Saliency Filters: Contrast Based Filtering for Salient Region Detection

Course : Computer Vision (CSE 578)

Prof. Anoop Namboodiri

Prof. Avinash Sharma

Centre for Visual Information Technology

IIIT Hyderabad, Spring 2019

Students: Ujjwal Tiwari | Harshit Kumar Sankhla | Udit Singh Parihar

Salient Team Members

Harshit Kumar Sankhla

Ujjwal Tiwari

Udit Singh Parihar

VISIONARIES !

Contents

- Introduction
- Applications of saliency estimation
- Contributions of this paper
- Detailed Approach
- Results & Failures
- Extension
- Results and comparison
- Deep Learning vs Traditional computer vision methods
- References

Terminology 101

- **Saliency** - Quality of being particularly noticeable.
- **Saliency Map** - A **saliency map** is an image that shows each pixel's unique quality. The goal of a saliency map is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse. Saliency is a kind of image segmentation.
- **Image Segmentation** - The process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse.

The Problem Statement

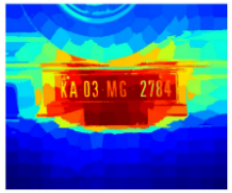
- The problem of saliency estimation deals with the automatic identification of image elements that catch the attention of a human observer.
- The goal of saliency estimation is to find a high level scene description. An understanding of what stands out in an image.
- Saliency estimation can also be viewed as a segmentation problem that also produces semantically meaningful Image segments.

Applications

Saliency estimation has found wide applications in computer vision such as:

1. Mobile robotics [Mobile robot Navigation]
2. Object Detection [Number Plate detection]
3. Automatic Thumb-nailing
4. Cropping/Resizing (Seam Carving-content aware image resizing)
5. Adaptive Compression

Number Plate Detection



Automatic Thumbnailing

- Mobile robot navigation

- Content aware resizing (Achanta + Susstrunk, 2009)

Previous Work and Motivation

Definition of contrast is based on various different types of image features such as colour variation of individual pixels, edges and gradient, spatial frequencies, histograms, multi-scale descriptors etc.

For this work we use two different contrast measures as features for dense estimation of saliency in an Image.

Contributions of the Paper

Contribution 1 : Computation of contrast based features for saliency estimation

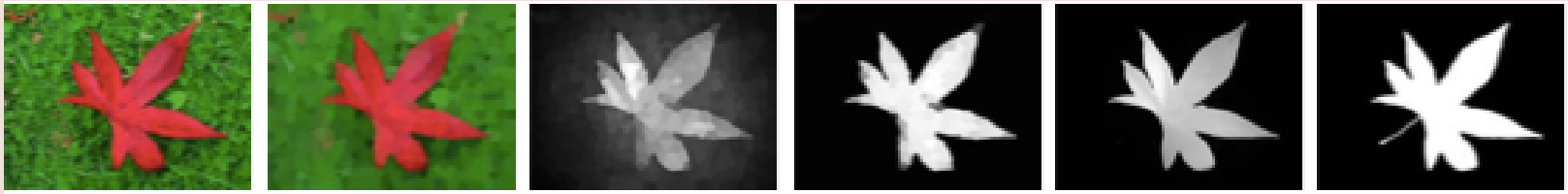
1. **Super-segmentation** : In this method, Image is decomposed into structurally representative elements that abstract away unnecessary details.
2. This allows for a very clear and intuitive definition of contrast.
3. Saliency estimate is based on two well defined contrast measures namely uniqueness and spatial distribution of element which is computed on top of the processed image in the first step.

Contributions of the Paper

Contribution 2 : Modelling the problem to optimise time complexity

1. Local and global contrast are handled by these measures in a unique way.
2. Formulation of these measures as single high dimensional Gaussian filtering framework accounts for a highly efficient implementation with linear complexity O(N).

Detailed Approach



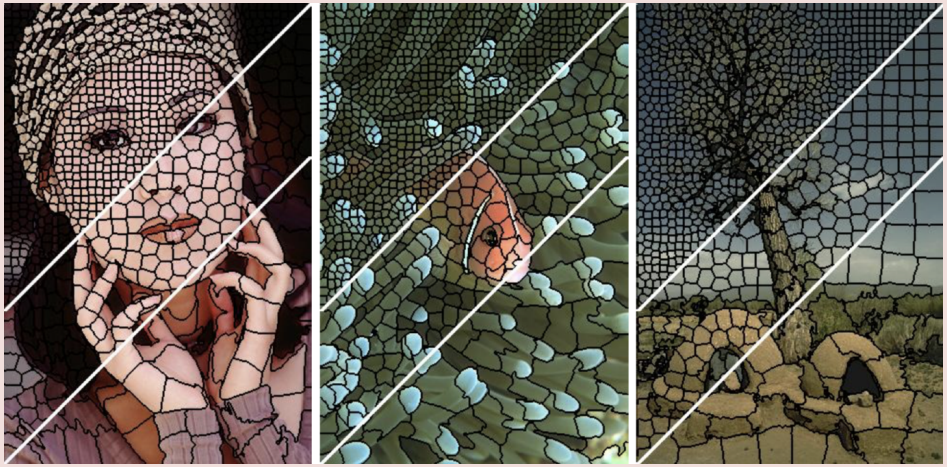
(a) Source image. (b) Abstraction. (c) Uniqueness. (d) Distribution. (e) Saliency. (f) Ground truth.

