CRV 2010

Unsupervised Feature Selection and Learning for Image Segmentation

M.S. Allili*, D. Ziou** N. Bouguila†, and S. Boutemedjet‡

*: Université du Québec en Outaouais.

**:Université de Sherbrooke.

†:Concordia University.

‡:École des Technologies suprérieures.

Plan

-Introduction

-Feature clustering for segmentation

-Feature selection for segmentation

-Some results

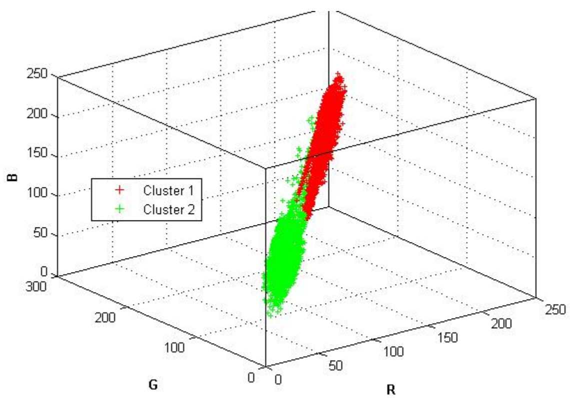
-Conclusions and future work

Introduction

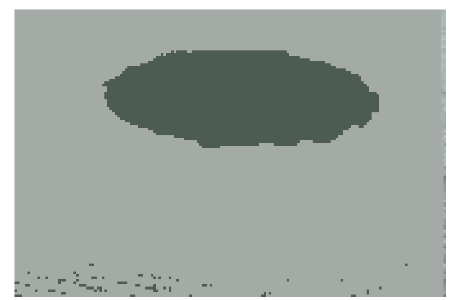
Feature clustering is a popular approach for segmentation and grouping.

For image egmentation, clustering is performed in order to subdivide the feature spcace into dense clusters, which (presumably) correspond to homogeneous regions in the image.

Original image



Cluters in feature space



Homogeneous regions

Motivation of the problem

Finite Mixture models (FMM) has been proved as a powerful technique for clustering. It has been successfully used for segmentation.

FMM try to model the distribution of each cluster in the data using a probability density function (pdf).

The image data can then be described by a mixture of these densities.

$$p(x) = \sum_{j=1}^{K} w_j \cdot p(x \mid \theta_j)$$

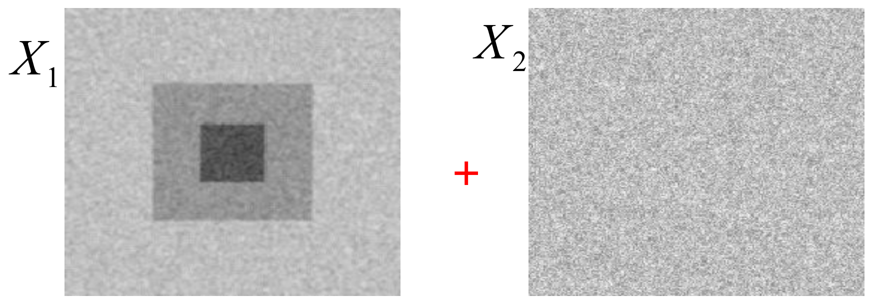
Motivation of the problem

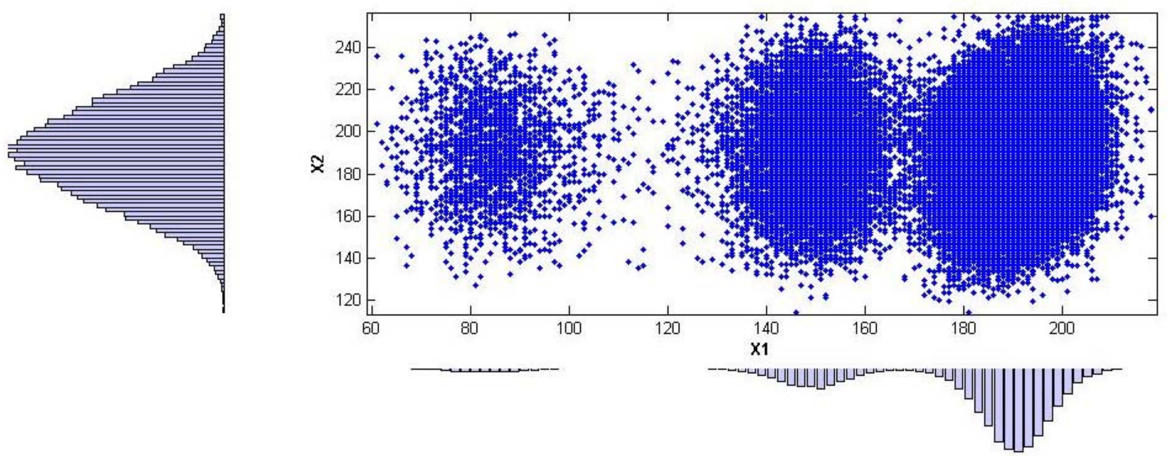
Motivation of the problem:

What is the best way to the combine several features to yield optimal regions (clusters) in the image?

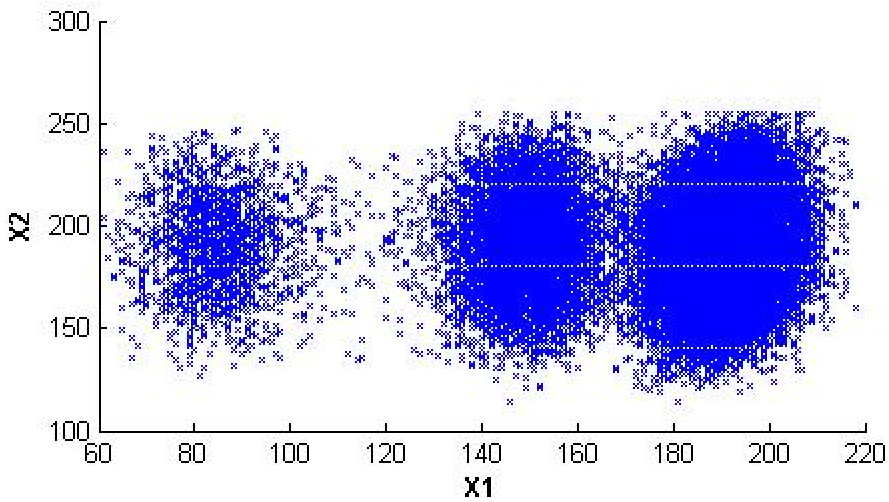
uniform

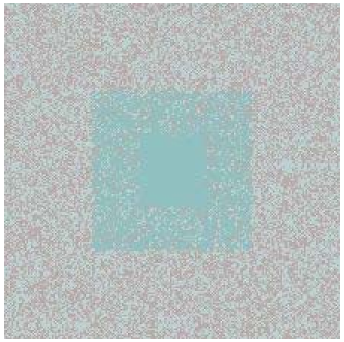
Example:



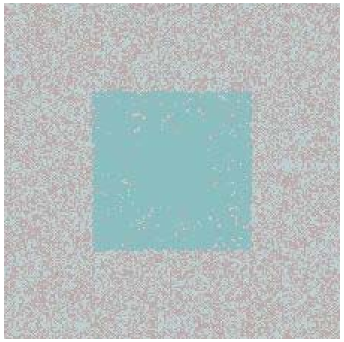


Motivation of the problem

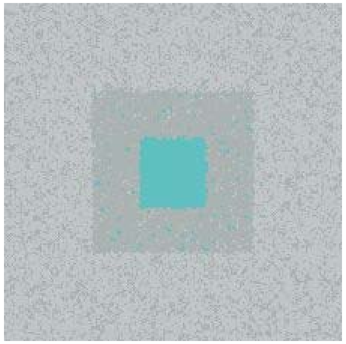




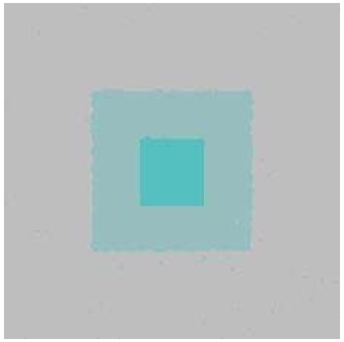
$\alpha_1 = 0.3 \quad \alpha_2 = 0.7$



$\alpha_1 = 0.4 \quad \alpha_2 = 0.6$

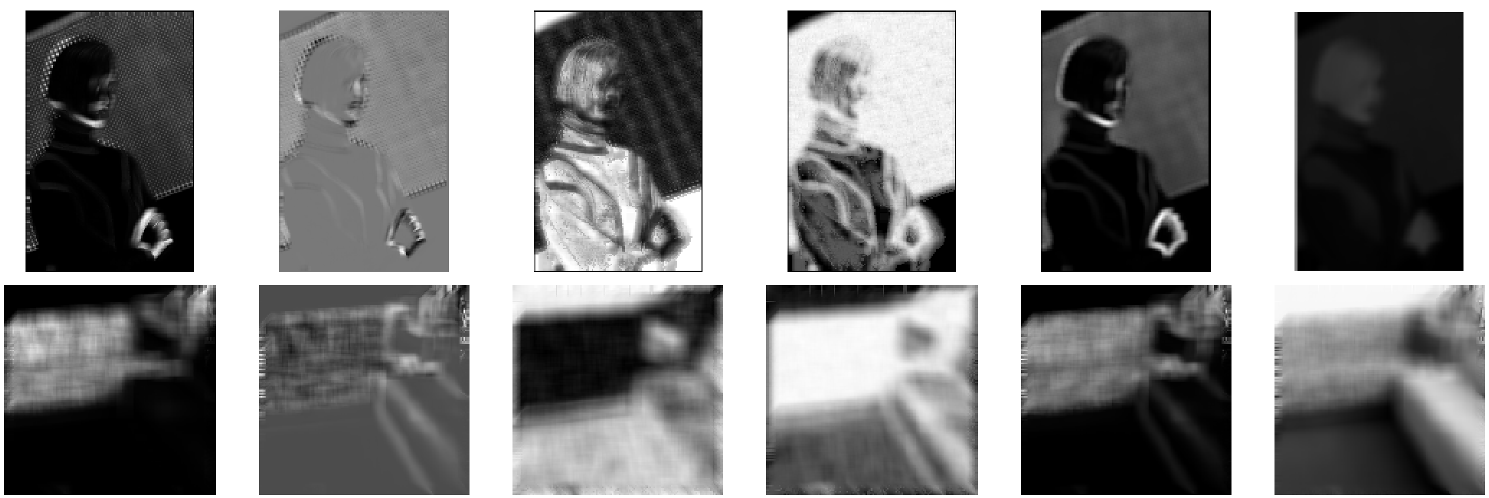


$\alpha_1 = 0.5 \quad \alpha_2 = 0.5$



$\alpha_1 = 0.8 \quad \alpha_2 = 0.2$

Motivation of the problem



CT CR EN HG VR ICM

Proposed solution

Introduce feature selection($saliency$) in segmentation.

We assume a feature vector $\vec{x} = (x_1, \ldots, x_d)^T$ calculated from the image (composed of color and texture).

Without taking into account the feature $saliency$, the mixture model fitting the data is given by :

$$p(\vec{x}) = \sum_{j=1}^{K} w_j \cdot p(\vec{x} \mid \theta_j)$$

Proposed solution: Feature Selection (saliency)

The relevance of a feature is a random binary variable $\phi_\ell \in \{0,1\}_{l=1}^{d}$. The distribution of ℓth feature in component j is

$$p(x_\ell \mid \theta_{j\ell}, \varphi_\ell) \approx \left[p(x_\ell \mid \theta_{j\ell})\right]^{\phi_\ell} \left[p(x_\ell \mid \varphi_\ell)\right]^{1-\phi_\ell}$$

$p(x_\ell \mid \theta_{j\ell})$ and $p(x_\ell \mid \varphi_\ell)$ are 2 univariate general Gaussians.

Mixture of GGDs with feature selection

$p(\phi_\ell = 1)$ denotes the probability of relevance of the ℓth feature

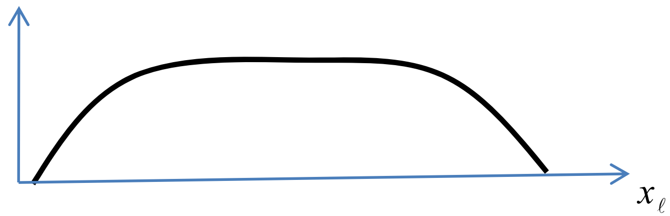
$$\rho_\ell = p(\phi_\ell = 1)$$

Now, the mixture model with feature selection becomes:

$$p(\vec{x}) = \sum_{j=1}^{M} w_j \prod_{\ell=1}^{d} \left[\rho_{\ell 1} p(x_\ell \mid \theta_{j\ell}) + (1-\rho_{\ell 1}) p(x_\ell \mid \varphi_\ell) \right]$$

