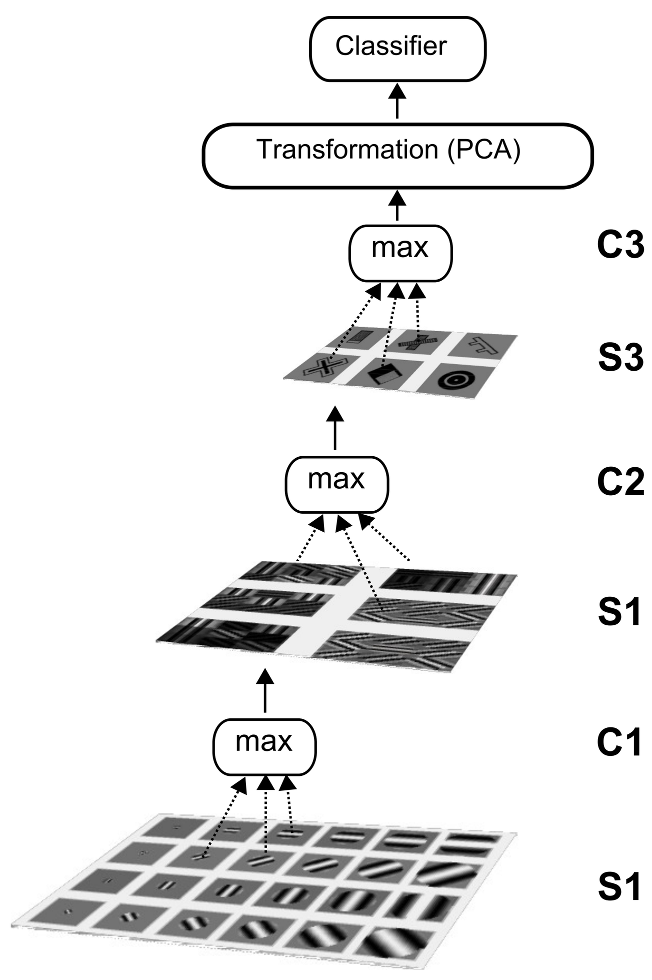
## Introduction

When we are looking and recognizing objects in natural scenes, shape and color information seem inseparably linked. However, neurobiological evidence suggests that color and shape processing in humans happen in a diverse number of distinct areas within the visual cortex, which provide specialized functionality for each submodality[2]. The exact amount of parallel processing and the point of merging these information channels into a multimodal representation is still disputed. Here, we approach the problem from a computational point of view and ask at which point modalities should be merged to maximize information in natural image statistics and test classification performance on natural tasks.

## Model

We use a simplified version of HMax, hierarchical feed-forward pattern recognition model [1].

