Model based clustering using color and depth information

Abul Hasnat (PhD student), **Olivier Alata**, Alain Trémeau

Lab. Hubert Curien, UMR CNRS 5516
Univ. Jean Monnet Saint-Etienne, Univ. Lyon, F-42000, Saint-Etienne, France

Journée SIERRA - Méthodes et modèles adaptatifs
the 25th of March, 2014

Outline

Outline

Exponential families (EF) of distributions

$$f(x|\theta) = \exp\{\langle\theta, t(x)\rangle - F(\theta) + g(x)\}, x \in \Omega_{X_s}$$

with

- $t(x)$, the sufficient statistic,
- $\theta \in \mathcal{P}_\Theta$, with dimension D (the order of the family), the natural parameter
- $F(.)$, the log-normalizer, $F(\theta) = \log \int_{\Omega_{X_s}} \exp(\langle\theta, t(x)\rangle + g(x))\,\mathrm{d}x$
- $g(.)$, the carrier measure.
- $\eta = \eta(\theta) = \nabla F(\theta)$, the expectation parameters.

Examples : Gaussian, Wishart, Poisson, Rayleigh, ... laws (see [1])

Exponential families (EF) of distributions

$$f(x|\theta) = \exp\{\langle\theta, t(x)\rangle - F(\theta) + g(x)\}, x \in \Omega_{X_s}$$

with

- $t(x)$, the sufficient statistic,
- $\theta \in \mathcal{P}_\Theta$, with dimension D (the order of the family), the natural parameter
- $F(.)$, the log-normalizer, $F(\theta) = \log \int_{\Omega_{X_s}} \exp(\langle\theta, t(x)\rangle + g(x))\,\mathrm{d}x$
- $g(.)$, the carrier measure.
- $\eta = \eta(\theta) = \nabla F(\theta)$, the expectation parameters.

Examples : Gaussian, Wishart, Poisson, Rayleigh, ... laws (see [1])

Example: univariate Gaussian distribution

- $t(x) = (x, x^2)$, $m = (\mu, \sigma^2)$, $\theta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right) = (\theta_1, \theta_2)$
- $F(\theta) = \left(-\frac{\theta_1^2}{4\theta_2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta_2}\right)\right)$
- $g(x) = 0$
- $\eta = (\eta_1, \eta_2) = \left(-\frac{\theta_1}{2\theta_2}, -\frac{1}{2\theta_2} + \frac{\theta_1^2}{4\theta_2^2}\right) = (\mu, \sigma^2 + \mu^2)$

EF and Bregman Divergence (BD) (see [1] and [2])

$$f\left(x\,|\theta\right) = \exp\left\{-d_{F^*}\left(t(x),\eta\right)\right\} b_{F^*}(t(x)), x \in \Omega_{X_s}$$

- d_{F^*}, the Bregman divergence associated with F^*, the conjugate function (Legendre dual) of F, which is a strictly convex function :

$$d_{F^*}(\eta_1,\eta_2) = F^*(\eta_1) - F^*(\eta_2) - \langle \eta_1 - \eta_2, \nabla F^*(\eta_2)\rangle = d_F(\theta_2,\theta_1)$$

  - For EF, $KL(\theta_1,\theta_2) = d_{F^*}(\eta_1,\eta_2) = d_F(\theta_2,\theta_1)$
  - $F^*(\eta) = \sup\limits_{t\in\mathcal{P}_\Theta} \{\langle \eta, t\rangle - F(t)\}$
  - $\theta = \theta(\eta) = \nabla F^*(\eta)$
  - $b_{F^*}(t(x)) = \exp\left(F^*(t(x)) + g(x)\right)$

EF and Bregman Divergence (BD) (see [1] and [2])

$$f\left(x \,|\theta\right) = \exp\left\{-d_{F^*}\left(t(x),\eta\right)\right\} b_{F^*}(t(x)), x \in \Omega_{X_s}$$

- d_{F^*}, the Bregman divergence associated with F^*, the conjugate function (Legendre dual) of F, which is a strictly convex function :

  $$d_{F^*}(\eta_1,\eta_2) = F^*(\eta_1) - F^*(\eta_2) - \langle \eta_1 - \eta_2, \nabla F^*(\eta_2)\rangle = d_F(\theta_2,\theta_1)$$

  - For EF, $KL(\theta_1,\theta_2) = d_{F^*}(\eta_1,\eta_2) = d_F(\theta_2,\theta_1)$
  - $F^*(\eta) = \sup_{t\in\mathcal{P}_\Theta} \{\langle \eta, t\rangle - F(t)\}$
  - $\theta = \theta(\eta) = \nabla F^*(\eta)$
  - $b_{F^*}(t(x)) = \exp\left(F^*(t(x)) + g(x)\right)$

Example: univariate Gaussian distribution

- $F^*(\eta) = -\frac{1}{2}\log\left(\eta_1^2 - \eta_2\right) + C$ avec $\eta = (\eta_1,\eta_2) = (\mu, \sigma^2 + \mu^2)$

Bregman divergence and information geometry [1, 2, 3]