Detailed Approach

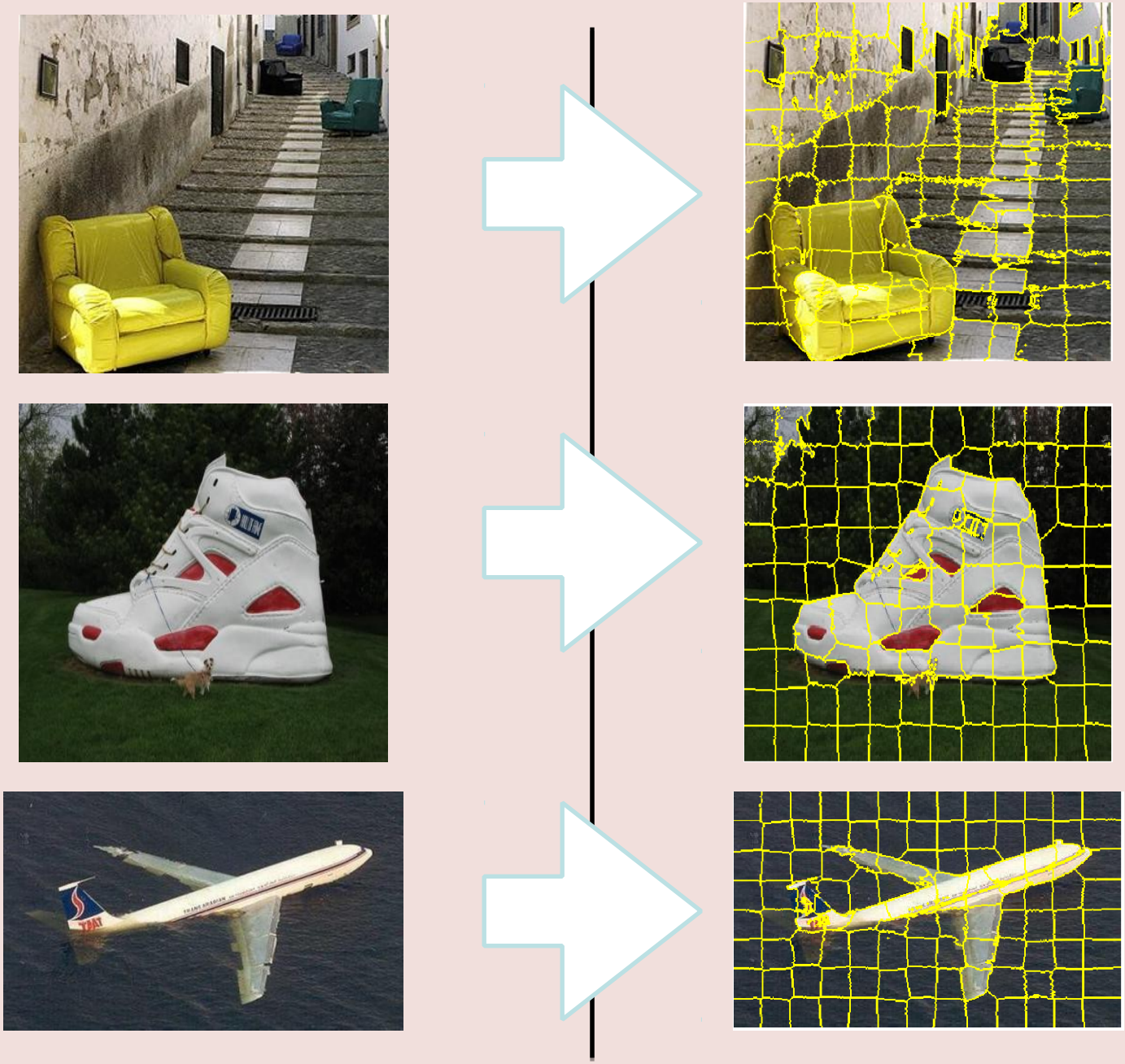
Step 1 - Abstraction

1. Decompose the image into basic elements that preserve relevant structure, but abstract undesirable detail.
2. Each element locally abstracts image by clustering pixels with similar properties.
3. Strong contours and edges in the image are preserved.
4. To achieve this decomposition we implement a localised over segmentation over the image using SLIC Super-pixels.
5. SLIC segments image by using K-means clustering in RGBXY space.
6. In this paper this is done in CIELab space using geodesic image distance. Geodesic distance guarantees connectivity, while retaining locality compactness and edge aware.

Detailed Approach



The algorithm proposed, **Simple Linear Iterative Clustering (SLIC)** performs a local clustering of pixels in the 5-D space defined by the *L,a,b* values of the CIELAB colour space and the *x,y* pixel coordinates. A novel distance measure enforces compactness and regularity in the superpixel shapes, and seamlessly accommodates grayscale as well as colour images. SLIC is simple to implement and easily applied in practice – the only parameter specifies the desired number of superpixels.



