

11/90

وزارة التعليم العالي
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR

2 ex

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT : Genie Électrique

PROJET DE FIN D'ETUDES

SUJET

Logiciel De Classification

Interactive

Multi dimensionnelle

Proposé par :

M. Diaf

Etudié par :

H. Tabti

H. Mechli

Dirigé par :

M. Diaf

Le travail que nous présentons dans ce mémoire à été effectué au centre de calcul de l'Ecole Nationale Polytechnique .

Nous tenons à remercier l'équipe du Centre d'Automatique de l'Université des Sciences et Techniques de LILLE ,pour nous avoir donné l'idée de base de la méthode que nous avons programmée .

Nous exprimons notre profonde gratitude à l'égard de Monsieur M.DIAF , Docteur ingénieur en Automatique , qui est à l'origine de ce projet , pour nous avoir accordés son soutien constant et pour ne nous avoir jamais épargné de son temps et de ses connaissances .

Nous remercions vivement , A.BENSENOUSSI Phd et chef de département Génie Electrique pour avoir accepté d'être président du jury .

Nos remerciements vont aussi à Monsieur O.STIHI Master et chargé de cours, Monsieur L.ZEBROWSKI Professeur à L'E.N.P , Monsieur ILLOUL Master et chargé de cours, et Madame MATOUGUI chercheur au CTDA pour avoir accepté de faire partie du jury .

Que Monsieur M.Bey trouve ici l'expression de nos profonds sentiments pour son aide technique .

SOMMAIRE

INTRODUCTION GENERALE	1
CHAPITRE I . SYSTHESE DE QUELQUES METHODES DE CLASSIFICATION AUTOMATIQUE	
1. Introduction	2
2. Perception et classification automatique	2
3. Analyse en composante principale	3
4. Approche metrique de la classification	4
5. Approche statistique de la classification	5
6. Quelques methodes de classification automatiques	6
6-1. Methode barycentrique	6
6-2. Methode deus nuees dynamiques	6
6-3. Decision bayesienne	8
7. Conclusion	9
CHAPITRE II. REPRESENTATION Module/Angle	
1. Introduction	10
2. Principe de la methode	10
3. Formulation	13
4. Conclusion	14

CHAPITRE III. PARTIE LOGICIEL

1	Introduction	15
2.	Choix d'implantation	15
3.	Les fonctions du logiciel	16
	a. Visualisation de la projection	16
	b. Visualisation de l'état de référence	17
	c. Sélection	18
	d. Classement	19
4.	Fonctionnement du logiciel	21
	4-1 Option fichier	23
	4-1-1 Option lire fichier	
	4-1-2 Option caractéristiques fichiers	
	4-1-3 Option caractéristiques classes	
	4-1-4 Option sauvegard des fichiers classes	
	4-1-5 Option help	
	4-2 Option visualisation	25
	4-2-1 Option visualisation de la projection	
	4-2-2 Option sélection	
	4-3 Option configuration	27
	4-3-1 Option configuration de la projection	
	4-3-2 Option configuration des coordonnées	
	4-4 Saisie	29
	4-5 Classement	30
	4-6 Quit	31

5. Structure des fichiers utilises	32
6. Applications et resultats	34
6-1. Test de validite de la methode	34
6-2. Etude d'un groupe d'abeilles	37
7. Conclusion Generale	43

CHAPITRE I

SYNTHESE DE QUELQUES METHODES D'ANALYSE DE DONNEES

INTRODUCTION GENERALE

De tout les temps , l'homme essaie d'inculquer au machines sa façon de percevoir , d'évaluer et d'interpréter l'environnement dans le quel il évolue .

Aujourd'hui dans certain domaines ces machines dépassent l'homme même si les algorithmes qui sont implantés dans ces machines sont établis par l'homme lui-même . C'est le cas ou on a à analyser un nombre très élevé de données .