- For different convex functions:
  - Quadratic form distances (Euclidean, Mahalanobis, ...) which are the only symmetric Bregman divergences.
  - Kullback-Leibler divergence (KL),
  - ...
- For two members, θ_1 and θ_2, of the same exponential family :

$$KL(\theta_1, \theta_2) = \int f(x|\theta_1) \log\left(\frac{f(x|\theta_1)}{f(x|\theta_2)}\right) \mathrm{d}x = d_F(\theta_2, \theta_1) = d_{F^*}(\eta_1, \eta_2)$$

- Meaning of Kullback-Leibler divergence:
  - Relative entropy between $f(x|\theta_1)$ and $f(x|\theta_2)$
  - Cencov [4] proved that the only Riemannian metric that "makes sense" for statistical manifolds is the Fisher information metric.
  - When θ_2 is closed to θ_1

$$2KL(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_I^2 (1 + o(1)),$$

with $\|.\|_I$ the norm based on the Fisher information metric.

BD and information geometry [3]

- Computation of centroids for K parametric distributions:
  - left-sided centroid (for natural parameters)

$$\theta_L = \arg\min_{\theta \in \mathcal{P}_\Theta} \frac{1}{K} \sum_{k=1}^{K} d_F(\theta, \theta_k) = \nabla F^* \left(\frac{\sum_{k=1}^{K} \alpha_k \nabla F(\theta_k)}{\sum_{k=1}^{K} \alpha_k} \right)$$

  - right-sided centroid (for natural parameters)

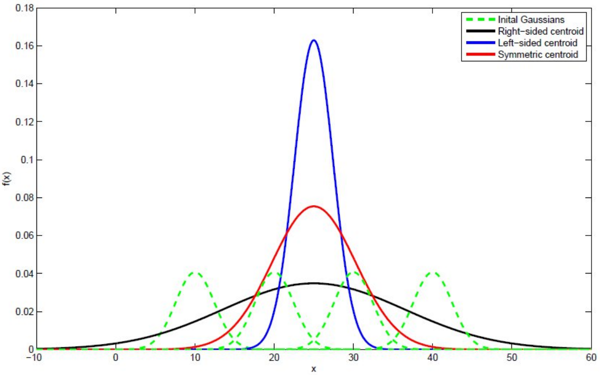
$$\theta_R = \arg\min_{\theta \in \mathcal{P}_\Theta} \frac{1}{K} \sum_{k=1}^{K} d_F(\theta_k, \theta) = \frac{\sum_{k=1}^{K} \alpha_k \theta_k}{\sum_{k=1}^{K} \alpha_k}$$

  - symmetric centroid (belongs to the geodesic link between θ_L and θ_R)

$$\theta_S = \arg\min_{\theta \in \mathcal{P}_\Theta} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} \left(d_F(\theta, \theta_k) + d_F(\theta_k, \theta) \right)$$

Computation of Bregman centroids

Example in a Gaussian univariate case [5]



- Initial set contains 4 univariate Gaussians $\sigma^2 = 6$
- Right-sided centroid
- Left-sided centroid
- Symmetric centroid

Mixture model

- $x = \{x_s\}_{s \in \Lambda}$, samples of **i.i.d.** (independently and identically distributed) random vectors with dimension p.

$$f(x) = \prod_{s \in \Lambda} f(x_s)$$

- **General hypothesis**: the distribution of each sample is a mixture with K components.

$$\begin{aligned} f(x_s) = f_K(x_s) &= \sum_{k \in K} f(x_s, k) && (1) \\ &= \sum_{k \in K} f(x_s | \theta_k) P(k) && (2) \end{aligned}$$

  - K number of components, **unknown a priori**.
  - $\alpha_k = P(k)$ "a priori" probability of kth component, $\sum_{k=1}^{K} \alpha_k = 1$.
  - θ_k natural parameters of component k and $\Theta_K = \{\alpha_k, \theta_k\}_{k=1,\cdots,K}$.

Approach developped in [1, 2, 5]

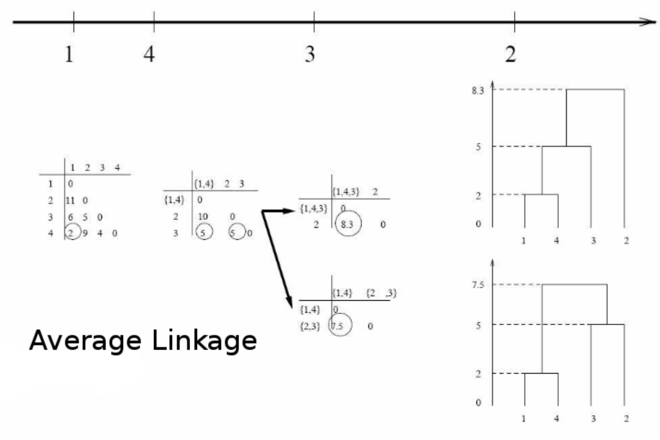
- Initialization:
  - Fix $K = K_{\max}$, the maximum number of components.
  - Bregman hard clustering (kmeans based on Bregman divergence)

