

---

# Un estimateur de la $L^2$ -mesure de dépendance probabiliste pour la réduction de dimension vectorielle

## Cas multiclasse

Wissal Drira, Faouzi Ghorbel

Laboratoire CRISTAL pôle GRIFT  
École Nationale des Sciences de l'Informatique  
Campus Universitaire de la Manouba, 2010 La Manouba, Tunisie  
Wissal.drira@gmail.com, Faouzi.ghorbel@ensi.rnu.tn

---

*RÉSUMÉ.* Nous évaluons ici les performances d'un estimateur de la  $L^2$ -mesure de dépendance probabiliste en vue de la réduction de dimension bidimensionnelle linéaire dans le cas multiclasse. Cette quantité qui présente un lien direct avec la probabilité d'erreur de classification, est construite à l'aide des séries de Fourier généralisées. Nous comparons l'algorithme proposé d'une part, à l'analyse discriminante linéaire introduite par Fisher et, d'autre part, à une version généralisée au cas multiclasse se basant sur l'extracteur linéaire récursif de la  $L^2$ -mesure de dépendance probabiliste. Dans le cas non gaussien, cette évaluation sera faite au sens de l'erreur des  $k$  plus proches voisins. L'estimateur à noyau des densités de probabilité est calculé dans le contexte du paramètre de lissage optimisé au sens de la moyenne quadratique intégrée. Ce dernier servira à l'estimation de la probabilité d'erreur de classification des mélanges de vecteurs gaussiens. Nous montrons sur un exemple de bases d'images de visages l'intérêt du réducteur de dimension proposé relativement aux méthodes conventionnelles.

*ABSTRACT.* We introduce an estimate of the  $L^2$ -probabilistic dependence measure constructed with the generalized Fourier series which is able to realize a linear vector feature dimensional reduction in the discriminate multi-class problem. It generalizes the Patrick-Fischer distance estimate generally used for dimensionality reduction of the feature space in the case of binary classification. It has a direct relationship with the probability classification error. We compare the proposed algorithm with the well known linear discriminate analysis (LDA) and with a generalized version of a multi class recursive linear extractor based on the  $L^2$ -probabilistic dependence measure (RID  $L^2$ -PMD). For vector Gaussian mixtures such comparison is done in the mean of the probability error of classification which is estimated by a multivariate Kernel probability density function. The corresponding smoothing parameters are optimized analytically in the sense of the Mean Integrated Square Error (MISE). The non Gaussian case is evaluated with the error of the  $k$  nearest neighborhood classifier. Finally we will illustrate the importance of the proposed method by testing it in the context of the face recognition.

*MOTS-CLÉS* :  $L^2$ -mesure de la dépendance probabiliste, réduction de dimension, paramètre de lissage.

*KEYWORDS*:  $L^2$ -probabilistic dependence measure, reduction dimension, smoothing parameter.

DOI:10.3166/TS.29.143-155 © 2012 Lavoisier

### Extended abstract

Researchers working with high dimensional data in the statistical pattern recognition confront the problem of the feature selection. Reduction dimension research in statistical field is witnessing a growing interest. It is well known that pattern recognition systems usually need many features to improve its performances. The definition of suitable representation requires generally a serious study in order to consider only pertinent and independent features. A large training and test samples are therefore constructed in order to design and evaluate the performances. In practice, the training sample serves to estimate conditional probability density functions of each class. The convergences of the different estimates have been well studied in the statistical literature according to the size  $N$  of the training sample (Silverman, 1986). It is well established that the sample size which is needed to estimate the probability density function as the histogram, the kernel or the orthogonal series, has to increase exponentially with the dimension  $D$  of the feature vector. In order to undergo over such limitation, the discriminate analysis for reducing the dimensionality is generally required. The convergence of such algorithms could be possible since the estimation of the criteria is realized in the reduced  $d$ -space ( $d \ll D$ ). However, in certain situations the popular Fisher discriminate analysis based on the scatter matrices (Fisher, 1936, Loog *et al.*, 2001) could give a not well reduced  $d$ -space. The following are two examples of such situations:

- one of the observation random vectors conditional to class distributions is multimodal,
- when the expectation values of each conditional observation to the label are much closed relatively to the traces of its covariance matrices,

Later, Patrick and Fischer have proposed a non parametric solution based on a probability density function distance. In the same work, they have introduced a kernel estimate for the Patrick-Fischer distance (Patrick, Fischer, 1969). A. Hillion *et al.* (1988) have considered the Gaussian kernel in the expression of Patrick-Fischer distance estimate for binary classification case. In the same work, a scalar feature extraction is applied to classify image by its texture. They showed experimentally the better performance of their method relatively to the Fisher one in the mean of the miss classification rate.

