Departement of Systems and Computational Biology
Albert Einstein College of Medicine

An ideal observer model for grouping and contour integration in natural images

Jonathan Vacher

With:
Ruben Coen-Cagli (Albert Einstein College of Medicine, New-York)
and Pascal Mamassian (LSP, École Normale Supérieure, Paris).

ECVP
29/08/2019

jonathan.vacher@einstein.yu.edu

https://jonathanvacher.github.io

Image Segmentation *vs* Visual Segmentation

Image Segmentation *vs* Visual Segmentation

Great progress with deep learning:

- mostly supervised learning ($\sim$ top-down approach)
- smart architecture but different from biological vision
- performance oriented
- work as a black box

Image Segmentation *vs* Visual Segmentation

Great progress with deep learning:

- mostly supervised learning ($\sim$ top-down approach)
- smart architecture but different from biological vision
- performance oriented
- work as a black box

Visual segmentation is more involved !

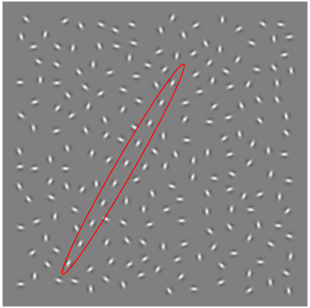
- variable but consistent across humans
- top-down + bottom-up processing

Our goal is to craft an open box model with all the ingredient of vision !

Grouping and contours integration: a basis for visual segmentation ?

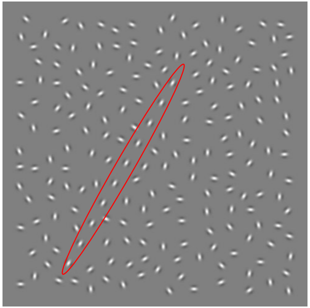
Grouping and contours integration: a basis for visual segmentation ?

Contour perception (Field et al. 1993)

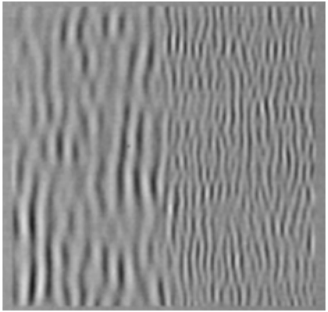


Grouping and contours integration: a basis for visual segmentation ?

Contour perception (Field et al. 1993)

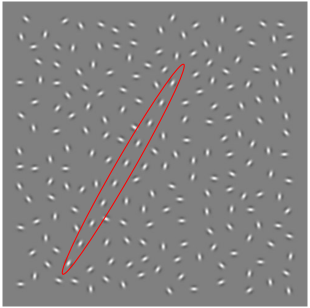


Texture perception (Landy et al. 2001)



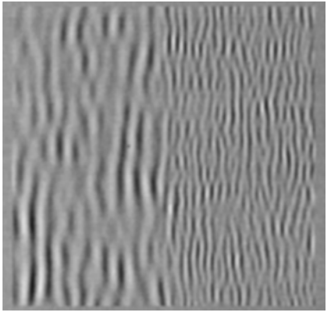
Grouping and contours integration: a basis for visual segmentation ?

Contour perception (Field et al. 1993)



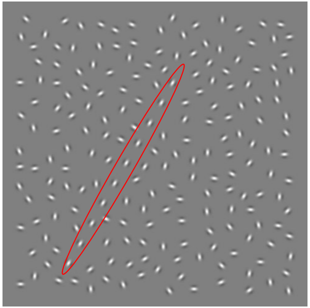
Artificial stimuli !

Texture perception (Landy et al. 2001)



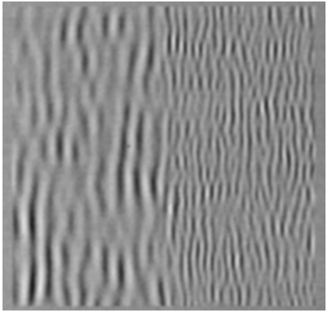
Grouping and contours integration: a basis for visual segmentation ?

Contour perception (Field et al. 1993)



Artificial stimuli !

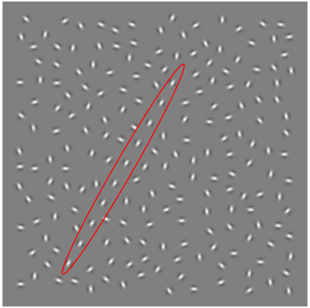
Texture perception (Landy et al. 2001)



Tractable models and well-controlled experiments . . .

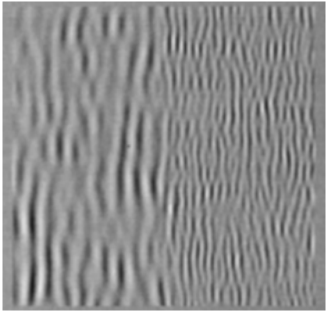
Grouping and contours integration: a basis for visual segmentation ?

Contour perception (Field et al. 1993)



Artificial stimuli !

Texture perception (Landy et al. 2001)



Tractable models and well-controlled experiments . . .
but how to generalize to natural images ?

Toward an ideal observer model for visual segmentation

An ideal observer for visual segmentation of natural images !

An ideal observer for visual segmentation of natural images !