  $\rightarrow$ First estimation of $\{\alpha_{l,K_{\max}}, \theta_{l,K_{\max}}\}_{l=1,\cdots,K_{\max}}$.
- Bregman soft clustering (An EM type algorithm based on Bregman divergence)

  $\rightarrow$ Final estimation of $\{\alpha_{l,K_{\max}}, \theta_{l,K_{\max}}\}_{l=1,\cdots,K_{\max}}$.
- Bregman hierarchical clustering (**fast simplification process**)
  - Choice of the Bregman divergence and associated cendroid
  - Choice of the linkage criterion

  $\rightarrow$ **Estimation of** K ($\hat{K}$ using IC) **and** $\{\alpha_{l,\hat{K}}, \theta_{l,\hat{K}}\}_{l=1,\cdots,\hat{K}}$.

Hierarchical clustering



IC and principle of parsimony

- A formulation justified by theory of information.

$$IC(K) = \underbrace{-2\log f\left(x \,\middle|\, \hat{\Theta}_K^{ML}, K\right)}_{Representation} + \underbrace{p\left(\left|\Theta_K\right|, N\right)}_{Complexity}$$

$$\hat{K}^{IC} = \arg\min_K IC(K)$$

ML for Maximum Likelihood. N is the number of samples.

IC and principle of parsimony

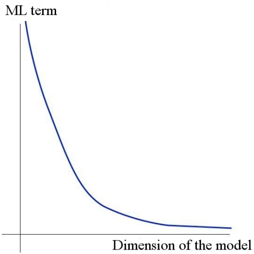
- A formulation justified by theory of information.

$$IC(K) = \underbrace{-2 \log f\left(x \,\middle|\, \hat{\Theta}_K^{ML}, K\right)}_{Representation} + \underbrace{p\left(\left|\Theta_K\right|, N\right)}_{Complexity}$$

$$\hat{K}^{IC} = \arg\min_K IC(K)$$

ML for Maximum Likelihood. N is the number of samples.

- Typical curves obtained in the case of nested models:



IC and principle of parsimony

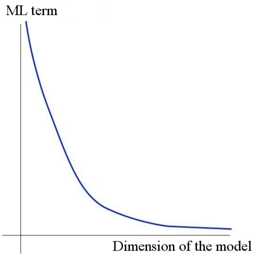
- A formulation justified by theory of information.

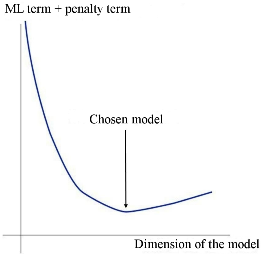
$$IC(K) = \underbrace{-2\log f\left(x \,\middle|\, \hat{\Theta}_K^{ML}, K\right)}_{Representation} + \underbrace{p\left(\left|\Theta_K\right|, N\right)}_{Complexity}$$

$$\hat{K}^{IC} = \arg\min_K IC(K)$$

ML for Maximum Likelihood. N is the number of samples.

- Typical curves obtained in the case of nested models:





Criteria for the estimation of K (see [6])

- "Classical" form:

$$IC(K) = -2\log\left(f\left(x\middle|\hat{\Theta}_K\right)\right) + C(N)\,P(K)$$

$P(K)$ number of free parameters.

  - AIC (Akaike Information Criterion, 1974) : $C(N) = 2$
    derived using KL divergence
  - BIC (Bayesian Information Criterion, Schwarz 1978) : $C(N) = \log N$
  - ϕ_β (El Matouat, Hallin 1996) : $C(N) = N^\beta \log\log(N)$ avec $\beta = \frac{\log\log N}{\log N}$.

- Other forms:
  - ICOMP (Information complexity criterion - Bozdogan 93).
  - ICL (Integrated Completed Likelihood - Biernacki, Celeux, Govaert 2000) :
    BIC + estimated mean entropy.
  - MML (Minimum Message Length - Figueiredo, Jain 2002).

- IC + Piecewise Linear Regression (PLR) (see [7]).

Outline

Von Mises-Fisher distribution

$x \in \mathbb{R}^p$ with $\|x\|_2 = 1$. Set of parameters: θ

Definition

$$f(x\,|\theta) = C_p(\kappa)\exp\left(\kappa\mu^T x\right)$$

$\theta = \{\kappa, \mu\}$ with κ concentration parameter and μ mean direction.

- **Von Mises distribution** for $p = 2$: $C_p(\kappa) = \frac{1}{2\pi I_0(\kappa)}$ and
  $\mu^T x = \cos(x - \mu)$
  $I_0(x)$ modified Bessel function of order 0.
- $p = 3$: $C_p(\kappa) = \frac{\kappa}{4\pi \sinh \kappa}$ using polar coordinates, $C_p(\kappa) = \frac{\kappa}{\sinh \kappa}$ otherwise.
- $p > 3$: $C_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)}$

Applications found in electric field, geology, bioinformatics, text mining, ...