Following this probabilistic approach for discriminate analysis, we propose new estimators of the  $L^2$ -probabilistic dependence measure based on Fourier series probability density function estimate. These estimates are given for the multiclass case and the non scalar feature extraction problem which we call it multivariate

reduction case. For some situations, the learning sample size for the convergence of the Bayesian classifier in the  $d$ -reduced space becomes more realistic. With this demarche, it becomes possible to try to attend Bayesian classifier for the high dimensions since  $L^2$ -probabilistic measure has a direct link with the Bayesian probability error of classification. After reducing the dimensionality we estimate the probability error of classification with the adjusted kernel probability density function in the mean of the optimization of its smoothing parameters. The plug in method is applied in this goal.

In the beginning of the paper, we recall the linear discriminate analysis principles and the methods based on the probabilistic distances, specifically the Patrick-Fischer distance which is used for dimensionality reduction in the case of binary classification and we introduce the  $L^2$ -probabilistic dependence measure estimate which is the generalization to the multi-class case. Such estimate is constructed with the generalized Fourier series which is able to realize a linear vector feature dimensional reduction in the discriminate multi-class problem.

After that, we compare the proposed algorithm with the Linear Discriminate Analysis (LDA) and with a generalized version of a multiclass extractor by a recursive approach using the scalar linear extractor method with a kernel estimate of the  $L^2$ -probabilistic dependence measure (RID  $L^2$ -PMD). Some Gaussian and non Gaussian case of the conditional observation vectors relative to the labels are simulated in order to illustrate the performance of the proposed estimates. In the Gaussian mixtures case such comparison is done in the mean of the probability error of classification which is estimated by a multivariate Kernel probability density function. The corresponding smoothing parameters are optimized analytically in the sense of the Mean Integrated Square Error (MISE). In the non Gaussian case is evaluated with the error of the  $k$  nearest neighborhood classifier.

Finally, some experiment results on real data sets are realized for 2D facial classification based on certain number of morphological face features. We show the better performance of the suggested method relatively to the conventional method (LDA). We observe also improvement of the proposed multivariate reduction relatively to the RID  $L^2$ -PMD in the context of the face recognition.

## 1. Introduction

En grande dimension, la convergence des estimateurs nécessite des tailles d'échantillon très élevées surtout lorsqu'il s'agit de l'estimation de distributions non paramétriques. Il est bien établi que les théorèmes de convergence des estimateurs des fonctions de densités de probabilité qu'il s'agisse des méthodes à noyau, de l'histogramme ou celles basées sur les fonctions orthogonales, requièrent des tailles d'échantillons qui augmentent exponentiellement avec la dimension  $D$  des vecteurs aléatoires. Dans le cadre de la classification, la  $d$ -réduction de dimension est un moyen très utilisé pour pallier cette difficulté. Les critères de séparabilité ayant un lien direct avec la probabilité d'erreur de Bayes telle que l'information mutuelle (Miguel, Murillo, 2007), la distance de Chernoff (Loog *et al.*, 2001) ou encore la

distance de Patrick-Fischer (Patrick, Fischer, 1969) lesquelles s'expriment en fonction des densités de probabilité conditionnelles, sont alors estimées dans l'espace réduit dont la dimension notée  $d$  est souvent très faible relativement à  $D$  (*i.e.*  $d \ll D$ ). Grâce à l'engouement de nouvelles applications telles que la reconnaissance de visage et l'indexation de bases d'images par le contenu..., un regain d'intérêt pour ce type d'approches s'est manifesté (Nenadic, 2007 ; Aladjem, 1997 ; Diaconis, Freedman, 1984) puisque les méthodes faisant appel aux matrices de dispersion ne donnent des résultats satisfaisants que dans les situations où les densités conditionnelles à une classe sont toutes uni-modales ou de vecteurs moyens suffisamment distants relativement aux dispersions.