Toward an ideal observer model for visual segmentation

An ideal observer for visual segmentation of natural images !

- To guide future model driven psychophysical experiments (ongoing work, not presented here)

An ideal observer for visual segmentation of natural images !

- To guide future model driven psychophysical experiments (ongoing work, not presented here)

Several constraints:

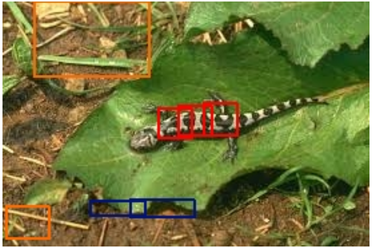
Toward an ideal observer model for visual segmentation

An ideal observer for visual segmentation of natural images !

- To guide future model driven psychophysical experiments (ongoing work, not presented here)

Several constraints:

- Image statistics



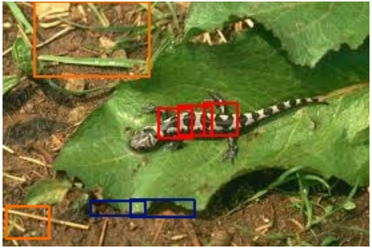
Toward an ideal observer model for visual segmentation

An ideal observer for visual segmentation of natural images !

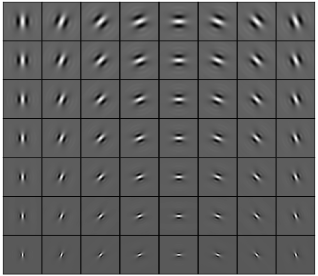
- To guide future model driven psychophysical experiments (ongoing work, not presented here)

Several constraints:

- Image statistics



- Cortical features



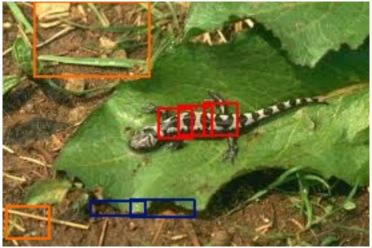
Toward an ideal observer model for visual segmentation

An ideal observer for visual segmentation of natural images !

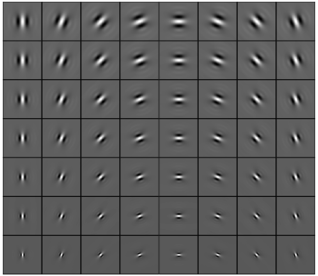
- To guide future model driven psychophysical experiments (ongoing work, not presented here)

Several constraints:

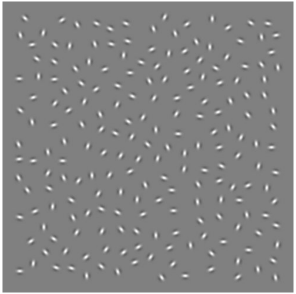
- Image statistics



- Cortical features



- Vision psychophysics

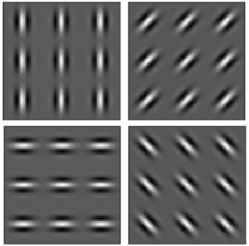


Representation and non-Gaussian statistics of natural images

How are images represented ?

Representation and non-Gaussian statistics of natural images

How are images represented ? $\Rightarrow$ decomposition in a wavelet basis (receptive fields) (Bell et al. 1997; Olshausen et al. 1996) ▶ $X = (X_1, \dots, X_n)^T = (\langle w_1, I \rangle, \dots, \langle w_n, I \rangle)^T$

$I =$  $(w_k)_k =$ 

Representation and non-Gaussian statistics of natural images

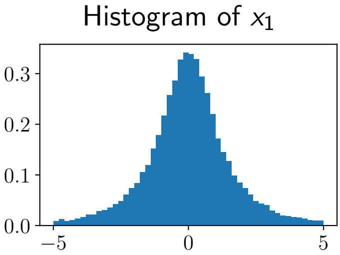
How are images represented ? $\Rightarrow$ decomposition in a wavelet basis (receptive fields) (Bell et al. 1997; Olshausen et al. 1996) ▶ $X = (X_1, \dots, X_n)^T = (\langle w_1, I \rangle, \dots, \langle w_n, I \rangle)^T$

What are the coefficient statistics ?

Representation and non-Gaussian statistics of natural images

How are images represented ? $\Rightarrow$ decomposition in a wavelet basis (receptive fields) (Bell et al. 1997; Olshausen et al. 1996) ▶ $X = (X_1, \dots, X_n)^T = (\langle w_1, I\rangle, \dots, \langle w_n, I\rangle)^T$

What are the coefficient statistics ? Non-Gaussian ! (Wainwright et al. 2000)



Representation and non-Gaussian statistics of natural images

How are images represented ? $\Rightarrow$ decomposition in a wavelet basis (receptive fields) (Bell et al. 1997; Olshausen et al. 1996) ▶ $X = (X_1, \ldots, X_n)^T = (\langle w_1, I \rangle, \ldots, \langle w_n, I \rangle)^T$

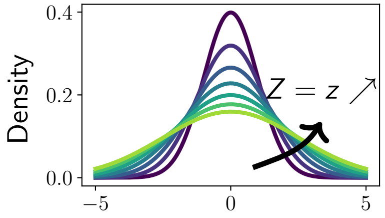
What are the coefficient statistics ? Non-Gaussian ! (Wainwright et al. 2000)