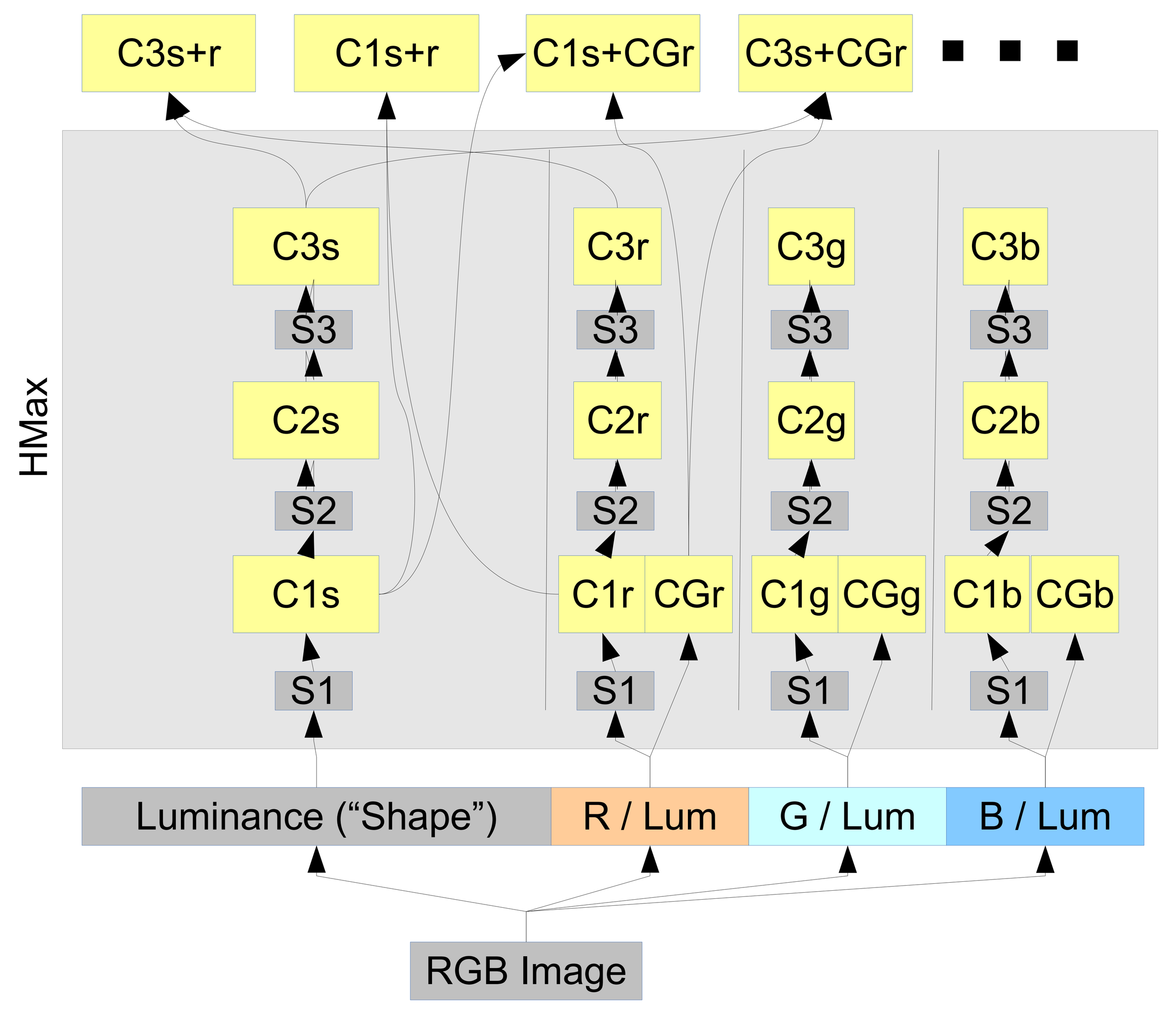
- Alternates simple cell layers (S) for patter matching and complex cell layers (C) for pooling.
- Simple cells match increasingly complex features
- Complex cells provide invariance over increasingly large locations and scales
- Mimics receptive field behaviour found in the ventral stream of primate visual cortex.
- Has been successfully used to replicate human psychophysics results on rapid animal detection.




In addition, we calculate global luminance histograms on shape (CGs) and all color channels (CGr, CGg, CGb) as low-level control features.

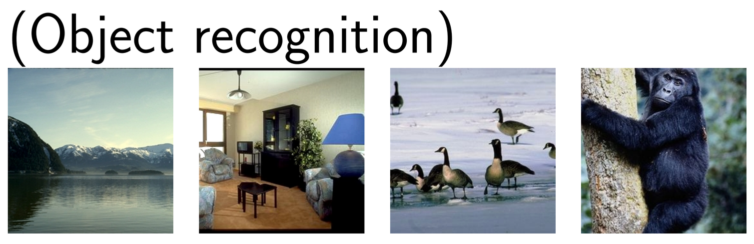
To test color processing, we split the input into luminance and color channels divided by luminance.

- Every channel is run through a separate HMax layer cascade and
- Features are extracted at various levels
- We test combining color and shape features of varying complexity
- Combined feature vectors are normalized to equal size (128 components) using PCA and then fed into a linear classifier (LSQ)