Nous avons proposé un nouvel estimateur pour la distance de Patrick-Fischer se basant sur les séries de Fourier usuellement utilisées en estimation des fonctions de densités de probabilité. Nous avons montré qu'une telle méthode peut réaliser une réduction de dimension vers un espace réduit de dimension supérieur à 2 (Drira, Ghorbel, 2010 ; 2012a). Dans cet article, nous suggérons d'étendre ces travaux au cas multiclasse par l'introduction d'un estimateur de la  $L^2$ -mesure de dépendance probabiliste à l'aide des fonctions orthogonales fournissant ainsi une 2D-réduction. Dans ce travail, la méthode proposée fera l'objet d'une comparaison avec l'analyse discriminante de Fisher et à une version généralisée au cas multiclasse de la méthode de Patrick-Fischer récursive (Drira, Ghorbel, 2012b ; Aladjem, 1996 ; 1997 ; Diaconis, Freedman, 1984) au sens du minimum de la probabilité d'erreur de Bayes et de l'erreur des  $k$  plus proches voisins selon que l'on considère un cas de mélange gaussien ou non. L'évaluation au sens de la probabilité d'erreur est réalisée par une estimation basée sur la méthode à noyau dans des conditions d'optimalité du paramètre de lissage et cela au sens de l'erreur quadratique moyenne intégrée (EQMI) (Saoudi *et al.*, 2009), alors que l'erreur des  $k$  plus proches voisins est estimée par les taux de mauvaise classification sur des échantillons tests. Ainsi la section 2 est consacrée à la formulation mathématique de la  $L^2$ -mesure de dépendance probabiliste basée sur la distance de Patrick-Fischer. La construction mathématique de l'estimateur que nous proposons d'introduire est traitée dans la section 3. L'évaluation par différentes simulations fait l'objet de la quatrième section. Enfin, l'algorithme proposé est appliqué à une base d'images de visages montrant ainsi son intérêt dans le cas de données réelles.

## 2. Formulation

La réduction de dimension linéaire a pour objectif de trouver une transformation linéaire  $W^*$  de l'espace des primitives de dimension  $D$  vers un espace vectoriel réduit de dimension  $d$  beaucoup plus faible. Elle est mise en œuvre par l'optimisation de critères s'exprimant en fonction des deux matrices de dispersion. Dans le cas de l'analyse de Fisher très souvent utilisée (Fisher, 1936), la matrice correspondant à  $W^*$  est formée par les  $D$  coordonnées des  $d$  vecteurs propres correspondant aux  $d$  plus grandes valeurs propres de la matrice,  $(\hat{S}_b)(\hat{S}_w)^{-1}$  où  $\hat{S}_b$  est une estimée de la matrice

interclasse et  $\hat{S}_w$  est l'estimée de la matrice intraclasse. Il est bien connu que ce type de méthodes ne fait intervenir que les moments statistiques d'ordre inférieur ou égal à 2 des vecteurs aléatoires mis en jeux. A titre d'exemple, dans le cas où une des densités de probabilité conditionnelles à une classe est multimodale, l'analyse discriminante linéaire (LDA) ne peut donner systématiquement satisfaction. Des méthodes basées sur des critères impliquant dans leurs définitions les densités de probabilité conditionnelles ont été étudiées dans la littérature telles que l'information mutuelle, la distance en probabilité, la mesure de dépendance probabiliste... Dans ce sens, nous nous intéressons à la mesure de dépendance probabiliste associée à la distance de l'espace fonctionnel des fonctions complexes de  $d$ -variables de carré intégrable, noté  $L^2(\mathbb{R}^d)$  dont l'expression pour  $d$  égale à 2 se réduit à :

$$I_p = \sum_{k=1}^K \left( \iint |\pi_k f_k(x, y) - f(x, y)|^2 dx dy \right)^{1/2} = \sum_{k=1}^K \|\pi_k f_k - f\|_{L^2} \quad (1)$$

où  $\pi_k$  et  $f_k$  représentent respectivement la probabilité *a priori* et la densité de probabilité conditionnelle à une classe  $k$  et  $f$  dénote la densité mélange.

### 3. Un extracteur linéaire bidimensionnel basé sur un nouvel estimateur de la $L^2$ -mesure de dépendance probabiliste

Dans cette section, nous généralisons le 2D-extracteur linéaire introduit dans (Drira et Ghorbel, 2010) au cas multiclasse par la proposition d'un estimateur de la  $L^2$ -mesure de dépendance probabiliste exprimée dans l'espace réduit de dimension 2. Nous supposons que tous les vecteurs aléatoires  $V = {}^t(X, Y)$  prennent leurs valeurs dans  $\mathbf{U}$  et que  $L^2(\mathbf{U})$  qui dénote l'espace des fonctions de carré sommable est muni d'une base de fonctions orthogonales  $e_{l,k}(v)$  avec  $v = {}^t(x, y)$  où  $k$  et  $l$  sont des entiers. Ainsi, l'estimateur par la méthode des fonctions orthogonales d'une densité de probabilité conjointe d'un couple aléatoire  $V$  s'écrit comme suit :

$$\hat{f}_{(V)}(v) = \sum_{l=1}^{m_N} \sum_{k=1}^{m_N} \hat{a}_{l,k} e_{l,k}(v) \quad (2)$$