Les méthodes d'analyse des données sont tellement nombreuses que l'utilisateur reste embarrassé dans son choix . Certaines méthodes de classification sont automatiques et d'autres interactives . Quant il s'agit de méthodes interactives l'utilisateur contribue à la classification en y ajoutant de son expérience .C'est de ces méthodes interactives qu'il est question dans notre projet .

Ainsi dans le premier chapitre nous présenterons des généralités sur la classification automatique .Dans le second chapitre nous décrirons la méthode *Module/Angle* que nous avons programmée .Cette méthode sera appliquée à différents types de données dont les résultats seront commentés dans le dernier chapitre .

I- INTRODUCTION

En classification automatique on constitue des *classes* ou des *partitions* à partir de tableaux de données. Cette étape constitue la phase d'apprentissage pour la reconnaissance de formes qui se résume finalement au problème d'affecter un *objet* (forme) dans une classe prédéterminée . On voit à travers ces définitions deux concepts distincts : la *classification* qu'on réserve à l'édification d'une partition ,et le *classement* qui consiste à affecter un élément à la classe correspondante ./1/

II- PERCEPTION ET CLASSIFICATION AUTOMATIQUE

L'homme a la faculté de percevoir une quantité excessive d'informations lui permettant de partitionner l'image et de discerner quelques objets la constituant. Cette *perception* comprend l'acquisition , l'évaluation et l'interprétation . /9/

La reconnaissance de formes permet aux machines d'avoir une perception relativement semblable . Ainsi une forme peut être décrite par un certain nombre N de paramètres X_1, X_2, \dots, X_N mesurés par des capteurs divers .On représente cette forme par le vecteur $X = [X_1, \dots, X_N]$ appelé *vecteur de forme* . Lorsqu'il s'agit de reconnaître plusieurs formes, les observations deviennent multidimensionnelles . Pour un observateur humain , discerner des groupements de points dans cet espace est impossible .

Ainsi plusieurs techniques ont été proposées pour transformer cet espace multidimensionnel en un espace bidimensionnel. Parmi ces techniques, on peut citer l'analyse factorielle des correspondances et l'analyse en composantes principales.

III- ANALYSE EN COMPOSANTE PRINCIPALE (ACP)

Parmi les techniques d'analyse de données qui transforment l'espace multidimensionnel en un espace réduit, nous allons décrire les étapes de l'ACP qui permet de réduire la dimension de l'espace de formes en éliminant les variables descriptives qui n'apportent pas beaucoup d'informations.

On part d'un tableau de N observations décrites pour P variables chacune et on calcule la moyenne M_1, M_2, \dots, M_p de chaque variable par la formule :

$$M_k = \frac{1}{N} \sum_{j=1}^n X_k(j) \quad j = 1, 2, 3 \dots, p$$

où $X_k(j)$ est la $j^{\text{ième}}$ variable le $k^{\text{ième}}$ individu.

On calcule ensuite la matrice de covariance Q dont chaque élément est calculé par :

$$q_{k,l} = \frac{1}{N} \sum_{j=1}^n (X_k(j) - M_k) * (X_l(j) - M_l)$$

La matrice Q étant symétrique et définie positive, ses valeurs propres sont réelles positives et correspondent à l'information contenue dans les nouveaux axes. Leur somme constitue la trace de la matrice Q .

On calcule les valeurs et les vecteurs propres de la matrice de covariance .

On ordonne par valeurs propres décroissantes. Le premier axe principal sera le vecteur propre associé à la plus grande valeur propre .

Si on faisant la somme des premières valeurs propres on obtient plus de 90 % de l'information totale , on rejette les autres valeurs propres et on ne garde que les premiers axes principaux. Les paramètres correspondants aux valeurs propres rejetées ne sont pas très discriminant . /3,8,16,17/

IV- APPROCHE METRIQUE DE LA CLASSIFICATION

On recherche à regrouper les observations telles que chaque individu de l'échantillon analysé ressemble plus aux individus de sa classe qu'à ceux des autres classes . Ce degré de ressemblance peut être une distance métrique par exemples . Parmi les méthodes utilisées on distingue des méthodes hiérarchiques qui permettent d'aboutir à des partitions par regroupements successifs . Dans ce genre de méthode , le nombre de classes est souvent supposé connu . On ne dispose d'aucune nature statistique pour la distribution des observations qui forment les classes et on ne tient compte d'aucune erreur .