Definition (Gaussian Scale Mixture)

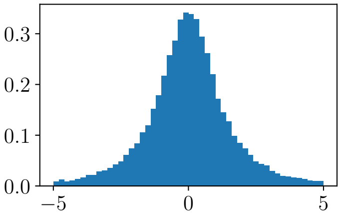
Gaussian vector of visual features ($G \sim \mathcal{N}(0, \Sigma)$) $\times$ Contrast between features ($Z \sim \mathcal{L}(\nu)$)

$$X = ZG.$$

Density of X_1 knowing $Z = z$



Histogram of x_1



Representation and non-Gaussian statistics of natural images

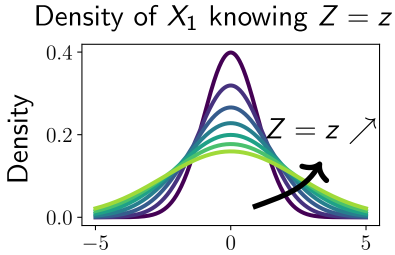
How are images represented ? $\Rightarrow$ decomposition in a wavelet basis (receptive fields) (Bell et al. 1997; Olshausen et al. 1996) ▶ $X = (X_1, \dots, X_n)^T = (\langle w_1, I\rangle, \dots, \langle w_n, I\rangle)^T$

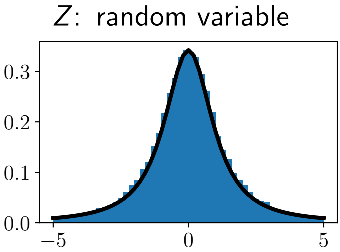
What are the coefficient statistics ? Non-Gaussian ! (Wainwright et al. 2000)

Definition (Gaussian Scale Mixture)

Gaussian vector of visual features ($G \sim \mathcal{N}(0, \Sigma)$) $\times$ Contrast between features ($Z \sim \mathcal{L}(\nu)$)

$$X = ZG.$$





How are images represented ? $\Rightarrow$ decomposition in a wavelet basis (receptive fields) (Bell et al. 1997; Olshausen et al. 1996) ▶ $X = (X_1, \dots, X_n)^T = (\langle w_1, I\rangle, \dots, \langle w_n, I\rangle)^T$

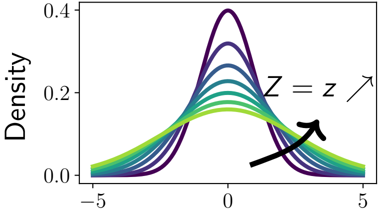
What are the coefficient statistics ? Non-Gaussian ! (Wainwright et al. 2000)

Definition (Gaussian Scale Mixture)

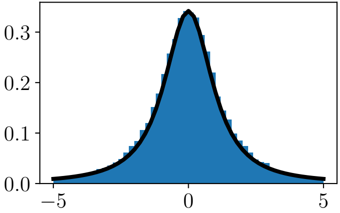
Gaussian vector of visual features ($G \sim \mathcal{N}(0, \Sigma)$) $\times$ *Contrast between features* ($Z \sim \mathcal{L}(\nu)$)

$$X = ZG.$$

Density of X_1 knowing $Z = z$

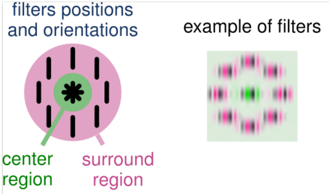


Z: random variable

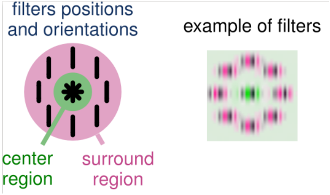


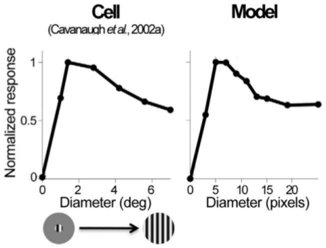
- ▶ Heavy-tailed distributions . . .
- ▶ and non-linear dependencies !

Gaussian Scale Mixture explains surround modulation in V1



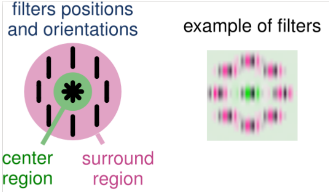
Gaussian Scale Mixture explains surround modulation in V1

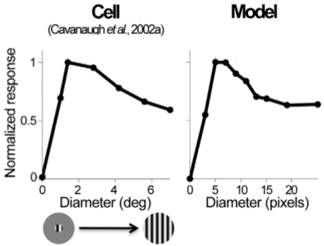




Taken from Coen-Cagli, Dayan, et al. 2012

Gaussian Scale Mixture explains surround modulation in V1





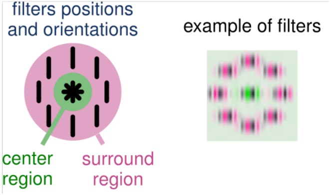
Taken from Coen-Cagli, Dayan, et al. 2012