$\hat{a}_{l,k}$  représente l'estimateur du coefficient de Fourier  $a_{l,k}$  qui s'exprime de la manière suivante :

$$\hat{a}_{l,k} = \frac{1}{N} \sum_{i=1}^N e_{l,k}(V_i) \quad (3)$$

avec  $\{V_i, i \in [1..N]\}$  est un échantillon de taille  $N$  distribué selon la loi du vecteur aléatoire  $V$  de dimension  $d$  et  $m_N$  représente le paramètre de troncature qui joue le rôle d'un facteur de lissage. La convergence en moyenne quadratique intégrée (MQI) de cet estimateur est obtenue lorsque la suite  $m_N$  est équivalente  $N^{1/s}$  avec  $s$  un nombre réel strictement supérieur à 2.

L'estimateur des fonctions orthogonales peut s'écrire aussi comme une sommation finie de noyau généralisant ainsi la méthode introduite par (Parzen, 1962) :

$$\hat{f}_{(v)}(v) = \frac{1}{N} \sum_{i=1}^N \tilde{K}_{m_N}(v, V_i) \quad (4)$$

avec :

$$\tilde{K}_{m_N}(v, V_i) = \tilde{K}_{m_N}(x, y, X_i, Y_i) = K_{m_N}(x, X_i) \cdot K_{m_N}(y, Y_i) \quad (5)$$

où  $K_{m_N}(x, X_j)$  est le noyau généralisé associé à une base de fonctions orthogonales  $e_l(x)$  donnée de l'espace des fonctions d'une variable réelle et de carré intégrable  $L^2(\mathbf{a}, \mathbf{b})$  et qui s'exprime comme suit :

$$K_{m_N}(x, X_j) = \sum_{l=1}^{m_N} e_l(x) e_l(X_j) \quad (6)$$

$(\mathbf{a}, \mathbf{b})$  dénote un intervalle borné ou infini de la droite réelle. Nous proposons ainsi de généraliser l'estimateur de la  $L^2$ -mesure de dépendance probabiliste associée à la distance de Patrick-Fischer à l'expression suivante :

$$\hat{I}_P = \frac{1}{(\sum_{k=1}^K N_k)^2} \sum_{k=1}^K \left( \begin{aligned} & \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} [\tilde{K}_{m_{Nk}}(V_i^k, V_j^k)] + \sum_{k=1}^K \sum_{j=1}^{N_k} \sum_{l=1}^K \sum_{j=1}^{N_l} [\tilde{K}_{m_N}(V_i^k, V_j^l)] \\ & - 2 \sum_{i=1}^{N_k} \sum_{l=1}^K \sum_{j=1}^{N_l} [\tilde{K}_{m_{Nk}}(V_i^k, V_j^l)] \end{aligned} \right)^{1/2} \quad (7)$$

où  $\{V_i^k, i = 1, \dots, N_k, k = 1, \dots, K\}$  désigne un échantillon d'apprentissage supervisé,  $N_k$  la taille de l'observation relative à la classe  $k$  et  $K$  dénote le nombre de classe.

#### 4. Évaluation des performances et simulations

Dans cette section, les performances du réducteur de dimension proposé sont évaluées selon un des deux critères suivants :

– la probabilité d'erreur de Bayes laquelle est estimée par la méthode à noyau dans des conditions d'optimalité du paramètre de lissage au sens de l'erreur quadratique moyenne intégrée (EQMI). Le paramètre de lissage est calculé dans le cas gaussien qui sera utilisé pour évaluer le cas de mélanges de vecteurs gaussiens (Saoudi *et al.*, 2009),

– l'erreur des  $k$  plus proches voisins que nous avons estimée par les taux de mauvaise classification sur des échantillons tests servira pour comparer les

différentes méthodes de réduction de dimension dans le cas où les vecteurs aléatoires conditionnels sont non gaussiens.

Nous avons généré des données de dimension  $D$  égale à 3 puis à 10 correspondant au cas gaussien-hétéroscédastique à trois classes. Malgré son imprécision, l'erreur apparente est quelquefois utilisée pour l'estimation de la probabilité d'erreur de Bayes puisqu'elle est souvent obtenue par la simple approximation des aires de recouvrement entre les densités de probabilité conditionnelles estimées pondérées par les proportions. Cela a pour intérêt de ne pas faire appel à un échantillon test. Dans le cas multivarié, l'extension de cette dernière méthode d'estimation requiert l'approximation des volumes de recouvrement entre les surfaces des fonctions densités de probabilité conditionnelles estimées.