Other directional distributions

$x \in \mathbb{R}^p$ with $\|x\|_2 = 1$. Set of parameters: θ

Watson distribution

$$f(x|\theta) = M\left(\tfrac{1}{2}, \tfrac{p}{2}, \kappa\right)^{-1} \exp\left(\kappa \left(\mu^T x\right)^2\right)$$

$\theta = \{\kappa, \mu\}$ with κ concentration parameter and μ mean direction. $M(.)$ Kummer function.

Other directional distributions

$x \in \mathbb{R}^p$ with $\|x\|_2 = 1$. Set of parameters: θ

Watson distribution

$$f(x|\theta) = M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right)^{-1} \exp\left(\kappa\left(\mu^T x\right)^2\right)$$

$\theta = \{\kappa, \mu\}$ with κ concentration parameter and μ mean direction. $M(.)$ Kummer function.

Fisher-Bingham or Kent distribution ($p = 3$)

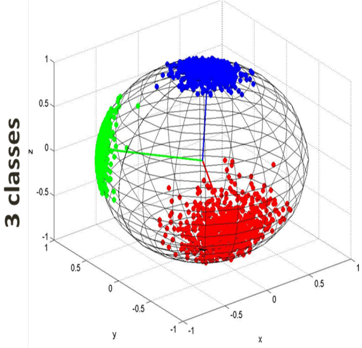
$$f(x|\theta) = c(\kappa, \beta) \exp\left(\kappa\mu_1^T x + \beta\left(\left(\mu_2^T x\right)^2 - \left(\mu_3^T x\right)^2\right)\right)$$

$\theta = \{\kappa, \beta, \mu_1, \mu_2, \mu_3\}$ with κ concentration parameter, β, $0 \le \beta < \kappa$, ellipticity of the contours of equal probability, μ_1 mean direction, μ_2 major axis and μ_3 minor axis.

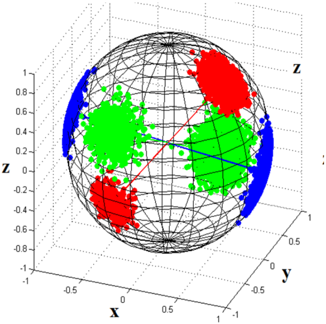
Examples with $p = 3$

Simulation of three different directional distributions:

von Mises Fisher distributions



Watson distributions



Application to depth images: surface normals are unit vectors in $\mathbb{R}^3$.

Directional distributions and EF

vMF distribution with $p = 3$, $x \in S^2 \subset \mathbb{R}^3$

- $t(x) = x$
- $\theta = \kappa\mu$, the natural parameters with $\kappa = \|\theta\|$ and $\mu = \frac{\theta}{\|\theta\|}$
- $F(\theta) = \log\left(\frac{\sinh(\kappa)}{\kappa}\right) = \log\left(\frac{\sinh(\|\theta\|)}{\|\theta\|}\right)$
- $g(x) = 0$
- $\eta = \left(\tanh(\kappa)^{-1} - \kappa^{-1}\right)\mu$

Directional distributions and EF

vMF distribution with $p = 3$, $x \in S^2 \subset \mathbb{R}^3$

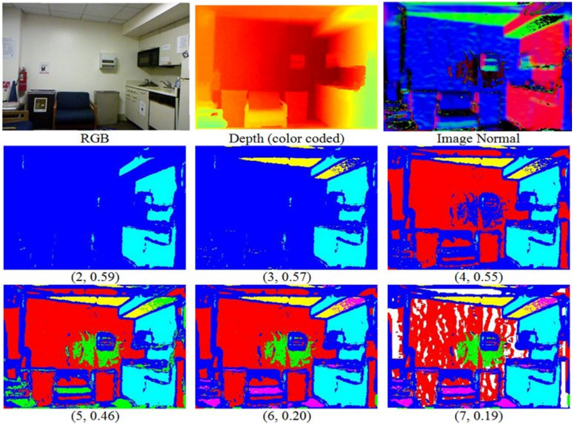
- $t(x) = x$
- $\theta = \kappa\mu$, the natural parameters with $\kappa = \|\theta\|$ and $\mu = \frac{\theta}{\|\theta\|}$
- $F(\theta) = \log\left(\frac{\sinh(\kappa)}{\kappa}\right) = \log\left(\frac{\sinh(\|\theta\|)}{\|\theta\|}\right)$
- $g(x) = 0$
- $\eta = \left(\tanh(\kappa)^{-1} - \kappa^{-1}\right)\mu$

Watson distribution with $x = [x_1, x_2, \ldots, x_p]^T \in S^{p-1} \subset \mathbb{R}^p$

- $t(x) = \left[x_1^2, \ldots, x_p^2, \sqrt{2}x_1x_2, \ldots, \sqrt{2}x_{p-1}x_p\right]^T$
- $\theta = \kappa\left[\mu_1^2, \ldots, \mu_p^2, \sqrt{2}\mu_1\mu_2, \ldots, \sqrt{2}\mu_{p-1}\mu_p\right]^T$ with $\kappa = \|\theta\|$
- $F(\theta) = \log\left(M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right)\right)$
- $g(x) = 0$
- $\eta = \frac{1}{p}\frac{M\left(\frac{3}{2}, \frac{p+2}{2}, \kappa\right)}{M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right)}\frac{\theta}{\kappa}$