Detailed Approach

Step 2 - Element Uniqueness

1. The definition of uniqueness is the rarity of segment *i*, given its position p_i and colour in CIELab c_i compared to all other segments *j*.
2. Global operator does not represent sensitivity to local contrast variation.
3. Evaluating this globally requires $O(N^2)$, *N* is the number of segments. Methods that down sample the image are prone to blurring the edges.
4. However, for Gaussian weights, this can be done in O(N). Z_i is the normalising factor and sigma controls the range of uniqueness operator.

$$U_i = \sum_{j=1}^{N} \|\mathbf{c}_i - \mathbf{c}_j\|^2 \cdot \underbrace{w(\mathbf{p}_i, \mathbf{p}_j)}_{w_{ij}^{(p)}} .$$

Uniqueness defined by colour similarity and cluster position

$$\begin{aligned} U_i &= \sum_{j=1}^{N} \|\mathbf{c}_i - \mathbf{c}_j\|^2 w_{ij}^{(p)} \\ &= \mathbf{c}_i^2 \underbrace{\sum_{j=1}^{N} w_{ij}^{(p)}}_{1} - 2\mathbf{c}_i \underbrace{\sum_{j=1}^{N} \mathbf{c}_j w_{ij}^{(p)}}_{\text{blur } c_j} + \underbrace{\sum_{j=1}^{N} \mathbf{c}_j^2 w_{ij}^{(p)}}_{\text{blur } c_j^2} . \end{aligned}$$

Expanding the terms, reducing time complexity

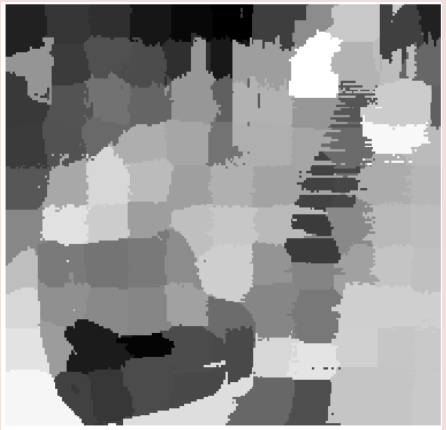
Detailed Approach

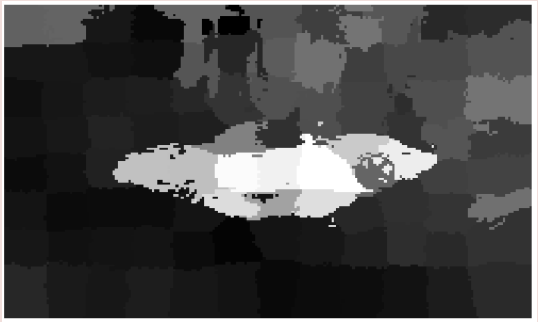
Step 3 - Element Distribution

1. Grouping based on compactness of salient features
2. Variance term Di based on spatial position and similarity between colours
3. The weightage term is set to be Gaussian for reducing quadratic complexity to linear.

$$D_i = \sum_{j=1}^{N} \|\mathbf{p}_j - \mu_i\|^2 \underbrace{w(\mathbf{c}_i, \mathbf{c}_j)}_{w_{ij}^{(c)}},$$

Uniqueness defined by colour similarity and cluster position





Detailed Approach

Step 4 - Saliency Assignment

1. We start by normalising both uniqueness U_i and distribution D_i to the range [0,...,1]
2. We compute saliency values S_i as a combination of uniqueness and distribution values for the clusters.
3. We use K = 6 as the scaling factor for the exponential.

