Unsupervised classification of weld defects in radiographic images based on mutivariate generalized Gaussian mixture

Nafaa Nacereddine¹, Aicha Baya Goumeidane¹, Djemel Ziou²

¹ Research Center in Industrial Technologies, Algiers, Algeria. Email: (n.nacereddine.a.goumeidane)@ctri.dz
² DMI, Sherbrooke University, Quebec, Canada. Email: djemel.ziou@usherbrooke.ca

Abstract An accurate modeling of unknown probability density functions (pdfs) of data, encountered in practical applications, can play an important role in machine learning, clustering and pattern recognition. Including Gaussian, Laplacian and uniform distributions as special cases, multivariate generalized Gaussian distribution (MGGD) is potentially interesting for modeling the statistical properties of computer vision applications. In fact, the GGD is an elliptically contoured distribution characterized not only by a mean vector and a scatter matrix, but also by a shape parameter determining the peakedness of the distribution and the heaviness of its tails making this distribution more flexible than multivariate Gaussian distribution (MGD) and thus, more suitable for modeling, among others, images or features extracted from these images. However, the expressions of the partial derivative equations (PDE) deriving the MGDD parameters handle highly nonlinear functions including piecewise, logarithm, gamma, psi, power, etc. So, a particular attention is required for the derivatives computation, especially, for the matrix differentiation. Here, the solutions are given by the Newton-Raphson method. In order to carry out the experiments, hundreds of weld defect regions, extracted from weld radiographic films provided by the International Institute of Welding, are used. These defects represent four weld defect classes (crack, lack of penetration, porosity, solid inclusion) and are indexed by eight measures of Shape Geometric Descriptor (SGD). The experimental results, in terms of confusion matrices and total classification rate, demonstrate an outstanding performance of the MGGD-based mixture model (MGGM) in the weld defect data modeling compared to the multivariate Gaussian mixture model (MGM).

Problematic

Methods

The d-dimensional Generalized Gaussian distribution (GGD) is defined as

\[ f(x; \mu, \Sigma, \beta) = \frac{1}{(2\pi)^{d/2} \beta \Gamma(d/2)} \exp \left( -\frac{1}{\beta} \sum_{i=1}^{d} (x_{i} - \mu_{i})^{2} \right) \]

where \( x \in \mathbb{R}^{d} \) is a random vector and \( \beta \) is the dimensionality of the probability space, \( \mu \in \mathbb{R}^{d} \) is the mean vector, \( \Sigma \) is a symmetric positive definite scatter matrix of size \( d \times d \), and \( \Gamma \) is a shape parameter. The \( \Gamma \) function is defined by \( \Gamma \left( \frac{d}{2} \right) = \frac{1}{\Gamma \left( \frac{d}{2} \right) \beta} \). The \( \Gamma \) norm of vector \( x \) is noted \( ||x||_{\beta} \).

For a multivariate Generalized Gaussian Mixture Model, the complete log-likelihood is given by

\[ L(\theta) = -\frac{1}{2} \log(\Delta(\Sigma)) - \frac{1}{2} \sum_{i=1}^{M} n_{i} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{M} n_{i} \log(\beta) - \frac{1}{2} \sum_{i=1}^{M} n_{i} \log(\Gamma(d/2)) - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \left( x_{ij} - \mu_{ij} \right)^{2} / \beta \]

where \( N \) is the number of data set and \( M \) is the number of modes or clusters in the mixture.

Using EM algorithm, in the M-step, the posterior probabilities at iteration \( t \) are given by

\[ P(i|x_{ij}, t) = \frac{f(x_{ij} ; \mu_{ij}, \Sigma_{ij}, \beta_{ij})}{\sum_{k=1}^{M} f(x_{ij} ; \mu_{kj}, \Sigma_{kj}, \beta_{kj})} \]

In the M-step, the estimates of the mixture parameters, at iteration \( t+1 \), are expressed by

\[ \mu_{ij}(t+1) = \frac{1}{n_{i}} \sum_{j=1}^{N} f(x_{ij} ; \mu_{ij}, \Sigma_{ij}, \beta_{ij}) \]

\[ \Sigma_{ij}(t+1) = \frac{1}{n_{i}} \sum_{j=1}^{N} f(x_{ij} ; \mu_{ij}, \Sigma_{ij}, \beta_{ij}) \]

\[ \beta_{ij}(t+1) = \frac{1}{n_{i}} \sum_{j=1}^{N} f(x_{ij} ; \mu_{ij}, \Sigma_{ij}, \beta_{ij}) \]

Experimental Results

As example, we represent in this figure the probability density function pdf graphics of a bivariate generalized Gaussian mixture, defined on the real values in \([0,1] \times [0,1] \) with the following parameters:

\[ M = 3 \quad \beta_{1} = 0.3 \quad \beta_{2} = 0.4 \quad \beta_{3} = 0.6 \]

\[ \mu_{1} = [0.6, 0.3] \quad \mu_{2} = [0.7, 0.7] \quad \mu_{3} = [0.3, 0.6] \]

\[ \Sigma_{1} = \Sigma_{2} = \Sigma_{3} = \begin{bmatrix} 0.015 & 0.005 \\ 0.005 & 0.015 \end{bmatrix} \]

\[ \beta_{1} = 2 \quad \beta_{2} = 3.5 \quad \beta_{3} = 1.5 \]

In the experiments, 344 weld defect regions, extracted from weld radiographic films provided by the International Institute of Welding, are used. These regions represent four weld defect classes: crack (Cr), lack of penetration (LP), porosity (Po) and solid inclusion (Sl).

The database is indexed by the following geometric shape measures: Compactness (C), Elongation index (EL), Rectangularity (R), Solidity (S), Symmetry (Sym), Deviation index to the largest inscribed circle (DI), Euclidean Lengthening (E) and Roundness (R). Then, for the used mixture models, we have: \( N = 344, M = 4, d = 2 \).

Concluding remarks We remark, from the confusion matrices given in the results, that the overall recognition for the multivariate GGM mixture weld defect clustering reaches more than 95% while it is just equal to 93% for the clustering based on the Gaussian distribution. This outstanding performance is due, especially, to the higher classification rate of the lack of penetration (92%) for MGGM compared to MGM with a recognition rate of 81.4% and this, despite the difficult discrimination between these two types of defects because of their high shape similarity. In this paper, the efficiency of proposed finite mixture model using the multivariate generalized Gaussian distribution, thanks to the flexibility of the latter, is shown on the classification of a weld defect radiographic image database composed of crack, lack of penetration, porosity and solid inclusion indications and indexed by a shape geometric descriptor.