Cette tâche n'est pas aisée et s'accompagne d'erreurs supplémentaires qui peuvent dépasser dans la plupart des cas la valeur de la probabilité d'erreur recherchée. Pour cela nous générons un échantillon test supervisé indépendant de l'échantillon d'apprentissage. Grâce à la comparaison des estimations des densités de probabilité conditionnelles pondérées par les proportions de chaque classe, nous calculons le taux d'erreur. Les densités de probabilité conditionnelles sont estimées à l'aide de la méthode des noyaux dans le cas d'optimalité des paramètres de lissages associés.

Dans le tableau 1, nous comparons les performances du réducteur à deux dimensions de la  $L^2$ -mesure de dépendance probabiliste (2D- $L^2$ PMD) à l'analyse discriminante de Fisher (LDA) et à celle à deux dimensions obtenue par la double itération de l'extracteur scalaire de la 1D  $L^2$ -mesure de dépendance probabiliste (R1D- $L^2$ PMD) (Aladjem, 1996 ; 1997). Dans ces résultats de simulations, la taille de l'échantillon test a été prise égale à celle de l'échantillon d'apprentissage  $N$  et dans des conditions d'équiprobabilité (*i.e.*  $N_1 = N_2 = \dots = N_k = N$ ).

Afin de comparer les différentes méthodes de réduction de dimension dans le cas non gaussien, nous nous sommes servis de l'erreur des  $k$  plus proches voisins, que nous avons appliquée sur des données de dimension  $D$  à trois classes dont deux sont distribuées selon des vecteurs gaussiens de matrices de covariances différentes. La troisième classe est construite de sorte à envelopper les deux autres et cela par la génération d'un vecteur gaussien tronqué par une boule centrée à l'origine. Le résultat correspondant est illustré par le tableau 2.

Il est important de signaler que les taux d'erreur affichés pour les deux méthodes optimisant la  $L^2$ -mesure de dépendance probabiliste dans le tableau 1, sont respectivement les résultats sur 10 réalisations différentes du  $\hat{I}_p$ . Seuls les taux d'erreur correspondant respectivement au maximum de ces dix valeurs de  $\hat{I}_p$  et le minimum des dix taux de mauvaise classification, sont retenus. Il est à constater qu'il y a cohérence entre les deux résultats dans la plupart des cas. Toutefois, nous remarquons que la valeur maximale obtenue sur  $\hat{I}_p$  au cours des 10 itérations ne donne pas nécessairement la plus faible estimation de la probabilité d'erreur de Bayes. Cette heuristique permet de tenter d'atteindre le classifieur idéal de Bayes. Il nous paraît raisonnable de noter que l'atteinte d'un tel objectif dépend de la convergence vers un maximum absolu ou non de la fonctionnelle  $\hat{I}_p$  laquelle est

conditionnée par la performance de l'algorithme numérique d'optimisation mis en jeux. Enfin, il est utile de noter que sans l'ajustement du paramètre de lissage de la méthode d'estimation du noyau, aucune supposition de ce genre ne peut être tentée.

*Tableau 1. Taux de mauvaise classification par la méthode du noyau optimisée dans le cas gaussien*

N	D	LDA	R1D-L <sup>2</sup> PMD	2D - L <sup>2</sup> PMD (Max I <sub>p</sub> )	2D - L <sup>2</sup> PMD (Min Pr)
1000	3	0.1300	0.0033	0.0000	0.0000
2000	3	0.0867	0.0167	0.0600	0.0000
3000	3	0.2600	0.3167	0.0000	0.0000
50	10	0.1933	0.2000	0.1400	0.1033
100	10	0.2023	0.2300	0.2110	0.1233
100	10	0.2366	0.3000	0.2100	0.2100
150	10	0.0756	0.2844	0.0841	0.0800
150	10	0.0466	0.1688	0.0644	0.0644
150	10	0.2505	0.0613	0.0621	0.0500
150	10	0.1977	0.1488	0.1601	0.1488
150	10	0.1355	0.0711	0.0066	0.0066

*Tableau 2. Taux d'erreur des k plus proches voisins dans le cas non gaussien*

N	D	LDA	R1D-L <sup>2</sup> PMD	2D-L <sup>2</sup> PMD
100	3	0.0867	0.2733 <sup>(*)</sup>	0.0033
150	3	0.0867	0.3300	0.0033
250	3	0.0900	0.0633	0.0000
300	3	0.2267	0.1367	0.0100
500	3	0.1133	0.0500	0.0067
1500	3	0.0700	0.3300	0.0000
50	10	0.1266	0.1800	0.1200
100	10	0.1733	0.1700	0.1800
150	10	0.1300	0.1666	0.0833