$$S_i = U_i \cdot \exp(-k \cdot D_i)$$

Final Saliency Assignment

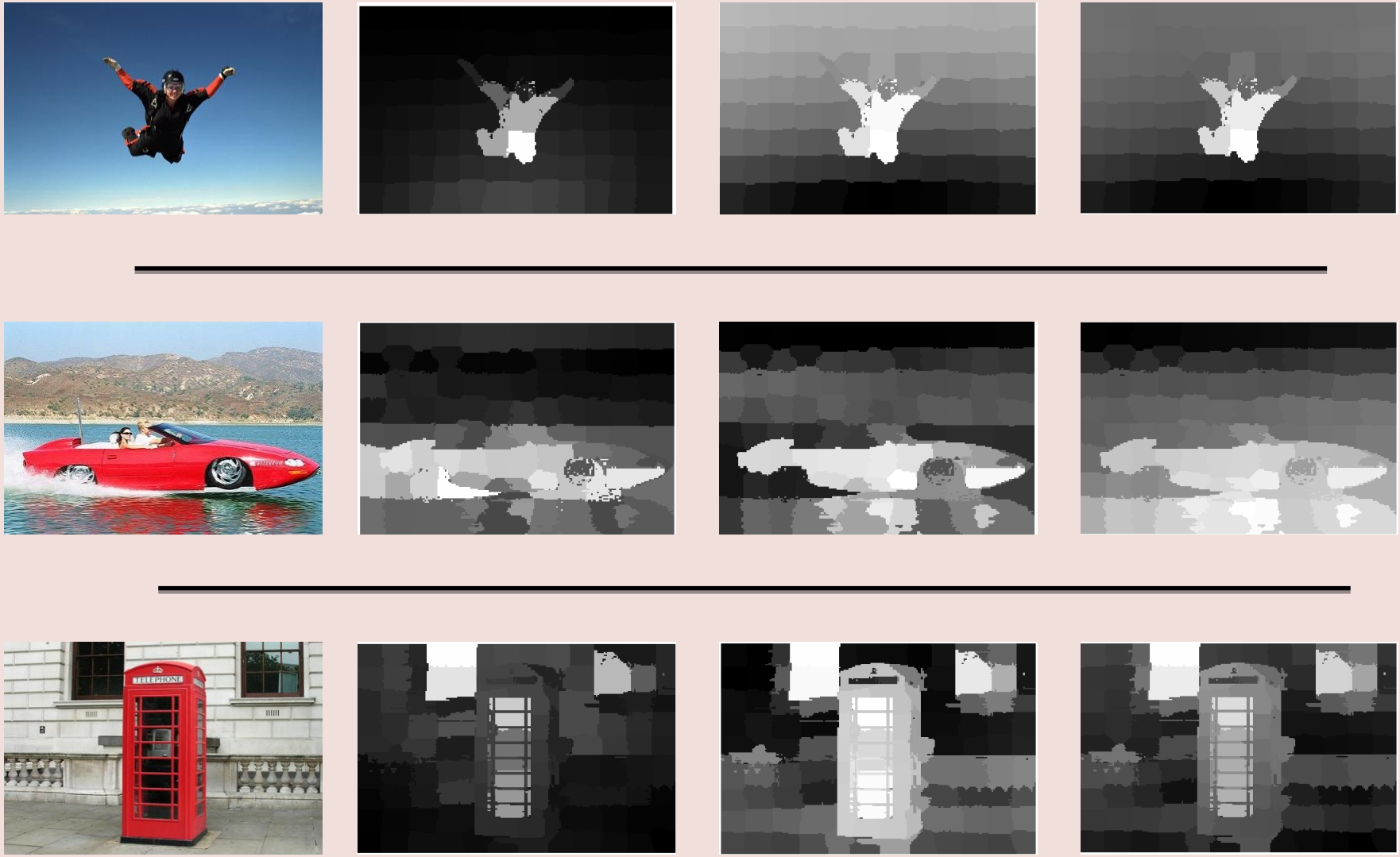
- As the final step, we need to assign a final saliency value of each image pixel, which can be interpreted as an upsampling of the per-element saliency S_i
- We define the saliency $\check{S}_i$ of a pixel as a weighted linear combination of the saliency S_j of its surrounding image elements.

$$\tilde{S}_i = \sum_{j=1}^{N} w_{ij} S_j .$$

$$w_{ij} = 1/Z_i \times \exp(-1/2 \times (\alpha \|c_i - c_j\|^2 + \beta \|p_i - p_j\|^2))$$

- By choosing a Gaussian weight as
  we can ensure that the upsampling process is both local and colour sensitive.
- Here alpha and beta are parameters controlling the sensitivity to colour and position.

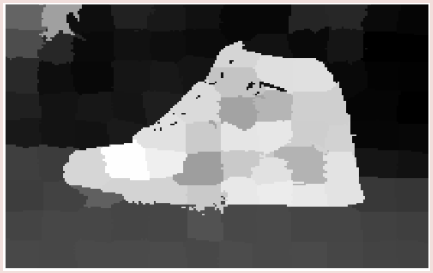
ORIGINAL UNIQUENESS DISTRIBUTION SALIENCY

