A Fixed-point Estimation Algorithm for Learning The Multivariate GGMM: Application to Human Action Recognition

31st IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2018)

Fatma Najar [1] Sami Bourouis [1] Nizar Bouguila [2] Safiya Belghith [1]

[1]ENIT University,Tunisia

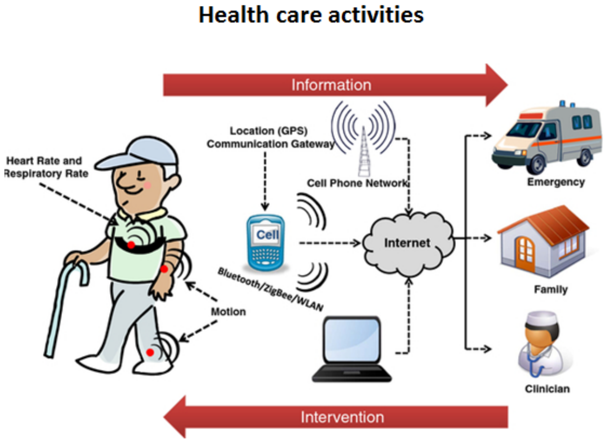
[2]Concordia University, Canada

Table of contents

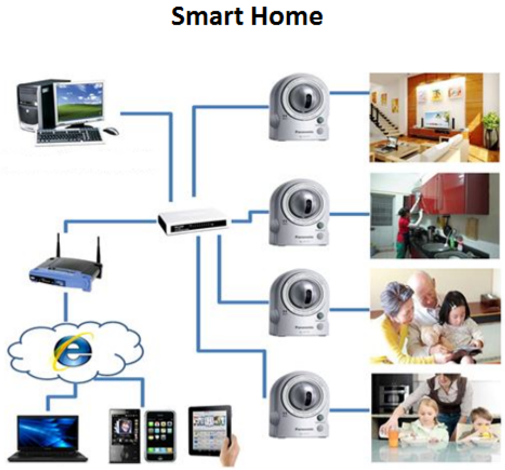
- **Human activity recognition**

Video surveillance systems

- **Human activity recognition**



- **Human activity recognition**



- **Human activity recognition**

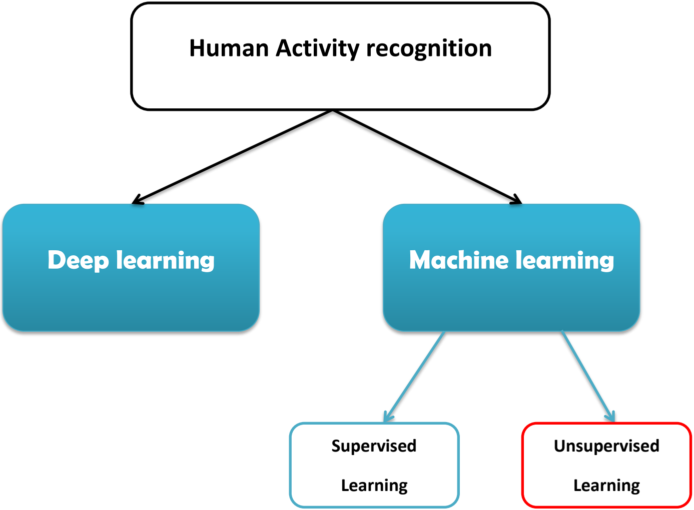
Complexity of action

Illumination changes

Background clutter

Change in scale

....


Human Activity recognition
Deep learning
Machine learning
Supervised Learning
Unsupervised Learning

Generalized Gaussian mixture model

$$p(\vec{X}_i|\vec{\mu}_j,\vec{\sigma}_j,\vec{\lambda}_j) = \prod_{k=1}^{d} B(\lambda_{jk}) \exp\left(-A(\lambda_{jk}) \left|\frac{X_{ik}-\mu_{jk}}{\sigma_{jk}}\right|^{\lambda_{jk}}\right)$$

- GGMM has been widely used for many applications
- Only diagonal covariance matrices have been used
- Assuming that features are independent

$$\Downarrow$$

Multivariate Generalized Gaussian distribution with Full Covariance matrix

$$p(X|\Sigma;\beta;\mu) = C(\beta)\frac{\beta}{m^{\frac{K}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2m^{\beta}}((X-\mu)^T\Sigma^{-1}(X-\mu))^{\beta}\right]$$

Fixed Point estimation algorithm

- Maximum Likelihood estimator computed by an FP algorithm
- For any shape parameter $\beta \in [0, 1]$, the MLE of MGGD' parameters are defined by :

$$\hat{\Sigma}_{k+1} = f(\Sigma_k) \quad (1)$$

where

$$f(\Sigma) = \sum_{i=1}^{T} \frac{K}{u_i + u_i^{1-\beta} \sum_{i \neq j} u_j^{\beta}} x_i x_i^T, \quad (2)$$