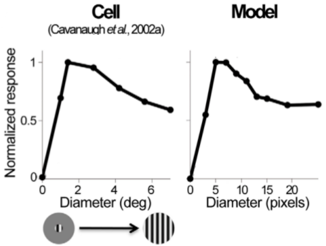
GSM $\Rightarrow$ normalization:

$$X = (X^{(c)}, X^{(s)}), \quad G = (G^{(c)}, G^{(s)})$$

$$G^{(c)} \propto \frac{X^{(c)}}{\sqrt{\nu + \sum_k w_k X_k^2}}$$

Gaussian Scale Mixture explains surround modulation in V1





Taken from Coen-Cagli, Dayan, et al. 2012

GSM $\Rightarrow$ normalization:

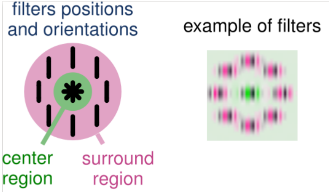
$$X = (X^{(c)}, X^{(s)}), \quad G = (G^{(c)}, G^{(s)})$$

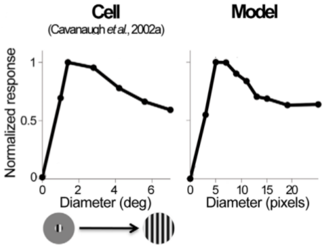
$$G^{(c)} \propto \frac{X^{(c)}}{\sqrt{\nu + \sum_k w_k X_k^2}}$$

Interpretation:

- G: vector of neurons responses (Coen-Cagli and Schwartz 2013; Orbán et al. 2016)
- Z: normalization (canonical across the brain Carandini et al. 2012)

Gaussian Scale Mixture explains surround modulation in V1





Taken from Coen-Cagli, Dayan, et al. 2012

GSM $\Rightarrow$ normalization:

$$X = (X^{(c)}, X^{(s)}), \quad G = (G^{(c)}, G^{(s)})$$

$$G^{(c)} \propto \frac{X^{(c)}}{\sqrt{\nu + \sum_k w_k X_k^2}}$$

Interpretation:

- G: vector of neurons responses (Coen-Cagli and Schwartz 2013; Orbán et al. 2016)
- Z: normalization (canonical across the brain Carandini et al. 2012)

- A normative model: vision is adapted to environmental statistics !

Images statistics and V1 physiology suggest Mixtures of GSMs

Key question: How are neurons pooled together (*i.e.* normalized together) ?

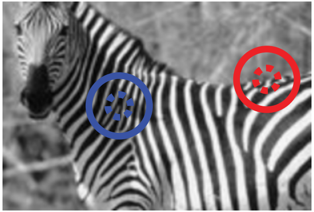
Center+Surround or Center/Surround pooling

Images statistics and V1 physiology suggest Mixtures of GSMs

Key question: How are neurons pooled together (*i.e.* normalized together) ?

Using a flexible pooling of neurons:

Center+Surround or Center/Surround pooling



Images statistics and V1 physiology suggest Mixtures of GSMs

Key question: How are neurons pooled together (*i.e.* normalized together) ?

Using a flexible pooling of neurons:

- qualitatively different image patches (Schwartz et al. 2006)

Center+Surround or Center/Surround pooling

Homogeneous:
pooled together

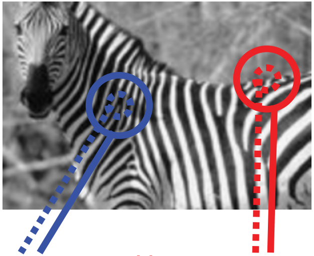
Images statistics and V1 physiology suggest Mixtures of GSMs

Key question: How are neurons pooled together (*i.e.* normalized together) ?

Using a flexible pooling of neurons:

- qualitatively different image patches (Schwartz et al. 2006)

Center+Surround or Center/Surround pooling



Homogeneous: pooled together

Heterogeneous: kept separate

Images statistics and V1 physiology suggest Mixtures of GSMs

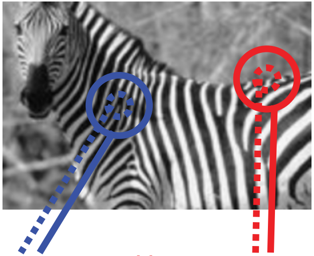
Key question: How are neurons pooled together (*i.e.* normalized together) ?

Using a flexible pooling of neurons:

- qualitatively different image patches (Schwartz et al. 2006)
- better explanation of some neurons activity (Coen-Cagli, Kohn, et al. 2015)

Center+Surround or Center/Surround pooling

- Homogeneous: strong surround suppression



- Heterogeneous: weak surround suppression

Homogeneous: pooled together

Heterogeneous: kept separate

Images statistics and V1 physiology suggest Mixtures of GSMs

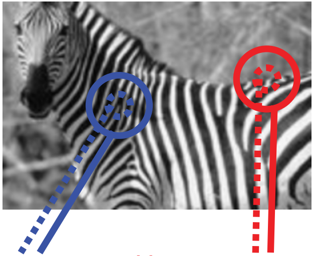
Key question: How are neurons pooled together (*i.e.* normalized together) ?

Using a flexible pooling of neurons:

- qualitatively different image patches (Schwartz et al. 2006)
- better explanation of some neurons activity (Coen-Cagli, Kohn, et al. 2015)

Center+Surround or Center/Surround pooling

- Homogeneous: strong surround suppression
- basis for segmentation ?