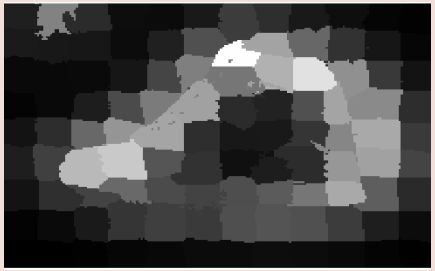


Varying Sigma for Uniqueness

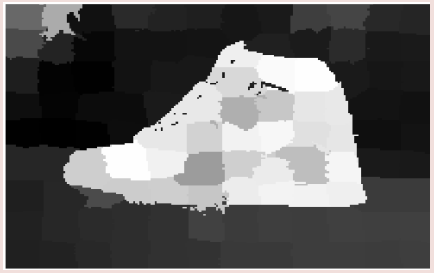
UNIQUENESS

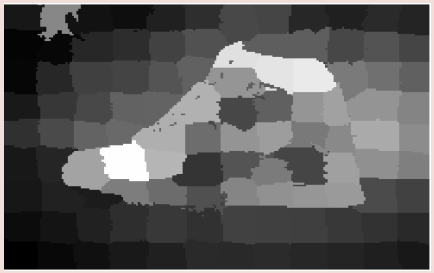
0.01



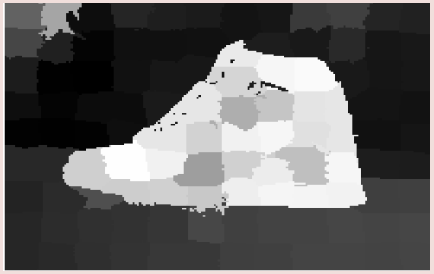


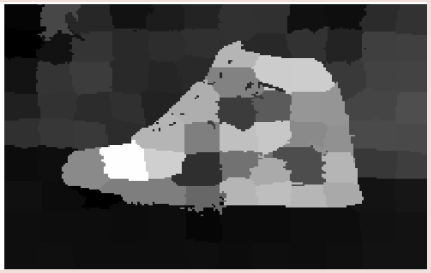
0.1



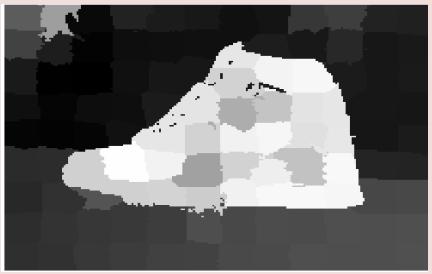


0.25





1



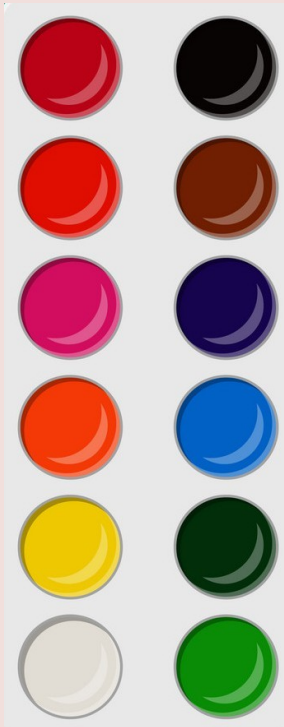
SALIENCY

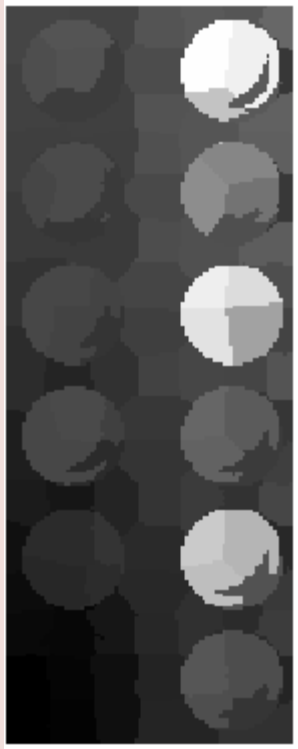
Why Both Uniqueness and Distribution ?

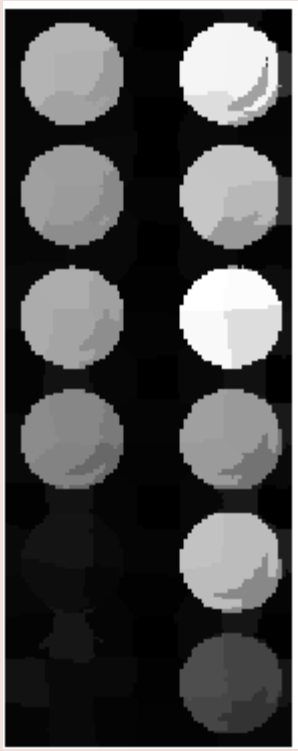
Without the distribution term, just the rarity of the colour is insufficient to correctly identify salient objects.