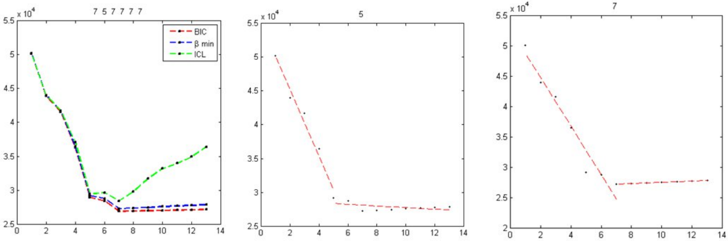
Results with an examplar of depth image (vMF)

Segmentation of a depth image generated by clustering image normals.



How to choose the number of components automatically ?

Curves of IC - Mixture of vMF with 7 classes

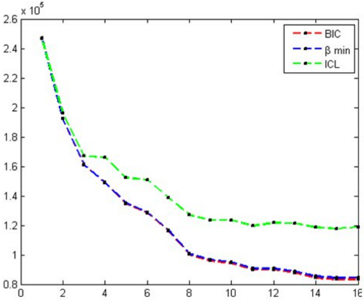


Curve: BIC, β_{min}, ICL   PLR, un-weighted   W-PLR [1,30]

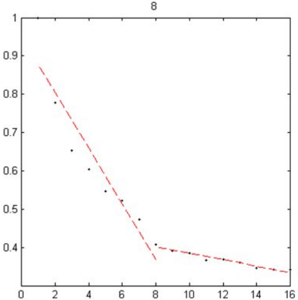
PLR and Weighted-PLR (proposed development)

$$\hat{k} = \underset{k}{\operatorname{argmin}} \left(\omega_l RMSE_{left} + \omega_r RMSE_{right} \right)$$

Curves of IC - depth image



Curve: BIC, β_{min}, ICL



W-PLR [1,30]

- Proposed W-PLR provides
  - **MoVMF case**: equivalent results as BIC, $\phi_{\beta_{\text{min}}}$ or ICL for simulated data
  - **Watson distribution case**: ICL is equivalent or better to W-PLR ($\omega_l = \omega_r = 1$)
  - Better component selection for real image data

Proposed methods

MBC-vMFMM and **MBC-WMM**

Validated by a study on synthetic data

(Model Based Clustering for von Mises-Fisher Mixture Model)

(Model Based Clustering for Watson Mixture Model)

		MBC-vMFMM	MBC-WMM
1st step	Initialization	kmeans++	DM
with $K = K_{\max}$	EM type procedure	BSC	BSC
2nd step	Distance	right-sided BD	right-sided BD
HAC	Linkage	Average	Average
	$\hat{K}$	IC & W-PLR	IC & W-PLR

DM: Diametrical clustering.

Dhillon et. Al, (2003), Diametrical Clustering for identifying anti-correlated gene clusters, Bioinformatics, vol 19, pp. 1612-1619.

Estimation of the number of components MBC-vMFMM

Comparison between MBC-vMFMM and a method containing state-of-the-art methods called MBC-MoVMF.

	Acc (%)		Comp. Time (sec)	
	MBC-MoVMF	**MBC-vMFMM**	MBC-MoVMF	**MBC-vMFMM**
3,ws	87.913	**99.992**	8.9187	**2.953**
5,ws	84.487	**99.995**	8.1757	**2.9494**
7,ws	76.991	**99.994**	7.8314	**2.8663**
3,nws	93.788	**99.039**	10.74	**2.9201**
5,nws	90.012	**97.156**	8.6715	**2.9004**
7,nws	80.709	**92.966**	7.9239	**2.8822**

Estimation of the number of components MBC-vMFMM

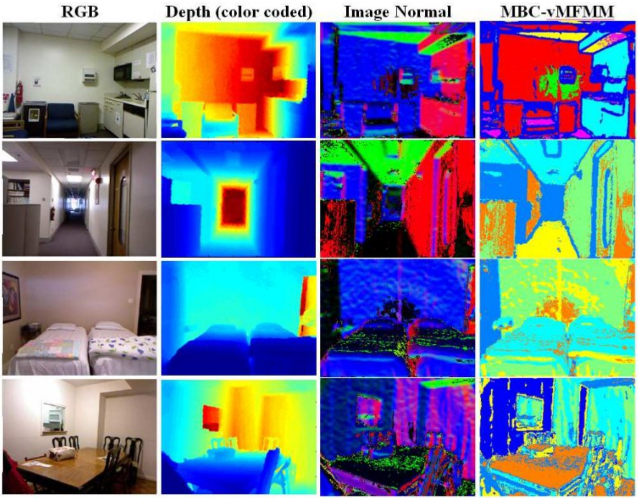
Comparison between MBC-vMFMM and a method containing state-of-the-art methods called MBC-MoVMF.