La meilleure performance de la L<sup>2</sup>-mesure de dépendance probabiliste est observée par des taux de mauvaise classification inférieurs ou égaux (tableau 1). Nous remarquons que les performances sont comparables lorsque les lois conditionnelles sont gaussiennes et de vecteurs moyens relativement distants par rapport aux valeurs des traces des matrices de covariance de chaque classe. Dans les cas où une classe est entourée ou enveloppée par la seconde, le plan discriminant obtenu par la L<sup>2</sup> mesure de dépendance probabiliste est le plus optimal au sens du taux d'erreur des k plus proches voisins (tableau 2).

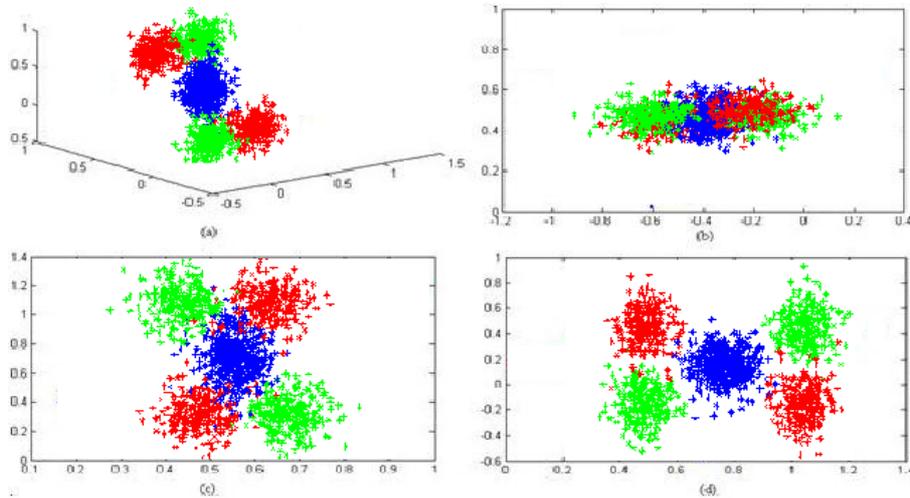


Figure 1. (a) Nuage des observations avant réduction ( $D = 3$ ), (b) projection sur le plan le plus discriminant selon la LDA, (c) projection sur le plan le plus discriminant selon l'algorithme Récursif 1D- $L^2$ PMD, (d) projection sur le plan le plus discriminant selon la 2D- $L^2$ PMD proposée

La figure 1(a) représente les trois nuages des observations de chaque classe dans l'espace initial ( $U = \mathbb{R}^3$ ) générés selon un mélange de trois classes équiprobables avec un vecteur gaussien tronqué suivant une boule centrée à l'origine pour la première et la seconde classe et un troisième vecteur gaussien à l'intersection des autres classes. Les figures 1 (b), (c) et (d) représentent respectivement les nuages projetés sur les plans discriminants obtenus par la méthode de la LDA, la version généralisée au multiclass obtenue par la double itération de l'extracteur scalaire de la 1D  $L^2$ -mesure de dépendance probabiliste (R 1D  $L^2$ -PMD) et la  $L^2$  mesure de dépendance probabiliste proposée (2D- $L^2$ PMD). Visuellement, il ressort de ces figures que l'estimateur de la  $L^2$  mesure de dépendance probabiliste proposé fournit un meilleur plan discriminant.

## 5. Application à la reconnaissance de visage

La reconnaissance de visage est actuellement l'un des sujets d'applications les plus actifs en biométrie. Le nombre de descripteurs retenus est souvent élevé. Pour réussir une telle tâche, une réduction de dimension est nécessaire pour la convergence des estimateurs des lois conditionnelles, des probabilités a priori et de la probabilité d'erreur dans l'espace réduit dont la dimension ne dépasse pas la valeur de 3 afin que la taille des échantillons devienne raisonnable. Ainsi, la mise en œuvre de la  $L^2$ -mesure de dépendance probabiliste devient applicable.