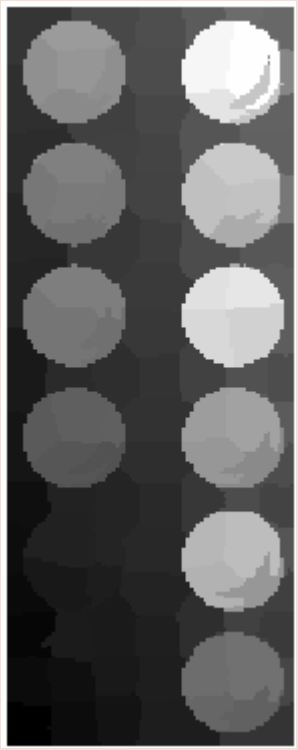
Observed Limitation

ORIGINAL → *UNIQUENESS* → *DISTRIBUTION* → *SALIENCY*



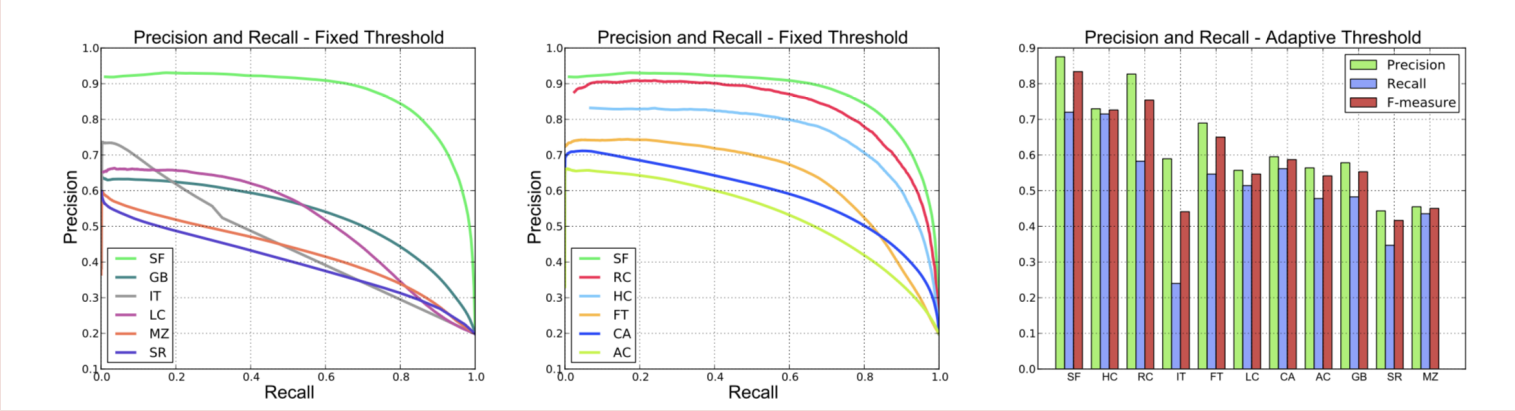






Colour similar to background have low uniqueness, due to low rarity factor. Distribution is able to recover most of the colours lost in uniqueness, due to compactness of region.

Precision and Recall



METHOD 1

Compare binary masks for every threshold in the range [0..255]. The resulting curves show that the algorithm consistently produces results closer to ground truth at every threshold and for any given recall rate

Precision and Recall

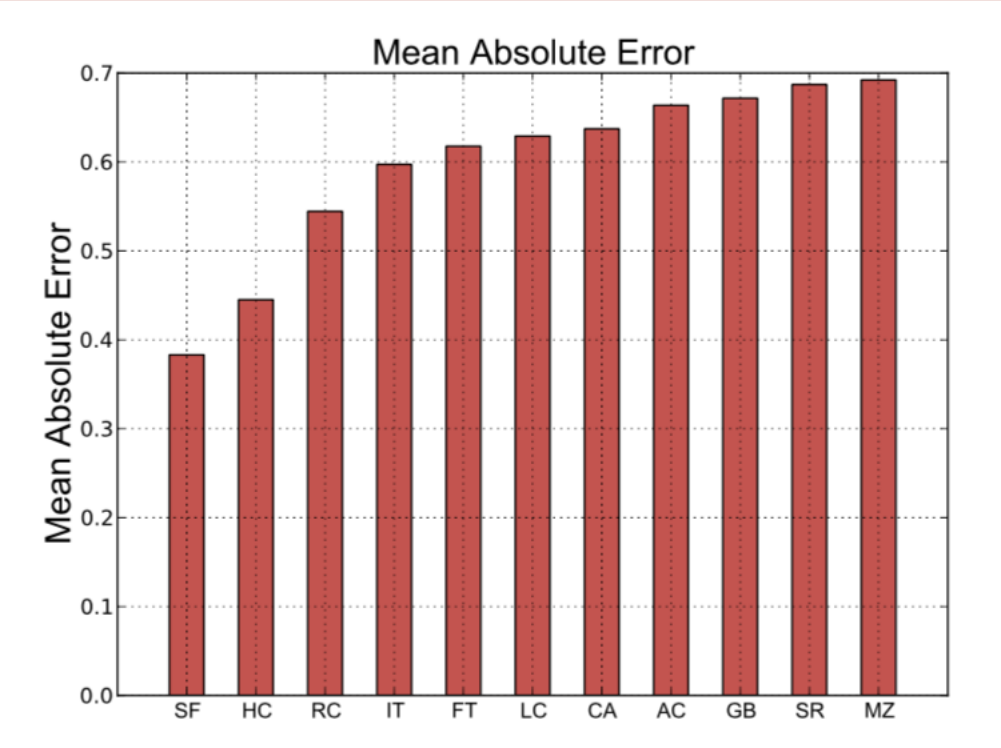
$$T_a = \frac{2}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y),$$

$$F_\beta = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}.$$

METHOD 2

Use the image dependent adaptive threshold, defined as twice the mean saliency of the image. In the above equations, W and H are the width and the height of the saliency map S, respectively. In addition to precision and recall we compute their weighted harmonic mean measure or *F- measure*

Mean Absolute Error



1. Estimates the negative saliency assignments
2. Number of pixels correctly marked as non-salient
3. Mean of absolute error of saliency map, S and ground truth

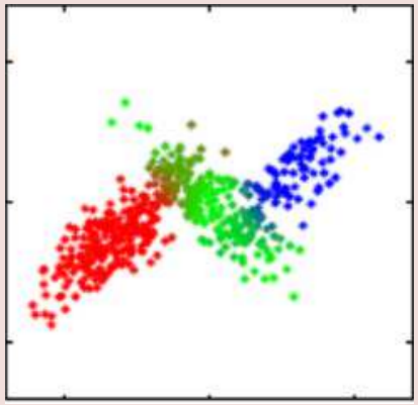
https://ieeexplore.ieee.org/document/937655

Dataset Used - database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics

EXPERIMENTATION WITH GMMs

WHY GMM ?