	Acc (%)		Comp. Time (sec)	
	MBC-MoVMF	**MBC-vMFMM**	MBC-MoVMF	**MBC-vMFMM**
3,ws	87.913	**99.992**	8.9187	**2.953**
5,ws	84.487	**99.995**	8.1757	**2.9494**
7,ws	76.991	**99.994**	7.8314	**2.8663**
3,nws	93.788	**99.039**	10.74	**2.9201**
5,nws	90.012	**97.156**	8.6715	**2.9004**
7,nws	80.709	**92.966**	7.9239	**2.8822**

100 % of correct number of components selection with MBC-vMFMM using :

- $\phi_{\beta_{\min}} IC$
- W-PLR on BIC curve with $\omega_l = 1$ and $\omega_r = 300$

With von Mises-Fisher MM (MBC-vMFMM using W-PLR)



With Watson MM (MBC-WMM using W-PLR)

Outline

A mixture model for RGB-D data

$x = \{(x_{s,D}, x_{s,C})\}_{s\in\Lambda}$, samples of **i.i.d.** (independently and identically distributed) random vectors with dimension $p = p_D + p_C$.

$$f\left(x_s\right) = \sum_{k\in K} P(\theta_k) f\left(x_{s,D}, x_{s,C} \,|\theta_k\right)$$

$x_{s,D}$ can contain spatial positions, depths, normals, ...
$x_{s,C}$ can be issued from different color spaces.

A mixture model for RGB-D data

$x = \{(x_{s,D}, x_{s,C})\}_{s\in\Lambda}$, samples of **i.i.d.** (independently and identically distributed) random vectors with dimension $p = p_D + p_C$.

$$f(x_s) = \sum_{k\in K} P(\theta_k) f(x_{s,D}, x_{s,C} \,|\theta_k)$$

$x_{s,D}$ can contain spatial positions, depths, normals, ...
$x_{s,C}$ can be issued from different color spaces.

Different hypothesis

- $x_{s,D}$ and $x_{s,C}$ are not independant
- $x_{s,D}$ and $x_{s,C}$ are independant
- What distributions can be used ?

A first proposal

- $x_{s,N}$ is the normal in s: vMF or W distributions
- $x_{s,P}$ is the position in s: Multidimensional Gaussian
- $x_{s,C}$ is the color in s using Lab: Multidimensional Gaussian
- Independance assumption $\theta_k = \{\theta_{N,k}, \theta_{P,k}, \theta_{C,k}\}$

$$f(x_s) = \sum_{k \in K} P(\theta_k) f_N(x_{s,N} | \theta_{N,k}) f_P(x_{s,P} | \theta_{P,k}) f_C(x_{s,C} | \theta_{C,k})$$

A first proposal

- $x_{s,N}$ is the normal in s: vMF or W distributions
- $x_{s,P}$ is the position in s: Multidimensional Gaussian
- $x_{s,C}$ is the color in s using Lab: Multidimensional Gaussian
- Independance assumption $\theta_k = \{\theta_{N,k}, \theta_{P,k}, \theta_{C,k}\}$

$$f\left(x_s\right) = \sum_{k \in K} P(\theta_k) f_N\left(x_{s,N} \left| \theta_{N,k}\right.\right) f_P\left(x_{s,P} \left| \theta_{P,k}\right.\right) f_C\left(x_{s,C} \left| \theta_{C,k}\right.\right)$$

Direct extension of previous method

- The combined BD is the sum of the three BD
- Centroids are computed independantly

First results on NYU Database

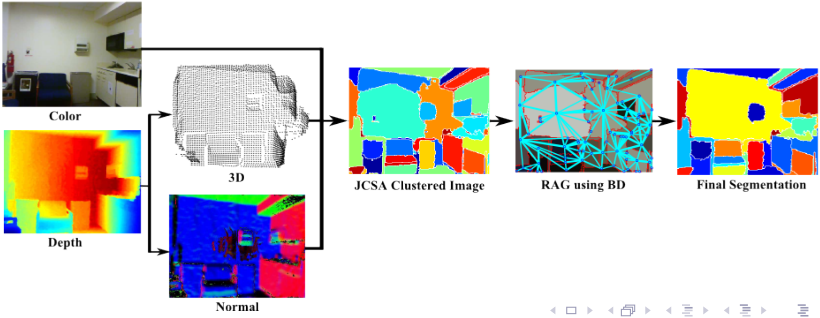
	PRI	VoI	GCE	BDE
N-WMM	**0,718**	3,523	0,511	**12,221**
N-VMFMM	**0,726**	3,770	0,540	**12,555**
C-MGMM	0,634	3,048	0,352	17,218
P-MGMM	0,702	**2,619**	**0,294**	25,471
N+C+P	0,694	**2,602**	**0,296**	23,516

Lower results than the most recent methods.