Figure 2. Exemples de visage de BioID

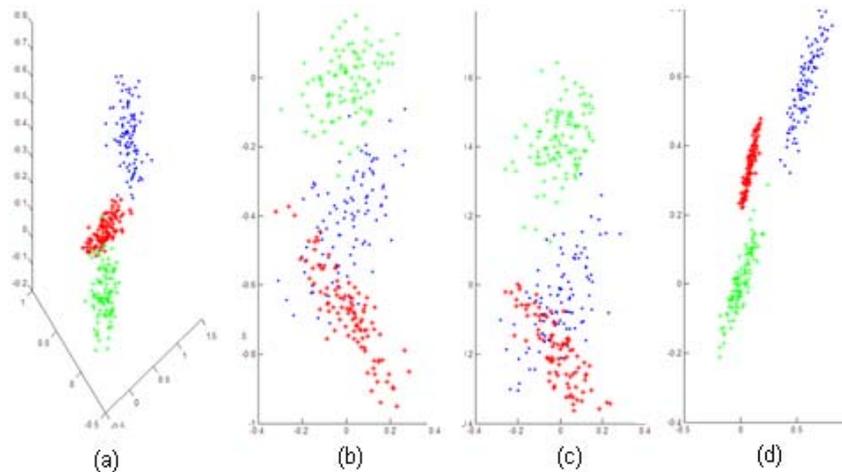


Figure 3. (a) Nuage avant réduction d'un sous-espace de 3 descripteurs pris au hasard parmi les 12 descripteurs ( $D = 12$ ), (b) projection sur le plan le plus discriminant selon la LDA, (c) projection sur le plan le plus discriminant selon l'algorithme Récursif  $1D-L^2PMD$ , (d) projection sur le plan le plus discriminant selon la  $2D-L^2PMD$  proposée

Pour le cas réel, nous avons utilisé l'ensemble de données « BioID » adopté dans (Jesorsky *et al.*, 2001). Cette base est constituée de 23 sujets : 1 521 images de vue frontale des images planes de visages en niveaux de gris. Pour chaque image, vingt points caractéristiques y sont indiqués manuellement. Ces images présentent

une résolution commune de 384 par 286 pixels (voir figure 2). Des primitives invariantes relativement aux déplacements du plan sont construites par le calcul de caractéristiques géométriques. Pour cet exemple, nous avons considéré des distances entre des points d'intérêt représentant respectivement la forme du menton, la largeur et la longueur du nez... Pour des raisons de réduction de la complexité algorithmique, la base des fonctions orthogonales trigonométriques a été choisie puisque les noyaux se réduisent au noyau de Dirichlet. Comme ces descripteurs sont exprimés en unité de pixel, une normalisation leur est préalablement effectuée de manière à contrôler l'effet de Gibbs provenant des oscillations des fonctions trigonométriques lesquelles induisent des divergences au niveau des bords du support des données dans l'estimation de la  $L^2$  mesure de dépendance probabiliste.

Les trois méthodes sont appliquées pour trouver les plans discriminants. La figure 3 illustre le résultat obtenu sur trois classes d'images des 3 personnes en prenant 12 primitives de distances. La classification est ensuite opérée sur l'espace réduit, confirmant ainsi les résultats des simulations de la section 4. Cet exemple tend à prouver l'intérêt de la  $L^2$  mesure de séparabilité probabiliste par rapport aux autres méthodes évoquées dans cet article.

## 6. Conclusion

Dans cet article, nous avons introduit un estimateur de la  $L^2$ -mesure de dépendance probabiliste par les fonctions orthogonales qui permet de réaliser une réduction de dimension bidimensionnelle dans le cas multiclasse. L'évaluation des performances au sens de la probabilité d'erreur est devenue possible grâce à l'introduction d'un estimateur de la probabilité d'erreur basé sur l'estimateur à noyau calculé dans les conditions d'optimalités du paramètre de lissage. Cette optimalité est faite au sens de la moyenne quadratique intégrée mais uniquement dans le cas de mélange de vecteurs gaussiens et au sens de l'erreur des  $k$  plus proches voisins dans le cas de mélange de vecteurs non gaussiens. Nous avons pu observer une meilleure performance de l'algorithme de la 2D-réduction linéaire multiclasse proposé par rapport à celui de Fisher et de la version bivariée obtenue par la double itération de l'extracteur scalaire de la 1D  $L^2$ -mesure de dépendance probabiliste (R 1D- $L^2$ PMD) généralisée au multiclasse.

En perspective de ce travail nous projetons d'étudier l'analyse discriminante tenant compte d'information sur le support des données à classifier et cela par l'application de ces deux estimateurs. En effet, pour chaque type d'espace de Hilbert, nous disposons de bases de fonctions orthogonales adaptées comme celles de Legendre, de Laguerre ou d'Hermite.

L'extension de l'algorithme proposé à des dimensions plus importantes afin de pouvoir l'exploiter dans des applications de la reconnaissance de formes telle que l'analyse de visages fera l'objet de nos travaux futurs.

