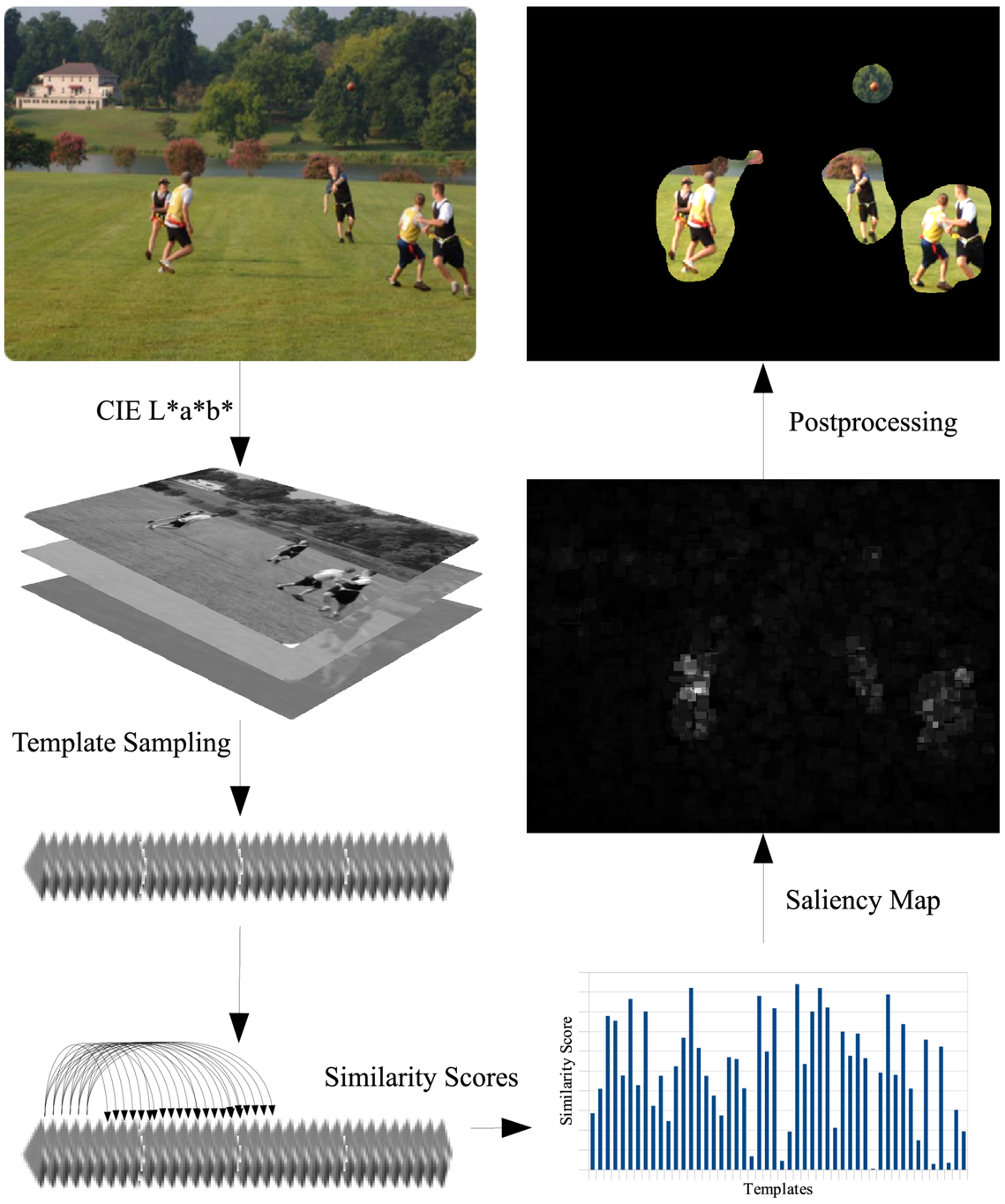
INSTITUTE FOR COGNITIVE SYSTEMS

Technische Universität München

SAMPLED TEMPALTE COLLATION FOR VISUAL ATTENTION AND OBJECT-BASED SEGREGATION

Andreas Holzbach, Gordon Cheng, Institute for Cognitive Systems, Technische Universität München. [andreas.holzbach@tum.de, gordon@tum.de]



Sampled Template Collation

Our model calculates the saliency map by sampling templates randomly over the image. Each template is then compared to the other templates by calculating a dissimilarity score. Higher scores mean lower similarity, lower responses higher similarity. Templates with a higher overall dissimilarity score therefore originate from areas in the image which stick out.