## Datasets

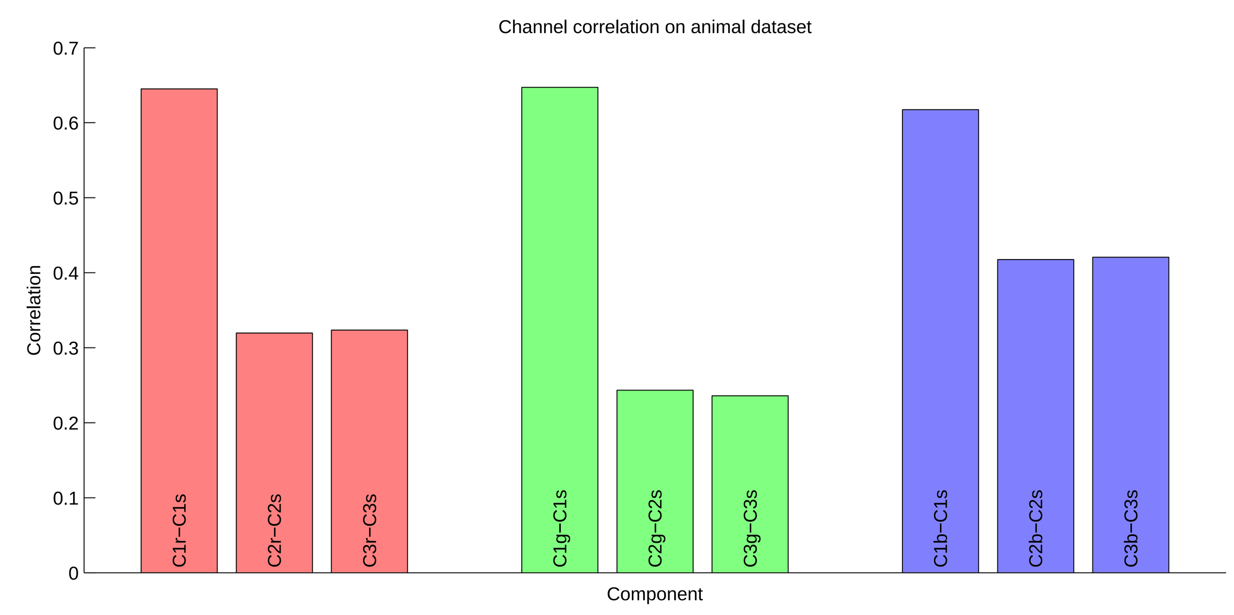
Dataset 1: Animal/Nonanimal[1]
(Object recognition)



Dataset 2: Streetview[3]
(Self-localization)