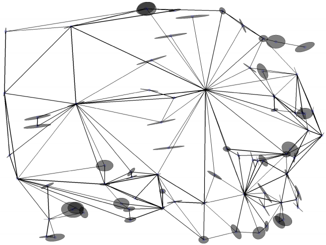
- **PRI**: Probabilistic Rand Index. Measuring the likelihood of a pair of pixels being grouped consistently in two segmentations.
- **VoI**: Variation of Information. Computing the amount of information of one result not contained in the other.
- **GCE**: Global Consistency Error. Measuring the extent to which one segmentation is a refinement of the other.
- **BDE**: Boundary Displacement Error. Computing the average displacement between the boundaries of two segmentations.

A RAG based method

- Another exploration of the hierarchy of models based on the following considerations:
  - The shadow zones are noisy.
  - Some planar regions deserve to be merged.
  - A need to take into account the boundaries of the RGB-D image in order to limit the merging processus.
- Block diagram of the proposed method :



A RAG based method - Example of obtained Graph



$R = \{r_i\}_{i=1,\dots,M}$, and undirected graph $G = (V, E)$

- v_i: node i corresponds to region i and is associated with mean direction, μ, and concentration κ of Watson distribution.
- e_{ij}: edge linking nodes i and j associated with two weights based on statistical dissimilarity, based on Watson distribution, and boundary strength ($I_G^{rgbd}(.)$ normalized gradient value):

$$w_d(v_i, v_j) = min(D_W(r_i, r_j), D_W(r_j, r_i))$$
$$w_b(v_i, v_j) = \frac{1}{|r_i \cap r_j|} \sum_{b \in r_i \cap r_j} I_G^{rgbd}(b)$$

A RAG based method - Proposed algorithm

$$cand(v_i) = \begin{cases} true, & \text{if } \kappa(r_i) > \kappa_{planar}, \\ false, & \text{otherwise.} \end{cases}$$

$$P_{ij} = \begin{cases} true, & \text{if } (a)\ cand(v_j) = true;\ and \\ & \quad (b)\ w_b(v_i, v_j) < th_{bound};\ and \\ & \quad (c)\ w_d(v_i, v_j) < th_{dist};\ and \\ & \quad (d)\ planar\ outlier\ ratio > th_{ratio} \\ false, & \text{otherwise.} \end{cases}$$

Input: $R = \{r_i\}_{i=1,...,M}$, $G = (V, E)$, $th_{boundary}$, th_{dist} and th_{ratio}
Output: Final segmentation after region merging.
Compute $cand(v_i)$ for $\{v_i\}_{i=1,...,M}$ using Eq. (18);
Set $i = 1$;
foreach i **do**
    if $cand(v_i)$ *is* true **then**
        while *no adjacent of* v_i *is left to check* **do**
            Sort e_{ij} in ascending order according to $w_b(v_i, vj)$;
            Evaluate each v_j with the *merging predicate* P_{ij} (Eq. (19)) ;
            if P_{ij} *is* true **then**
                Merge two nodes v_i and v_j and update the RAG;
                Start over again from sorting the adjacents.
            else
                Check the next node
            end
        end
    end
end

Results on NYU Database

	PRI	VOI	BDE	RC
OWT-UCM	0.89	2.60	**8.87**	0.56
GBS	0.77	2.37	16.04	0.45
JCSA-HRM	**0.90**	**2.32**	10.01	**0.57**

Methods

- **OWT-UCM**: Oriented Watershed Transform - Ultrametric Contour Map (PAMI 2011, CVPR 2012)
- **GBS**: Graph-Based Segmentation for Colored 3D Laser Point Clouds (2010).
- **JCSA-HRM**: Joint Color-Spatial-Axial clustering with Hierarchical Region Merging (proposed method).

Evaluation criteria

- For PRI, VoI and BDE, see previous slide.
- **RC**: Region (or segmentation) Covering evaluates the region overlaps between two segmentations (evaluation of the pixel-wise classification task in recognition).

Outline

Conclusion and perspectives

Conclusion

- An "unsupervised" clustering procedure based on Bregman divergence
  - for Normals (MoVMF and MoW)
  - for Normals+Positions+Colors (Independance assumption)
  - estimation of the number of component with IC and W-PLR based on BIC curve
  - accurate results on NYU Database using a RAG based algorithm

Conclusion and perspectives

Conclusion

- An "unsupervised" clustering procedure based on Bregman divergence
  - for Normals (MoVMF and MoW)
  - for Normals+Positions+Colors (Independance assumption)
  - estimation of the number of component with IC and W-PLR based on BIC curve
  - accurate results on NYU Database using a RAG based algorithm

Perspectives

- Extension to Kent mixtures
- MRF (or CRF, ...) based spatial regularization
- Still improving the management of the hierachy of models
- Exploring the different possible combinations of N, P, C, ...
- Semantic analysis

References

[1] F. Nielsen and V. Garcia.
Statistical exponential families: A digest with flash cards,
http://arxiv.org/abs/0911.4863v2, 2011.

[2] A. Banerjee, S. Merugu, I. S. Dhillon and J. Ghosh.
Clustering with Bregman Divergences,
Journal of Machine Learning Research, 6, pp. 17051749, 2005.