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \, p(\boldsymbol{x} | \boldsymbol{\theta}_k)$$



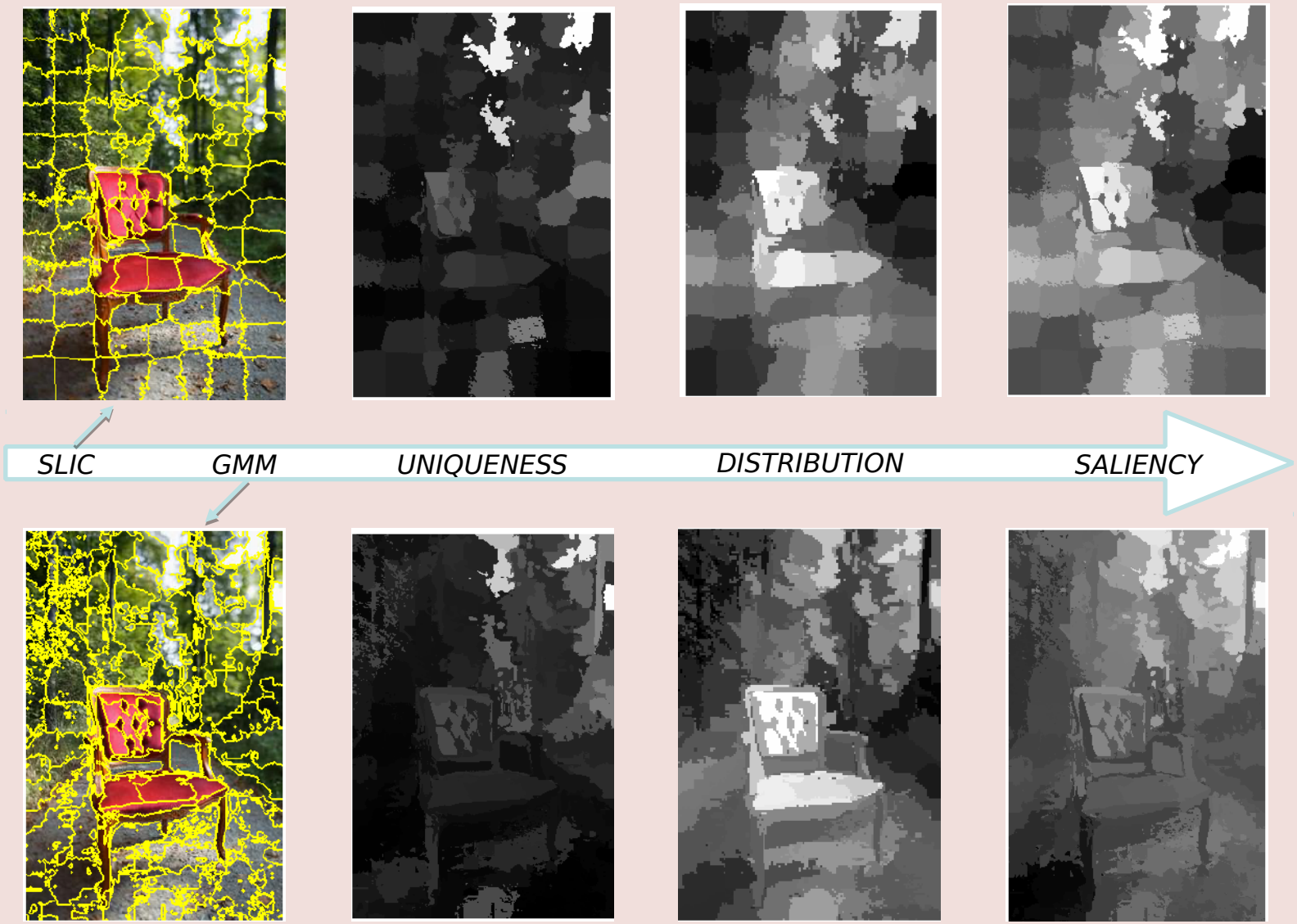
K - MEANS LIMITATIONS:

1.Point totally belongs to cluster or not, hard labelling.
2. Fails when clusters are unequal size.
3. Overlapping clusters.

GAUSSIAN MIXTURE MODELS:

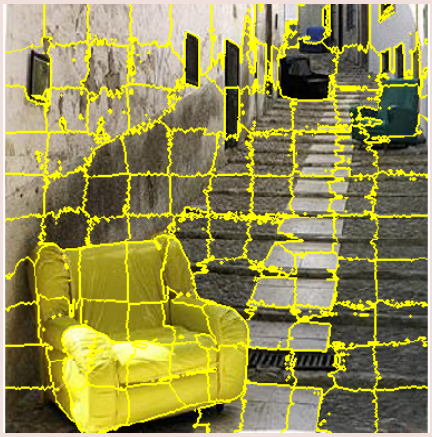
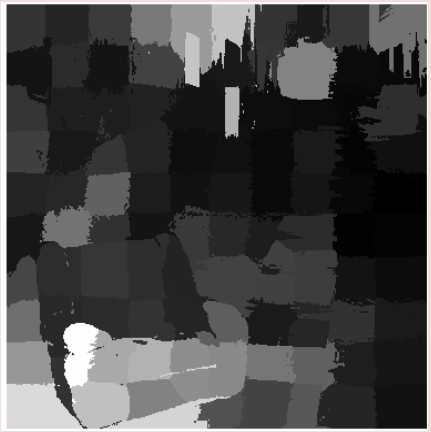
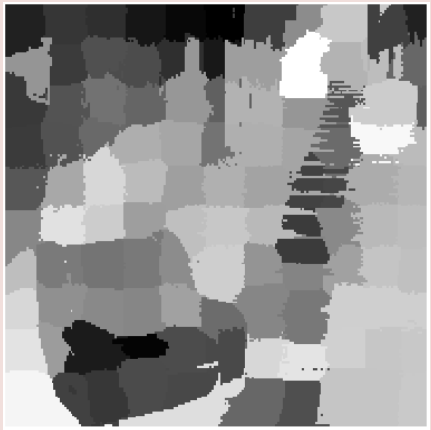
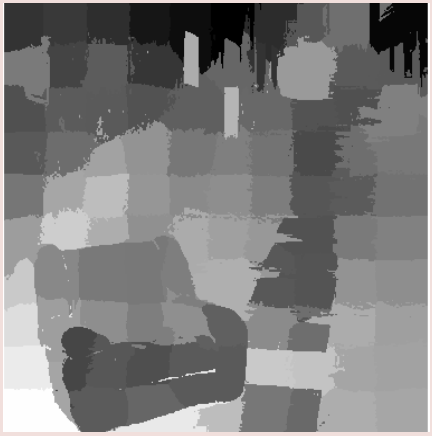
1. Works on overlapping, unequal size data.
2. Feature vectors experimented on RGB, RGBXY and LABXY space.