- A Newton-Raphson method for shape parameter :

$$\hat{\beta}_{k+1} = \hat{\beta}_k - \frac{\alpha(\hat{\beta}_k)}{\alpha'(\hat{\beta}_k)} \quad (3)$$

1. **Initialization step** : Initializing model's parameters with the k-means algorithm followed by the method of moment applied to each cluster.
2. Repeat until convergence of the log-likelihood :
   - **Expectation step** : Computing responsibilities

$$p(j|X_i) = \frac{p_j p(X_i|\Sigma_j;\beta_j;\mu_j)}{\sum_{m=1}^{M} p_m p(X_i|\Sigma_m;\beta_m;\mu_m)} \tag{4}$$

   - **Maximization step**
     - *Mean estimation*

$$\hat{\mu}_j = \frac{\sum_{i=1}^{T} p(j|X_i)|X_i - \mu_j|^{\beta_j - 1} X_i}{\sum_{i=1}^{T} p(j|X_i)|X_i - \mu_j|^{\beta_j - 1}} \tag{5}$$

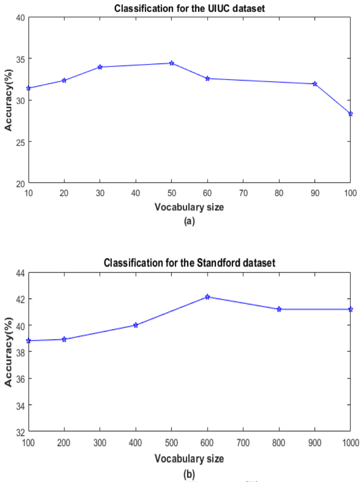
     - *Covariance estimation of each cluster* : Normalizing the dataset $(X_n = X - \mu_j)$, then evaluating the covariance matrix using equations 1 and 2.
     - *Shape estimation* : The shape parameter is determined using equation 3.
3. Assign each data point to the nearest cluster through the Bayes' rule.

1. Extract features using dense SIFT descriptors of 16×16 pixel patches computed over a grid with spacing of 8 pixels.
2. Quantize the image features into visual words using the bag of words (BOW) technique on the basis of the K-means algorithm.
3. Each image is represented as a frequency histogram over the V visual words.
4. Application of a probabilistic Latent Semantic Analysis (pLSA) to the obtained histograms in order to represent each image by a D-dimensional vector where D is the number of latent aspects.
5. Classifying the overall images to their right activities using our FP-MGGMM algorithm.

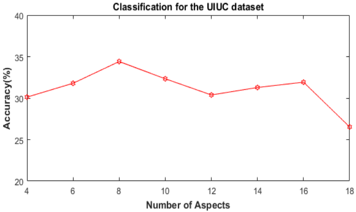
Figure – Sample images from the UIUC sports event dataset

Figure – Sample images from the Stanford 40 Action dataset

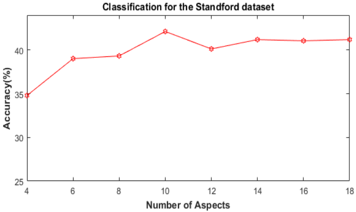
- Impact of different visual vocabulary sizes on the classification accuracy



- Impact of Number of aspects on the classification accuracy



(a)



(b)

- Comparative study between our proposed algorithm (FP-MGGMM) and GMM, GGMM (diagonal covariance matrix)

Algorithm	UIUC dataset	Stanford dataset
GMM	30.52	34.80
GGMM	31.69	35.20
FP-MGGMM	**34.41**	**42.13**

Table – The average classification accuracy rate for different mixture models

- FP-MGGMM offers the highest average accuracy rate (it is about 34% for UIUC and 42% for Stanford)
- It outperforms GGMM which assume that dimensions of the observed data are independent.

- The consideration of the full covariance matrix through the Fixed-point algorithm helps in improving the expected performances.

$$\Downarrow$$

More features used in the covariance matrix to describe the actions, better classification performances can be obtained.

Conclusion

- A novel unsupervised Fixed-point estimation algorithm for learning the multivariate generalized Gaussian mixture model that uses the full covariance matrix.
- Applied the proposed algorithm to Human activity recognition
- Evaluated the performance of the proposed framework through two publicly available datasets : UIUC Sport Event dataset and Stanford 40 Action.
- Obtained results are encouraging and show that our model outperforms the GMM and GGMM which are based only on the diagonal covariance matrix.
- **Future work :** Improvement of obtained results by taking into account more relevant visual features and also by adopting a semi-supervised or a weak-supervised setting.

Thank you for your attention !

Fixed Point Iteration

$$x^2 - x - 1 = 0$$

$$x_{n+1} = 1 + \frac{1}{x_n}$$

Pick $x_1 = 2$

$$x_2 = 1 + \frac{1}{2} = 1.5$$

$$x_3 = 1 + \frac{1}{1.5} = 1.666$$

$$x_4 = 1 + \frac{1}{1.666} = 1.6$$