Color Space

We use a L2 norm to calculate the difference of lightness L and color-opponent dimensions a and b between two templates T1 and T2 in Lab Color Space .

$$l = ||T1_L - T2_L||_{L_2} = \sqrt{\sum (T1_L - T2_L)^2}$$
$$a = ||T1_a - T2_a||_{L_2} = \sqrt{\sum (T1_a - T2_a)^2}$$
$$b = ||T1_b - T2_b||_{L_2} = \sqrt{\sum (T1_b - T2_b)^2}$$

Distance Weight

We include a distance weight to the dissimilarity score to account for local salient areas. Templates which are closer together have a higher weight than templates which are e.g. on the opposite side of the image.

$$w = 1 - \frac{d(T_1, T_2)}{max(d)}$$

Entropy

We integrate the self-information of a template in our model, as areas which would be salient because of their lightness and color uniqueness - e.g. a small area of a blue sky in the top of an image are not salient to a human subject.

$$H(X) = -\sum_{m=1}^{M} p_m \log p_m$$

Overall dissimilarity score

The per-template dissimilarity score s is calculated by:

$$s = l(a + b) * w * H(T_1)H(T_2)$$

Computational complexity

$$O(n \log n) + O(n * k)$$
$$= O(n \log n) + O(n)$$
$$= O(max(n \log n, n)).$$

Object-Based Segregation

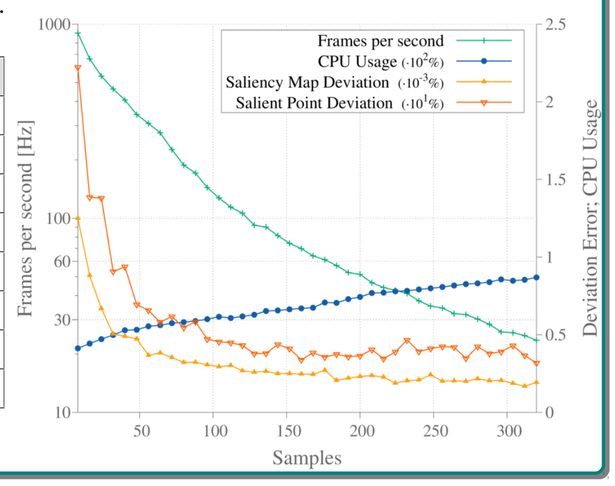
Object-based attention describes a pattern-specific attentional filtering in the visual cortex. Activity patterns in early visual areas are strongly biased in favour of the attended object. This phenomenon contributes towards the recognition of objects in higher cortical areas.
In our model one single seed template is taken from the area with the highest salient point computed by the previous STC procedure. All following sampled templates are then compared to this seed template using a similar metric as for the visual attention processing:

$$\mathrm{s} = (\mathrm{a} + \mathrm{b}) + (\alpha * l) + |H(T_1) - H(T_2)|$$

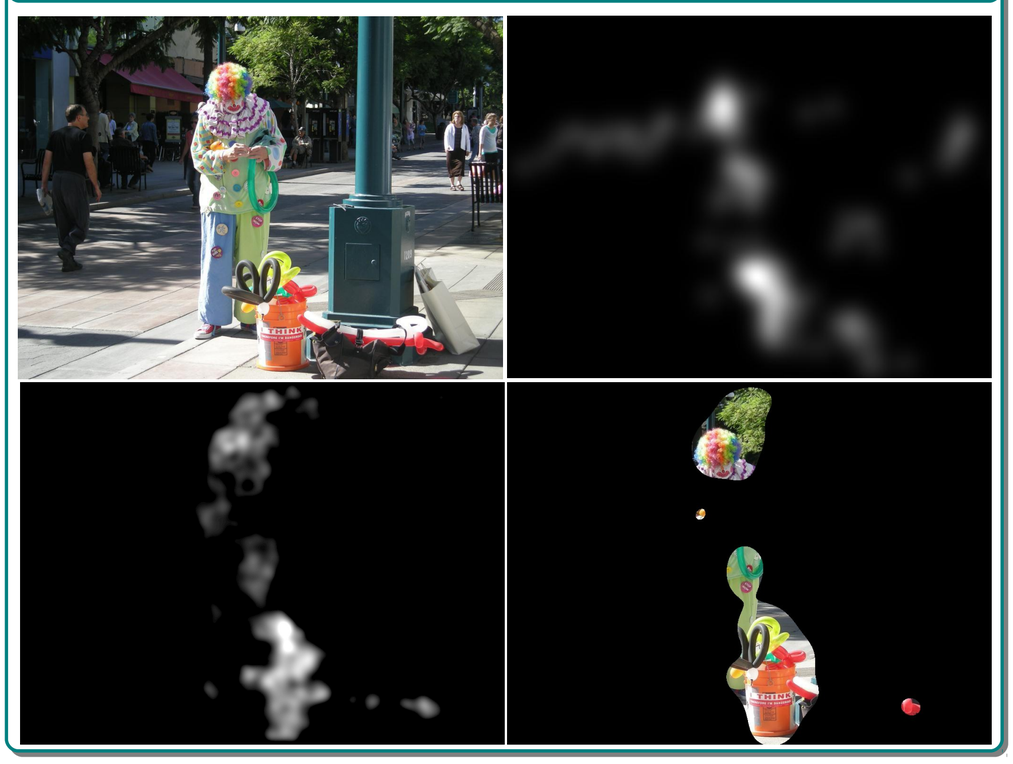
Saliency Benchmark Results

Table 1: Results of Judd's saliency benchmark.