Use a secondary mixture to model the non-relevant features.

$$p(x_\ell \mid \varphi_\ell) = \sum_{k=1}^{K} \pi_k p(x_\ell \mid \varphi_{k\ell})$$



Mixture of GGDs with feature selection

We end up with the following general model:

$$p(\vec{x}) = \sum_{j=1}^{M} w_j \prod_{\ell=1}^{d} \left[\rho_{\ell 1} p(x_\ell \mid \theta_{j\ell}) + (1-\rho_{\ell 1}) \left(\sum_{k=1}^{K} \pi_{k\ell} p(x_\ell \mid \varphi_{j\ell}) \right) \right]$$

Problem:

Estimate the model parameters (θ, φ, K, M)?

Model selection

Solution: Use the minimum message length principle:

$$(\Theta^*, M^*, K^*) = \arg\min_{\Theta, M, K} \{MessLen(\Theta, M, K)\} \qquad \Theta = (\theta, \varphi)$$

$$MessLen = -\log p(\Theta) - \log p(X \mid \Theta) + \frac{1}{2}\log|I(\Theta)| + \frac{c}{2}(1 - \log(12))$$

$c = M + d + 3dM + 4dK$ is the total number of parameters

Model Learning

We use the Expectation-Maximization (EM) algorithm to minimize iteratively $MessLen$ w.r.t (Θ, K, M).

Over-initialize EM with

$$M = M_{\text{Max}} \text{ and } K = K_{\text{Max}}.$$

After convergence, find optimal values of (K, M) and Θ including the relevance of the features ρ_ℓ.

Performance testing

$Boundary\ \ alignment\ \ error(\varepsilon_1):$

$$\varepsilon_1 = \frac{1}{N}\sum_{(u,v)\in I}\min\{\varepsilon_{u,v}(S_G,S_T),\varepsilon_{u,v}(S_T,S_G)\}$$

where $\varepsilon_{u,v}(S_G,S_T) = \dfrac{|S_G - S_T|}{|S_G|}$ and $\varepsilon_{u,v}(S_T,S_S) = \dfrac{|S_T - S_G|}{|S_T|}$

$Amount\ \ of\ \ over/under\text{-}segmentation\ \ (\varepsilon_2):$

$$\varepsilon_2 = \text{ \# of over-segmented segments.}$$

S_G : Segment in the ground truth.

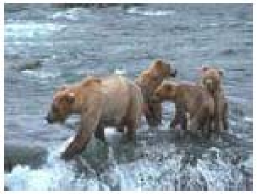
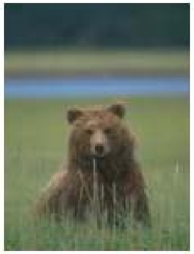
S_T : Segment in a tested segmentation.

Some results

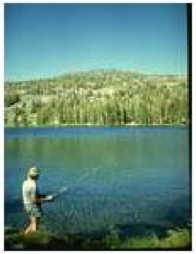
We tested our model for image and video segmentation.

We used the Berkeley Benchmark dataset and known test videos:

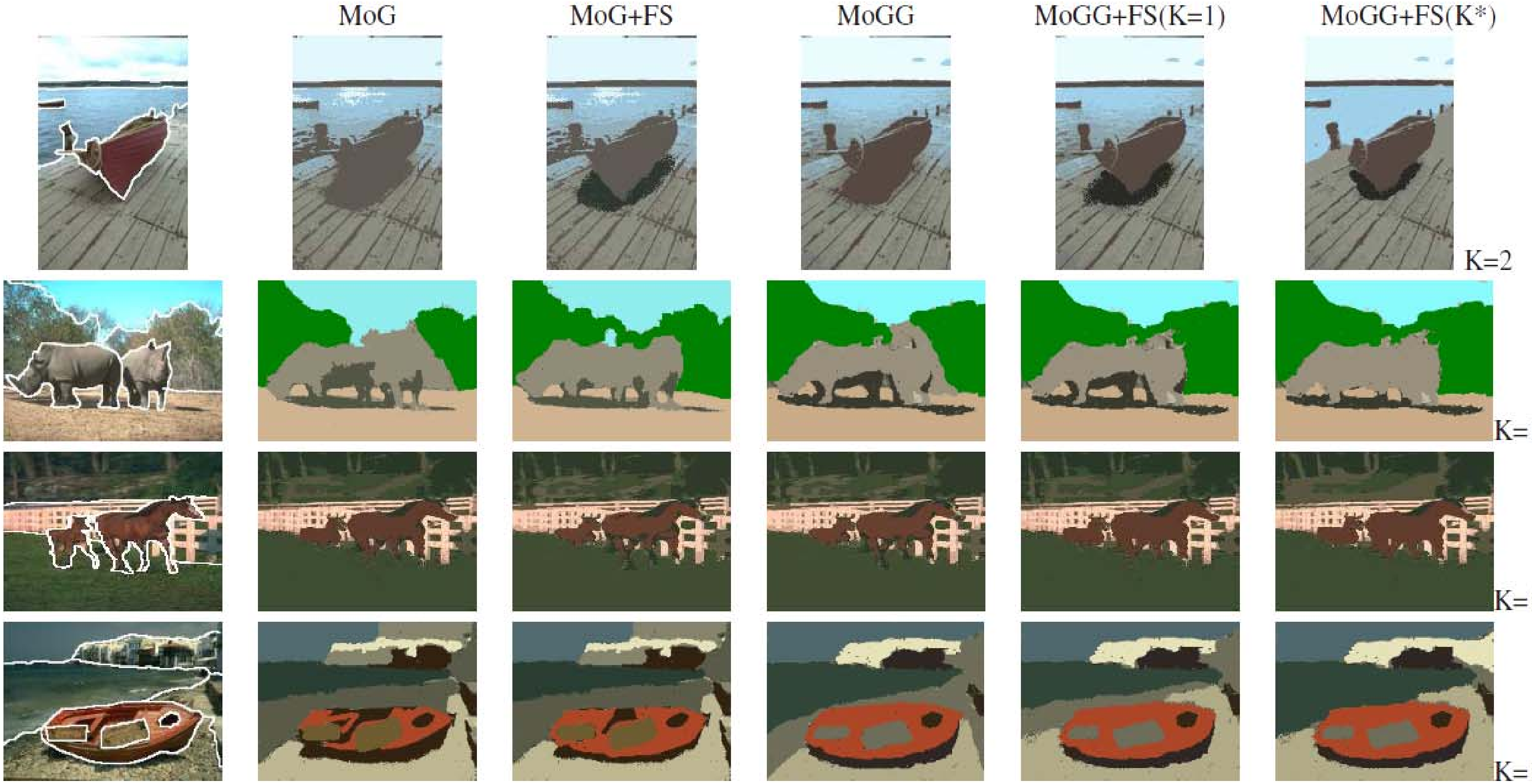







Some results



Some results

Some results

Image segmentation

$[\bar{\epsilon}_1, \bar{\epsilon}_2]$				
MoG	MoG+FS	MoGG	MoGG+FS (K=1)	MoGG+FS (K=*)
[0.21, 23]	[0.15, 16]	[0.19, 20]	[0.13, 13]	**[0.10, 11]**

Video segmentation

Video	Size	$[\bar{\epsilon}_1, \bar{\epsilon}_2]$				
		MoG	MoG+FS	MoGG	MoGG+FS(K=1)	MoGG+FS(K=*)
Akiyo	300 frames	[0.23,22.5]	[0.16,17.8]	[0.22,20.5]	[0.12,13.2]	[0.12,13.2]
suzie	150 frames	[0.25,27.4]	[0.18,19.1]	[0.24,23.7]	[0.15,17.5]	[0.12,11.3]
Grandma	870 frames	[0.19,20.5]	[0.15,14.8]	[0.17,19.3]	[0.13,14.5]	[0.12,13.9]
Irene	300 frames	[0.27,28.2]	[0.19,20.8]	[0.23,26.6]	[0.16,17.3]	[0.13,14.4]

Conclusion and future work

- We introduced an unsupervised model for feature selection and clustering for segmentation.

- Feature selection is important for mitigating over-segmentation due to uniform (non-relevant) features.

- Use more features for more accurate segmentation.

- Use feature selection for object segmentation.

Thank you!

Questions?