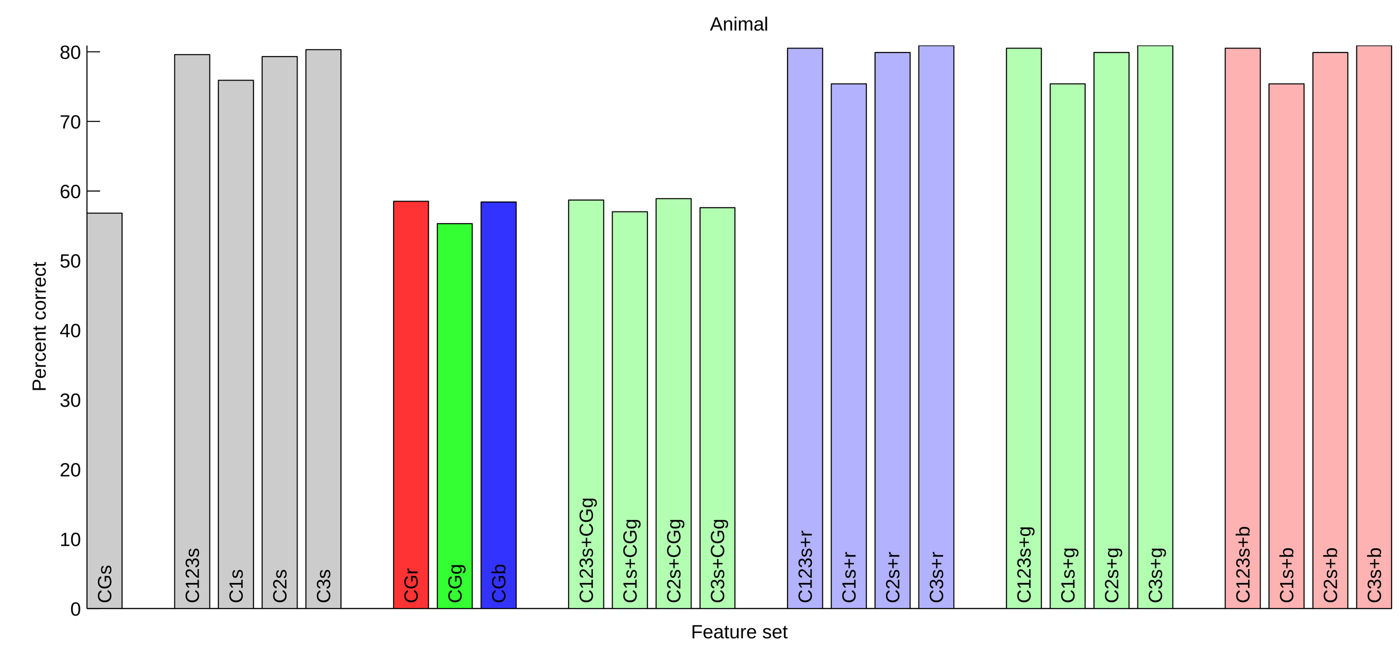
## Results

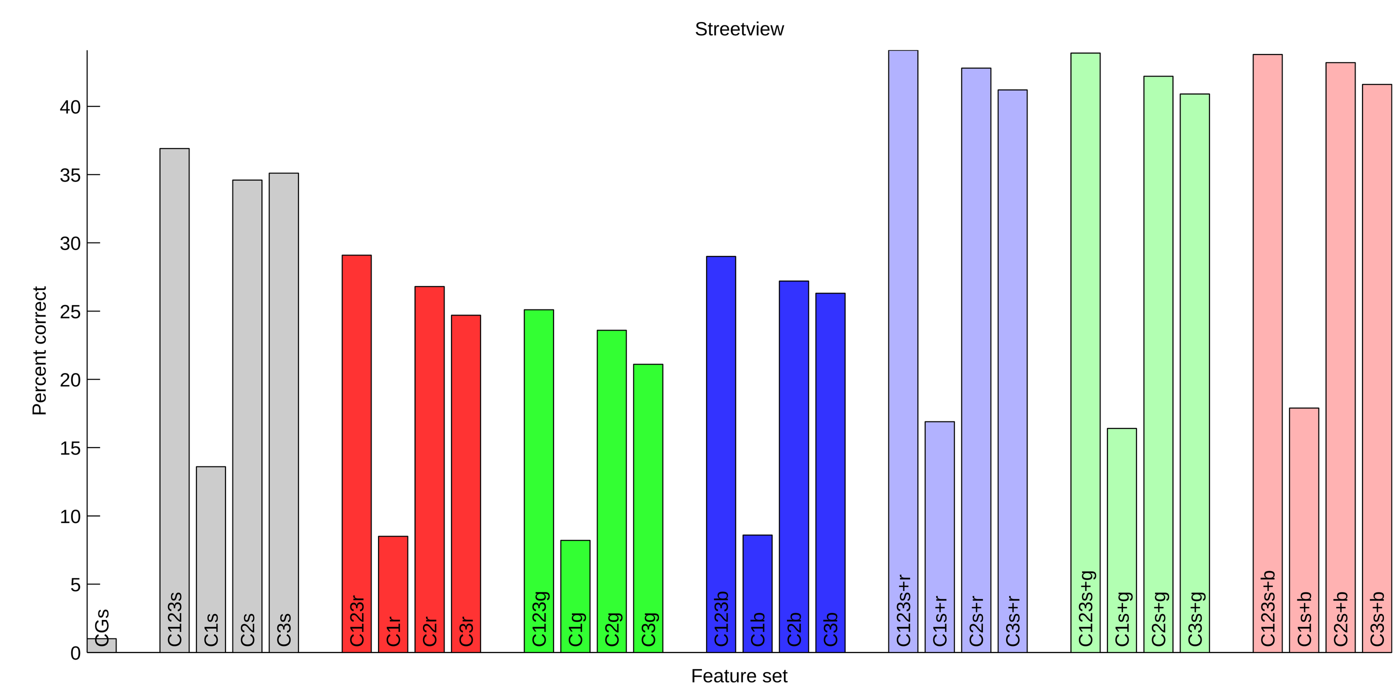
### Feature correlation




- At a low level, color and shape features are highly linearly correlated, hinting that they should be processed as compound features
- However, at higher stages, more different features emerge

### Classification performance




- Performance on low-level color features (both C1 and CG) is poor
- For animal detection, fusion of shape and color on low levels (C1) decreases performance. Fusion on high level increases it.
- Fusion of color information from higher processing levels (C3r,g,b)
- All performance gains on color fusion are very low

## Discussion

- Intuitively, adding color information to the classifier should increase performance a lot - but this is not the case
- **Fusion of color information from higher processing levels (C3r,g,b) helps more than lower levels. However, the effect is marginal.**
- **Simple fusion of color and shape by feature merging in PCA does not increase classification performance significantly.**

We conclude that color information needs to be processed separately from shape and that different models are required for each modality. This may explain why separate eras process information in the primate visual cortex.

## References

[1] Serre, T., et al. "'A Feedforward Architecture Accounts for Rapid Categorization."' PNAS 104, no. 15 (2007): 6424-6429.
[2] Kruger, N., et al. "Deep Hierarchies in the Primate Visual Cortex: What Can We Learn For Computer Vision?." IEEE transactions on pattern analysis and machine intelligence (2012).
[3] Eberhardt, S., et al. "'Low-level global features for vision-based localization."' KI 2013: WS on Vis. and Spat. Cog. (in press)

Universität Bremen
DFG Deutsche Forschungsgemeinschaft
UNI FREIBURG