L'évaluation des performances au sens de la probabilité d'erreur est devenue possible grâce à l'introduction d'un estimateur de la probabilité d'erreur basé sur

l'estimateur à noyau calculé dans les conditions d'optimalités du paramètre de lissage. Cette optimalité a été limitée uniquement au cas de mélange de vecteurs gaussiens. Nous proposons dans un travail ultérieur de généraliser cette approche au cas non gaussien.

## Bibliographie

- Aladjem M. E. (1996). Two class pattern discrimination via recursive optimization of Patrick-Fisher distance. *Proc. of the 13<sup>th</sup> International Conference on Pattern Recognition*, vol. 2, p. 60-64.
- Aladjem M. E. (1997). Linear discriminant analysis for two-classes via removal of classification structure. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, p. 187-192.
- Aladjem M. E. (1991). PNM: A program for parametric and nonparametric mapping of multidimensional data. *Computers in Biology and Medicine*, vol. 21, p. 321-343.
- Diaconis P., Freedman D. (1984). Asymptotics of graphical projection pursuit, *The Annals of Statistics*, vol. 12, p. 793-815.
- Drira W., Ghorbel F. (2010). Réduction de dimension par un nouvel estimateur de la distance de Patrick Fisher à l'aide des fonctions orthogonales. *42<sup>e</sup> Journées de Statistique*. Marseille, 24-28 Mai.
- Drira, W., Ghorbel F. (2012a). Dimension reduction by an orthogonal series estimate of the probabilistic dependence measure. *International Conference on Pattern Recognition Applications and Methods ICPRAM*, Portugal.
- Drira W., Ghorbel F. (2012b). *Vers un classifieur de Bayes pour les grandes dimensions obtenu par analyse discriminante basée sur un estimateur non paramétrique de la  $L^2$ -mesure de dépendance probabiliste*, édité par F. Ghorbel, S. Derrode et O. Alata édition Arts Pi Tunis.
- Fisher R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, vol. 7, p. 179-188.
- Hillion A., Masson P., Roux C. (1988). Une méthode de classification de textures par extraction linéaire non paramétrique de caractéristiques. *Traitement du signal*, vol. 5, n° 4.
- Jesorsky O., Kirchberg K., Frischholz R. (2001). Robust Face Detection Using the Hausdorff Distance. *Proc. Third International Conference on Audio and Video based Person Authentication*. Halmstad, Sweden, 6-8 June. p. 90-95. <https://support.bioid.com/downloads/facedb/index.php>
- Loog M., Duin R.P.W., Haeb-Umbach R. (2001). Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria. *IEEE transaction on PAMI*, vol. 23, n° 7.
- Miguel J., Murillo L., (2007). Maximization of Mutual Information for Supervised Linear Feature Extraction. *IEEE transactions on Neural Networks*, vol. 18, n° 5.
- Nenadic Z. (2007). Information Discriminant Analysis: Feature Extraction with an Information-Theoretic Objective. *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 29, n° 8.
- Parzen E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, vol. 33, n° 3, p. 1065-1076.
- Patrick E.A., Fisher F.P. (1969). Non parametric feature selection. *IEEE Trans. On Inf. Theory*, vol. IT-15, p. 577-584.

- Saoudi S., Troudi M., Ghorbel F., (2009). An Iterative Soft Bit Error Rate Estimation of Any Digital Communication Systems Using a Nonparametric Probability Density Function. *EURASIP Journal on Wireless Communications and Networking*, vol. 2009.
- Silverman BW. (1986). *Density Estimation for Statistics and Data Analysis*, London, Charman and Hall.

**Wissal Drira** est ingénieur en Informatique de l'Ecole Nationale des Sciences de l'Informatique (ENSI) et titulaire du diplôme de Master PRISM option image. Elle est actuellement doctorante au Groupe de Recherche en Images et Formes de Tunisie (GRIFT) du laboratoire CRISTAL et assistante contractuelle à l'ISAMM (Institut Supérieure des Art Multimédias de la Mannouba). Ses travaux actuels concernent la réduction de dimension et la classification par des approches stochastiques.

**Faouzi Ghorbel** est diplômé de l'École nationale supérieure des télécommunications de Bretagne, Docteur de l'université de Rennes I et Habilité à diriger des Recherche (HDR) de la même université (95). Ancien Professeur à Telecom Lille I (de 1991 à 1997), il est actuellement Professeur à l'École nationale des sciences de l'informatique de Tunis, depuis 1996, président de l'Association de Recherche Tunisienne des Sciences Pour l'Image (ARTS PI) depuis de 2004 et directeur du Laboratoire CRISTAL (Centre de Recherche en Réseau Image Système Architecture et Multimédia) depuis 2007.