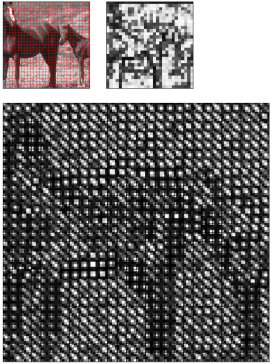
- Heterogeneous: weak surround suppression
- basis for contour detection ?

Homogeneous: pooled together

Heterogeneous: kept separate

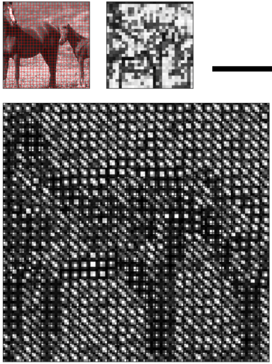
Natural images statistics suggest Mixtures of GSMs

- Using a fixed pooling of neurons:



Natural images statistics suggest Mixtures of GSMs

- Using a fixed pooling of neurons:

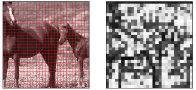


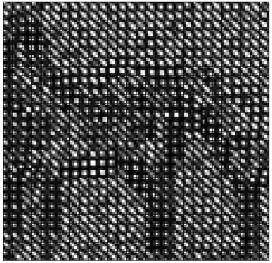
Normalization statistics vary accross an image

$Z \sim \mathcal{L}(\nu)$

Natural images statistics suggest Mixtures of GSMs

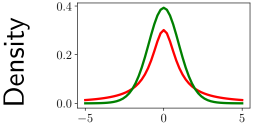
- Using a fixed pooling of neurons:





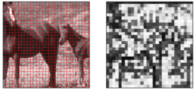
Normalization statistics vary accross an image

$Z \sim \mathcal{L}(\nu)$



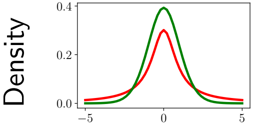
Natural images statistics suggest Mixtures of GSMs

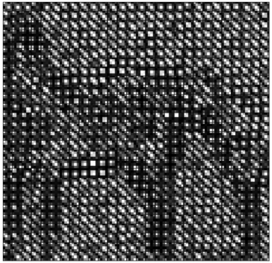
- Using a fixed pooling of neurons:



Normalization statistics vary accross an image

$Z \sim \mathcal{L}(\nu)$





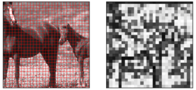
Covariances vary accross an image

$G \sim \mathcal{N}(0, \Sigma)$

jonathan.vacher@einstein.yu.edu https://jonathanvacher.github.io

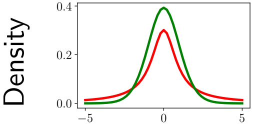
Natural images statistics suggest Mixtures of GSMs

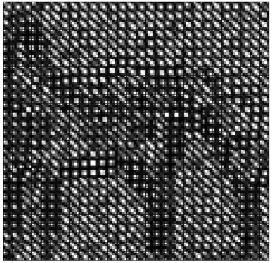
- Using a fixed pooling of neurons:



Normalization statistics vary accross an image

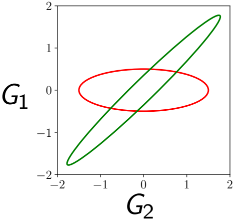
$Z \sim \mathcal{L}(\nu)$



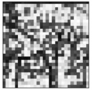


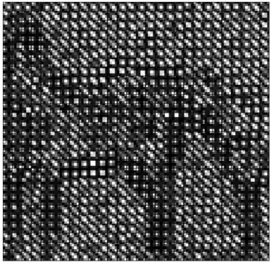
Covariances vary accross an image

$G \sim \mathcal{N}(0, \Sigma)$

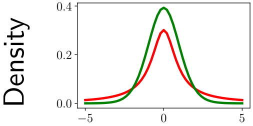


- Using a fixed pooling of neurons:



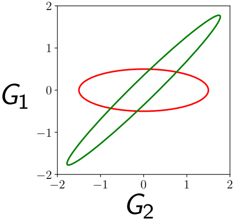


Normalization statistics vary accross an image
$Z \sim \mathcal{L}(\nu)$



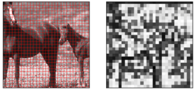
Natural images are not stationary !

Covariances vary accross an image
$G \sim \mathcal{N}(0, \Sigma)$

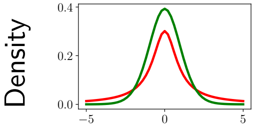


Natural images statistics suggest Mixtures of GSMs

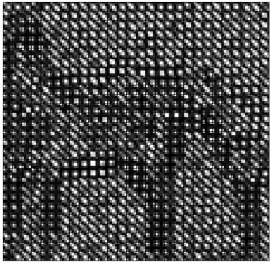
- Using a fixed pooling of neurons:



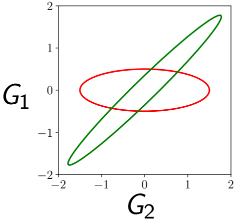
Normalization statistics vary accross an image
$Z \sim \mathcal{L}(\nu)$



Natural images are not stationary !