[3] F. Nielsen and R. Nock.
Sided and Symmetrized Bregman Centroids,
IEEE Trans. on Information Theory, 55(6), pp. 2882-2904, 2009.

[4] N. N. Cencov.
Statistical decision rules and optimal inference,
translated from Russian (Nauka, Moscow, 1972) in "Translations of Mathematical Monographs", 53, American Mathematical Society, 1982.

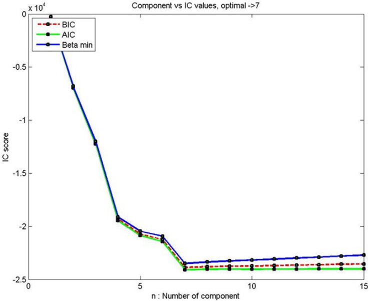
[5] V. Garcia and F. Nielsen
Simplification and hierarchical representations of mixtures of exponential families,
Signal Processing, 90, pp. 3197-3212, 2010.

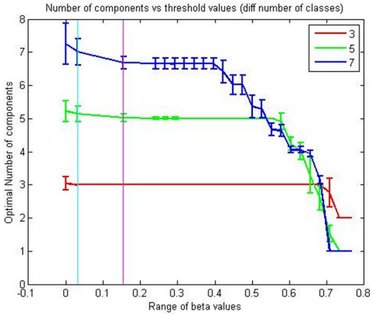
[6] Olivier Alata and Ludovic Quintard.
Is there a Best Color Space for Color Image Characterization or Representation based on Multivariate Gaussian Mixture Model ?,
Computer Vision and Image Understanding, 113(8), pp.867–877, 2009.

[7] S. Salvador and P. Chan.
Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms,
in 16th IEEE Int. Conf. on Tools with Artificial Intelligence, 2004, pp. 576-584.

Merci pour votre attention ! Des questions ?

Curves of ϕ_β IC





Data containing 7 classes. Data containing 3, 5 or 7 classes.

- $K_{\max} = 15$
- IC can be computed using Bregman divergence and a hard clustering of data using the estimated models with decreasing number of components.

Clustering accuracy - K is known - vMF-MM

$N = 10000$

Methods:

- kmeans++ (KMPP)
- Gaussian Mixture Model (GMM)
- Spherical kmeans (SPKM)
- vMF-MM [Banerjee & al. 2005]
- H-vMF-MM (MBC-vMFMM without number of components estimation)

	KMPP	GMM	SPKM	vMF-MM	**H-vMF-MM**
3 cl, ws	93.41	91.71	98.23	98.92	**99.99**
5 cl, ws	90.76	83.93	97.07	97.6	**99.99**
3 cl, nws	89.58	90.5	92.25	93.07	**99.05**
5 cl, nws	85.76	86.06	93.64	94.96	**97.16**

Clustering accuracy - K is known - Watson MM

$N = 10000$

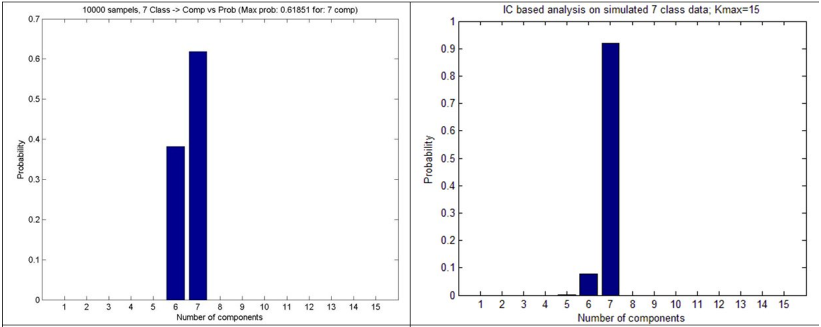
Methods:

- Diametrical clustering (DM) [Dhillon & al. 2003]
- Mixture of Watson Distributions (MWD) [Bijral & al. 2007]
- HWMM (MBC-WMM without number of components estimation)

	DM	MWD	**HWMM**
2 cl, ws	99.9998	99.9998	**100**
3 cl, ws	98.9932	98.0559	**99.9992**
4 cl, ws	90.9388	98.1295	**99.9909**
2 cl, nws	97.1702	**97.2206**	97.1998
3 cl, nws	**96.7246**	96.0239	92.207
4 cl, nws	97.9254	96.7252	**98.0496**

Importance of the initialisation

Estimation of the number of components with IC



Kmeans | Kmeans++

Synthetic data containing 7 classes

Estimation of the number of components (MBC-WMM)

	Cl. Acc	Time	BIC	$\phi_{\beta_{\min}}$	ICL
2, ws	100	12.84	66	66	66
3, ws	99.99	12.30	100	100	100
4, ws	99.99	12.58	90	90	90
5, ws	99.96	12.84	98	98	98
7, ws	99.62	13.37	98	98	98
2, nws	97.23	12.47	98	100	98
3, nws	96.43	12.64	100	100	100
4, nws	98.06	13.11	100	100	100
5, nws	97.22	13.30	100	100	100
7, nws	91.97	14.97	100	100	100