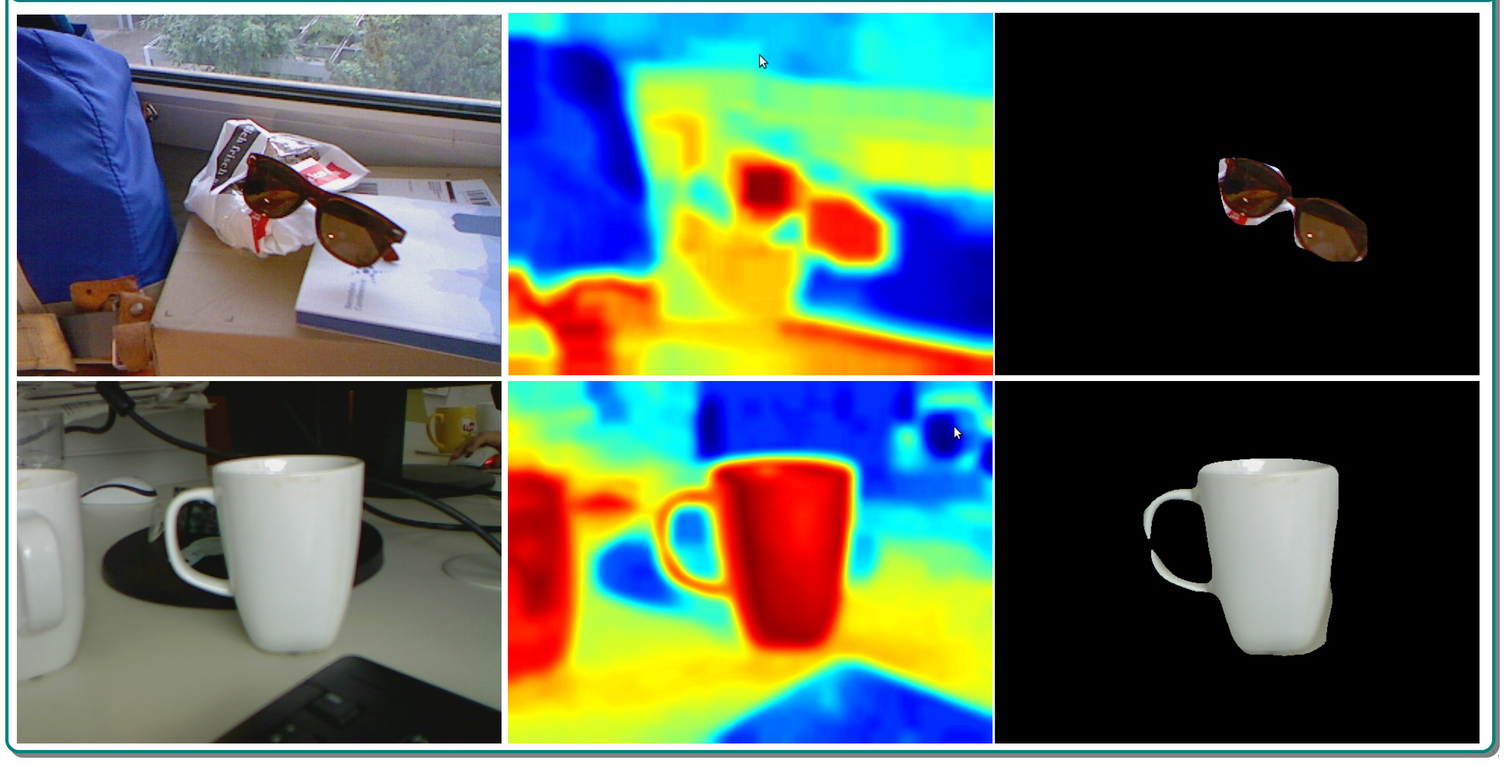
Model	ROC	Similarity
GBVS	0.801	0.472
Our Model /w CB	**0.794**	**0.477**
Multi-Resolution AIM	0.772	0.471
Center Based	0.783	0.451
Our Model /wo CB	**0.687**	**0.357**
Torralba	0.684	0.343
Itti & Koch	0.562	0.284
Chance	0.503	0.327



Saliency Maps Example



Object-Based Segregation Examples



References: [1] Desimone, R and Duncan, J, *Neural mechanisms of selective visual attention.*, 1995. [2] Judd, T. et al., *A Benchmark of Computational Models of Saliency to Predict Human Fixations*., 2012