Covariances vary accross an image
$G \sim \mathcal{N}(0, \Sigma)$



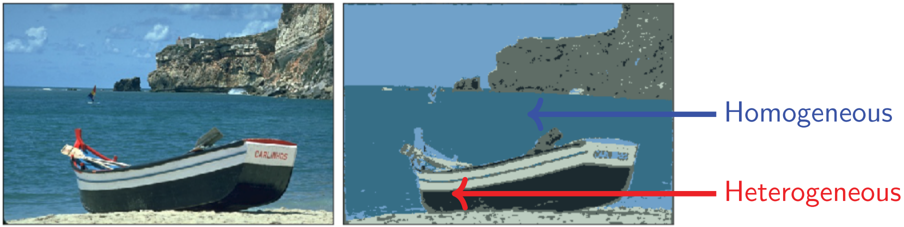
Two reasons for Mixture of GSMs:
- Natural images are non-stationary (statistics vary across space)
- Flexible pooling seems necessary (for image stats and physiology)

Naive Mixture of Gaussian Scale Mixtures for image segmentation

Using wavelet feature vectors at a single pixel location we obtain:

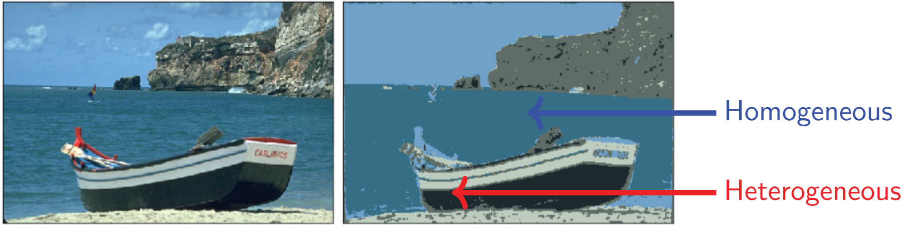
Naive Mixture of Gaussian Scale Mixtures for image segmentation

Using wavelet feature vectors at a single pixel location we obtain:



Naive Mixture of Gaussian Scale Mixtures for image segmentation

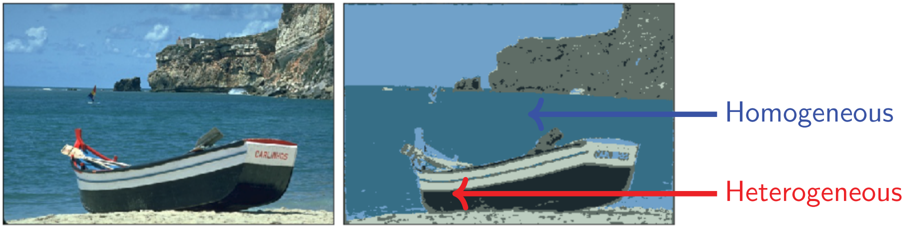
Using wavelet feature vectors at a single pixel location we obtain:



Decomposing the components:

Naive Mixture of Gaussian Scale Mixtures for image segmentation

Using wavelet feature vectors at a single pixel location we obtain:

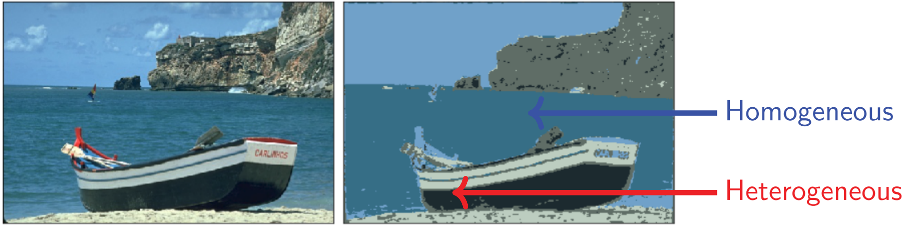


Decomposing the components:

- homogeneous: smoothing ($\sim$ proximity grouping), see our pre-prints: texture-based segmentation algorithm (Vacher, Mamassian, et al. 2019) + extension using hierarchical features (Vacher and Coen-Cagli 2019). Can achieve state-of-the-art performances.

Naive Mixture of Gaussian Scale Mixtures for image segmentation

Using wavelet feature vectors at a single pixel location we obtain:



Decomposing the components:

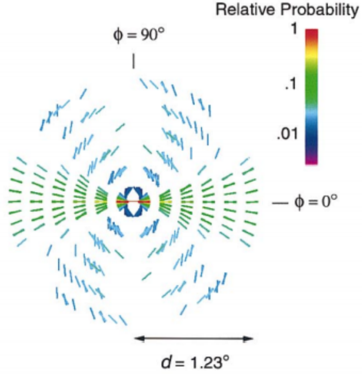
- homogeneous: smoothing ($\sim$ proximity grouping), see our pre-prints: texture-based segmentation algorithm (Vacher, Mamassian, et al. 2019) + extension using hierarchical features (Vacher and Coen-Cagli 2019). Can achieve state-of-the-art performances.
- heterogeneous: forcing the covariance structure ($\sim$ flexible pooling)

Heterogeneous component !

Heterogeneous: kept separate

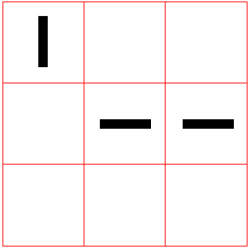
Heterogeneous component !