SLIC vs GMM - 1 Distinction from Background

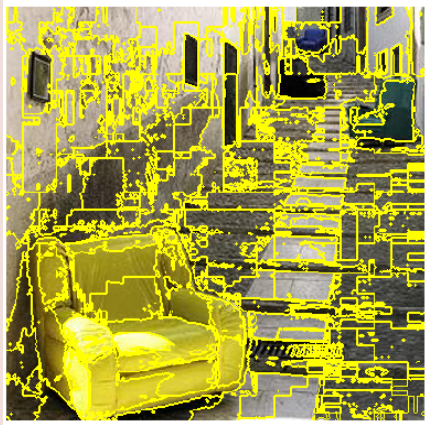


SLIC vs GMM - 2 Internal Structure

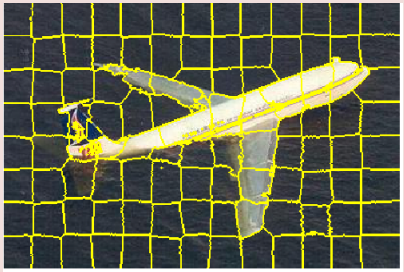
SLIC vs GMM - 3 Distant Objects

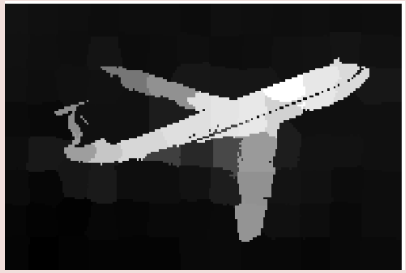
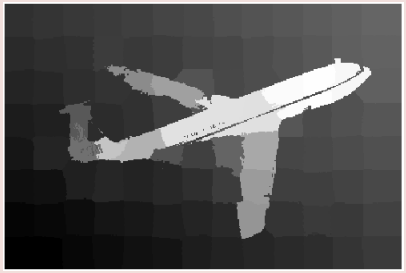





SLIC *GMM* *UNIQUENESS* *DISTRIBUTION* *SALIENCY*

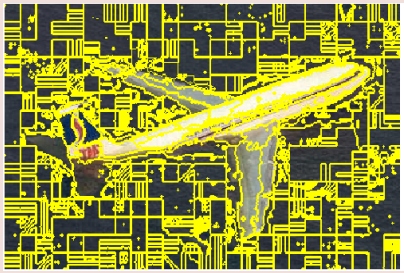


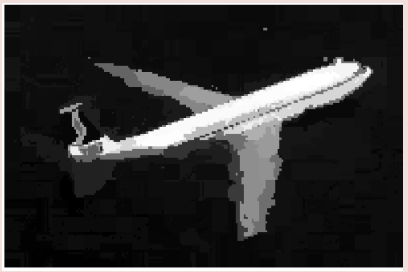
SLIC vs GMM - 4 Sharpness Along Edges



SLIC *GMM* *UNIQUENESS* *DISTRIBUTION* *SALIENCY*





Qualitative Comparison of Traditional Methods with Deep Learning Models

- Formulation driven vs Data driven approach: High computation cost. (Both memory and time complexity.)

- Interpretability of both approach and results in a traditional methods.

- Deep learning method need a lot of data to train! Data is abundant but task specific data and standard data is expensive and hard to find.

- Debugging: Due to this interpretability traditional methods are easier to debug.

- However, traditional methods like these are extremely biased towards contrast prior, background prior. As formulations are rigid they scaled to diverse set of images that do not conform to the

References

1. Saliency Filters : contrast based saliency estimation. [LINK]

2. Efficient Salient Region Detection with Soft Image Abstraction. [LINK]

3. A model of saliency-based visual attention for rapid scene analysis. [LINK]

4. SLIC Superpixels. Technical report. [LINK]

5. Learning to detect a salient object. [LINK]

6.Improved saliency detection based on superpixel clustering and saliency propagation. [LINK]

7. Global contrast based salient region detection. [LINK]

KEEP
CALM
AND
STAY
SALIENT