Afin de tenir compte de la probabilité d'erreur qui doit être minimale , on utilise souvent l'approche statistique . /7,16,17/

V - APPROCHE STATISTIQUE DE LA CLASSIFICATION

Les méthodes utilisant cette approche nécessite une normalisation des données avant tout traitement. Si on connaît le nombre de classes en présence dans un échantillon et si l'on connaît la fonction de densité de probabilité, on peut déterminer la probabilité d'erreur associée au classement d'une observation quelconque dans chacune des classes . /7,8/

Parmi ces techniques, la plus utilisée est la décision bayésienne, on peut aussi citer les méthodes d'estimation par maximum de vraisemblance dans le cas multidimensionnel et la méthode de nuées dynamiques.

VI- QUELQUES METHODES DE CLASSIFICATION

1-Méthode barycentrique

Le principe de cette méthode est de donner en première étape, un nombre de classes fixe k . En deuxième étape on calcule le barycentre $G(Y)$ de chacune des classes tel que :

$$G(Y) = \frac{1}{N} \sum_{j=1}^N X_j$$

Où N est le nombre de points de chaque classe et $G(Y)$ le barycentre de la classe Y .

On calcul ensuite la distance de chaque point aux différents centres de classes, en utilisant la distance euclidienne définie par :

$$d(X,Y) = [X - G(Y)]^t [X - G(Y)]$$

Où $d(X,Y)$ est la distance du vecteur X à la classe Y et $G(Y)$, le vecteur moyen de la classe Y .

On attribue finalement chacun de ces points à la classe dont il est le plus proche ./17/

2 : Méthode des nuées dynamiques

L'algorithme des nuées dynamique est une procédure itérative. Initialement le nuage est divisé en k régions ou classes ($C_1, C_2 \dots, C_k$) d'une façon aléatoire et chaque classe comporte ainsi n_i ($i=1,2,3,\dots,k$) points.

Le résultat de la classification est en principe indépendant de cette solution initiale. A partir de celle-ci on calcule les centres de gravité L_i de chaque classe tels que :

$$L_i = \frac{1}{N} \cdot \sum_{j=1}^N X_i^{(j)}$$

puis on affecte tous les points aux classes, de telle façon qu'un point X sera affecté à la classe C_i , si sa distance à cette classe est inférieure à celles entre ce point et les autres classes C_j .

$$X \in C_i \text{ si } D(X, C_i) < D(X, C_j) \quad \forall j \neq i$$

Le résultat de cette étape est une partition $P = [P_1, P_2, P_3, \dots, P_k]$ du nuage tout entier

On vérifie ce résultat pour le critère d'optimisation W tel que :

$$W = \sum_{i=1}^k \sum_{x \in P_i} R(x, P_i)$$

avec
$$R(x, P_i) = \frac{1}{n_i} \cdot \text{Card}(P_i) \cdot D(x, P_i)$$

L'itération est arrêtée si $\frac{W^n - W^{n-1}}{W^n} < 10^{-3}$

Dans le cas contraire, on redéfinit les classes C_1, C_2, \dots, C_k formées de n_1, n_2, \dots, n_k points chacune, en prenant n_i points les plus proches de P_i ($i=1, 2, 3, \dots, k$). Ces n_i points sont ceux qui minimisent la fonction R . /2,7/

3- Décision bayésienne

Cette méthode statistique est couramment employée dans la classification automatique. A chaque classe C_i recherchée, définie par ses échantillons et donc par un nuage de points dans R , on associe une variable aléatoire dont on estime la loi de densité. Si on a k classes à rechercher, on aura k variables aléatoires possibles, suivant k loi de densité de probabilité.