Edge co-occurrence (Geisler et al. 2001)



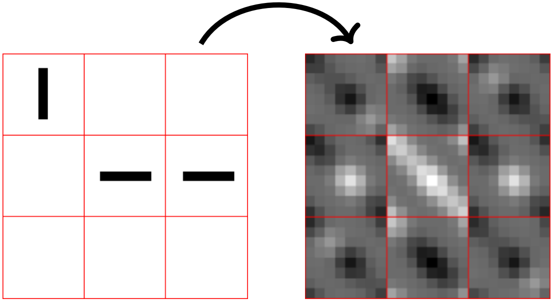
Heterogeneous component !

- Train a GSM on many natural images using center-surround feature vectors



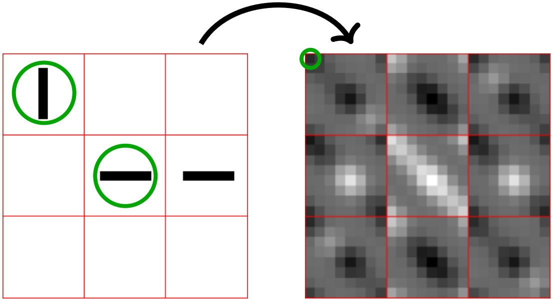
Heterogeneous component !

- Train a GSM on many natural images using center-surround feature vectors



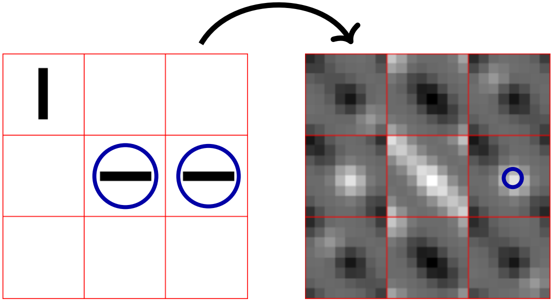
Heterogeneous component !

- Train a GSM on many natural images using center-surround feature vectors



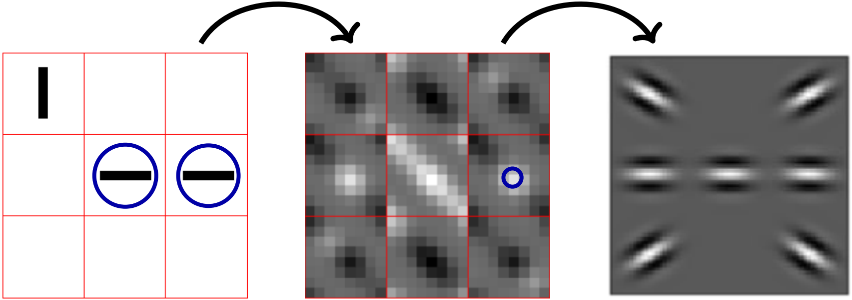
Heterogeneous component !

- Train a GSM on many natural images using center-surround feature vectors



Heterogeneous component !

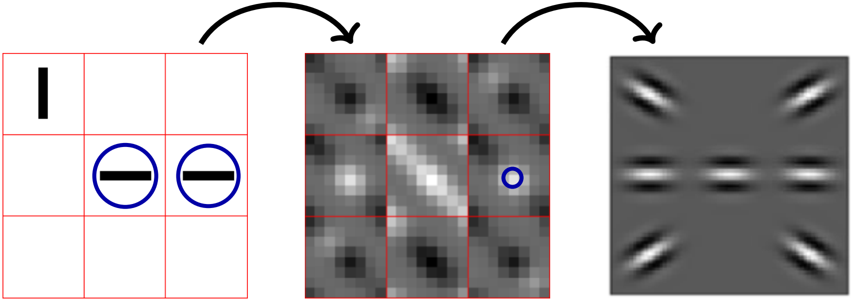
- Train a GSM on many natural images using center-surround feature vectors



- The covariance contains the association field structure !

Heterogeneous component !

- Train a GSM on many natural images using center-surround feature vectors



- The covariance contains the association field structure !
- However not strong enough to distinguish contour from non-contour $\Rightarrow$ enforcing association field

Contour component *vs* “garbage” component

Contour component *vs* "garbage" component

- Forcing the association field structure $\Rightarrow$ block diagonal covariance

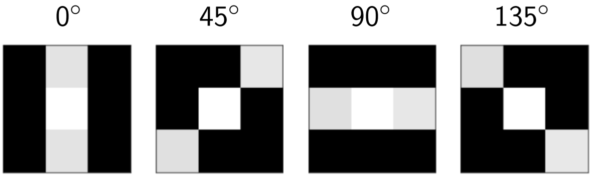
Contour component *vs* “garbage” component

- Forcing the association field structure $\Rightarrow$ block diagonal covariance
- Equivalent to specify a linear subspace *i.e.* a basis $\Rightarrow$ Principal Components (PCA)

Contour component *vs* "garbage" component

- Forcing the association field structure $\Rightarrow$ block diagonal covariance
- Equivalent to specify a linear subspace *i.e.* a basis $\Rightarrow$ Principal Components (PCA)

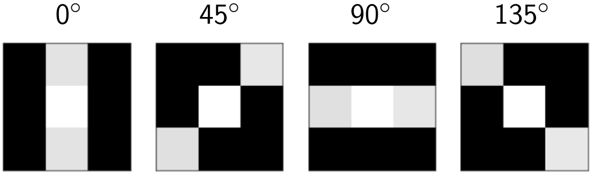
Principal component of the block diagonal covariance trained on human labeled edges only:



Contour component *vs* “garbage” component

- Forcing the association field structure $\Rightarrow$ block diagonal covariance
- Equivalent to specify a linear subspace *i.e.* a basis $\Rightarrow$ Principal Components (PCA)

Principal component of the block diagonal covariance trained on human labeled edges only:

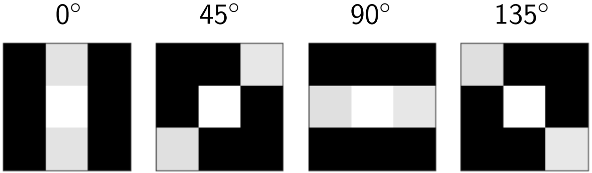


- In practice, it’s better to project the feature vectors onto this subspace before training.

Contour component *vs* “garbage” component

- Forcing the association field structure ⇒ block diagonal covariance
- Equivalent to specify a linear subspace *i.e.* a basis ⇒ Principal Components (PCA)

Principal component of the block diagonal covariance trained on human labeled edges only:



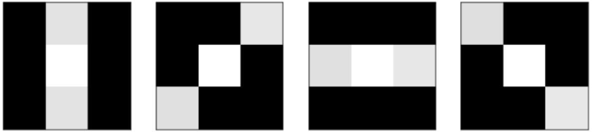
- In practice, it’s better to project the feature vectors onto this subspace before training.
- Similar to the template matching framework (Geisler 2018; Sebastian et al. 2017)

Contour component *vs* "garbage" component

- Forcing the association field structure $\Rightarrow$ block diagonal covariance
- Equivalent to specify a linear subspace *i.e.* a basis $\Rightarrow$ Principal Components (PCA)

Principal component of the block diagonal covariance trained on human labeled edges only:

0° 45° 90° 135°



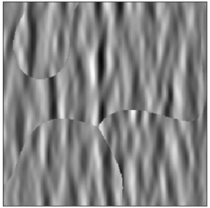
- In practice, it's better to project the feature vectors onto this subspace before training.
- Similar to the template matching framework (Geisler 2018; Sebastian et al. 2017)
- Also, coherent with expected long edge receptive fields in the higher visual cortex (V2, V4) (Hosoya et al. 2015; Liu et al. 2016)

Contour component *vs* "garbage" component: results

Encouraging results on different types of images:

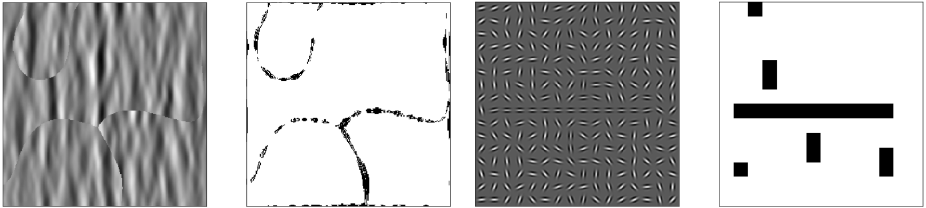
Contour component *vs* "garbage" component: results

Encouraging results on different types of images:



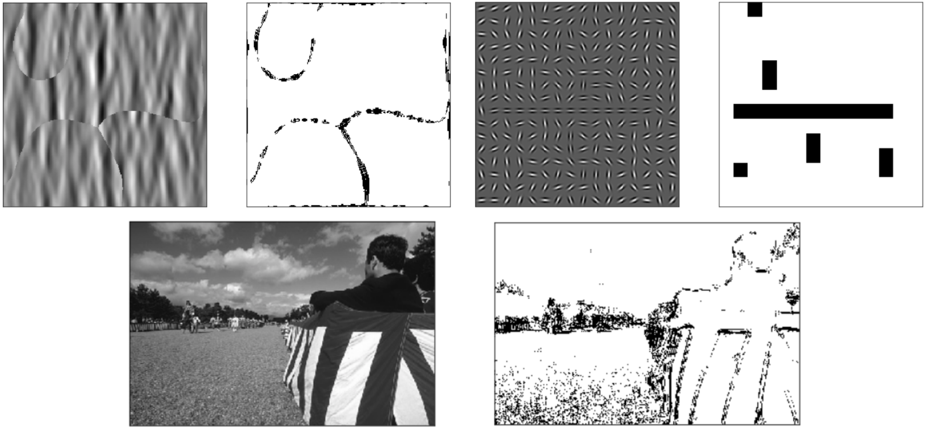
Contour component *vs* “garbage” component: results

Encouraging results on different types of images:



Contour component *vs* “garbage” component: results

Encouraging results on different types of images:



Conclusion

- probabilistic model that accounts for image statistics, physiology and psychophysics
- need to improve the contour-based model and quantify its performance

Conclusion

- probabilistic model that accounts for image statistics, physiology and psychophysics
- need to improve the contour-based model and quantify its performance
- Next step: compare model predictions to human segmentation maps (see you next year!)

Conclusion

- probabilistic model that accounts for image statistics, physiology and psychophysics
- need to improve the contour-based model and quantify its performance
- Next step: compare model predictions to human segmentation maps (see you next year!)

Thanks to Pascal Mamassian and Ruben Coen-Cagli