Etant donné un point x , on l'associe à la classe C_i , donc à la variable aléatoire associée qui a une plus forte chance d'être une réalisation. Pour cela on emploie la règle de BAYES

$$P(C_i/X) = \frac{P(X/C_i) * P(C_i)}{P(X)}$$

C_i : Variable aléatoire associée à la classe i

On associe X à la classe C_i pour laquelle $P(C_i/x)$ est maximum. Comme on considère en générale $P(C_i/x) = \text{cte}$, on recherche alors la valeur de i pour laquelle $P(X/C_i)$ est maximum [1,3]. Cette règle peut s'écrire également.

$$X \in C_i \quad \text{Si} \quad P(X/C_i) > P(X/C_j) \quad \forall \quad j \neq i$$

VII- CONCLUSION

Dans les problèmes de classification automatique , plusieurs éléments peuvent influencer sur la qualité des résultats . Parmi ces éléments , la quantité de données qui influent sur le temps de calcul de l'ordinateur qui est lié à la puissance de la machine et l'espace mémoire nécessaire pour l'exécution de ces programmes .

CHAPITRE II

LA REPRESENTATION

MODULE / ANGLE

I- INTRODUCTION

La représentation *Module/Angle* est le résultat d'une approche qui imite un observateur humain dans un environnement complexe .En effet ,pour observer un nuage de points dans un espace multidimensionnel sous différentes directions l'observateur se déplace dans cet espace . Il est ainsi caractérisé par un point et une direction d'observation .

Les informations recueillies sont la distance euclidienne entre l'observateur et le point observé et l'angle entre le point observé et la direction du regard de l'observateur.

La possibilité de faire évoluer l'observateur dans l'espace multidimensionnel permet d'obtenir des classes qui évoluent en fonction de la position de cet observateur . Le résultat de la classification est noté par l'opérateur selon son expérience .

II - PRINCIPE DE LA METHODE

Après avoir placé l'observateur *Obs* dans l'espace multidimensionnel en fixant sa position et la direction de son regard , on calcule la distance qui le sépare d'un point *P* . Cette distance est donnée par $y = d(\text{Obs}, P)$. Le résultat est porté sur l'axe des abscisses . En ordonnées, on porte l'angle α entre le point *P* et la direction du regard de l'observateur . La détermination de l'angle est faite par rapport au référentiel. Celui-ci est défini par son origine . Sa projection est exprimée en coordonnées cartésiennes dans le repère à *N* dimensions .

Les coordonnées cartésiennes V_1 du vecteur regard dans l'espace à N dimensions sont données par les formules (I) si dessous . Elles sont établies en partant d'abord d'un plan , nous avons alors :

$$V_2 = R.\cos \alpha_2$$

$$V_1 = R.\sin \alpha_2$$

où V_1 et V_2 sont les projections du vecteur référence sur les axes du plan et R la norme du vecteur de regard .

En passant à l'espace à trois dimensions , ces projections deviennent :

$$V_3 = R.\cos \alpha_3$$

$$V_2 = R.\sin \alpha_3 \cos \alpha_2$$

$$V_1 = R.\cos \alpha_3 \sin \alpha_2$$

En généralisant à un espace à N dimension , nous aurons finalement :

$$V_n = R.\cos \alpha_n$$

$$V_{n-1} = R.\sin \alpha_n . \cos \alpha_{n-1}$$

$$V_{n-2} = R.\sin \alpha_n . \sin \alpha_{n-1} . \cos \alpha_{n-2}$$

..

..

... (I)

..

..

$$V_2 = R.\sin \alpha_n . \sin \alpha_{n-1} . \sin \alpha_{n-2} \dots \cos \alpha_2$$

$$V_1 = R.\sin \alpha_n . \sin \alpha_{n-1} . \sin \alpha_{n-2} \dots \sin \alpha_2$$

Dans ces formules, rappelons que R est la norme du vecteur de regard, α_i l'angle entre chaque projection du vecteur regard sur un plan formé par deux axes quelconques de l'espace multidimensionnelle et l'un de ces axes.

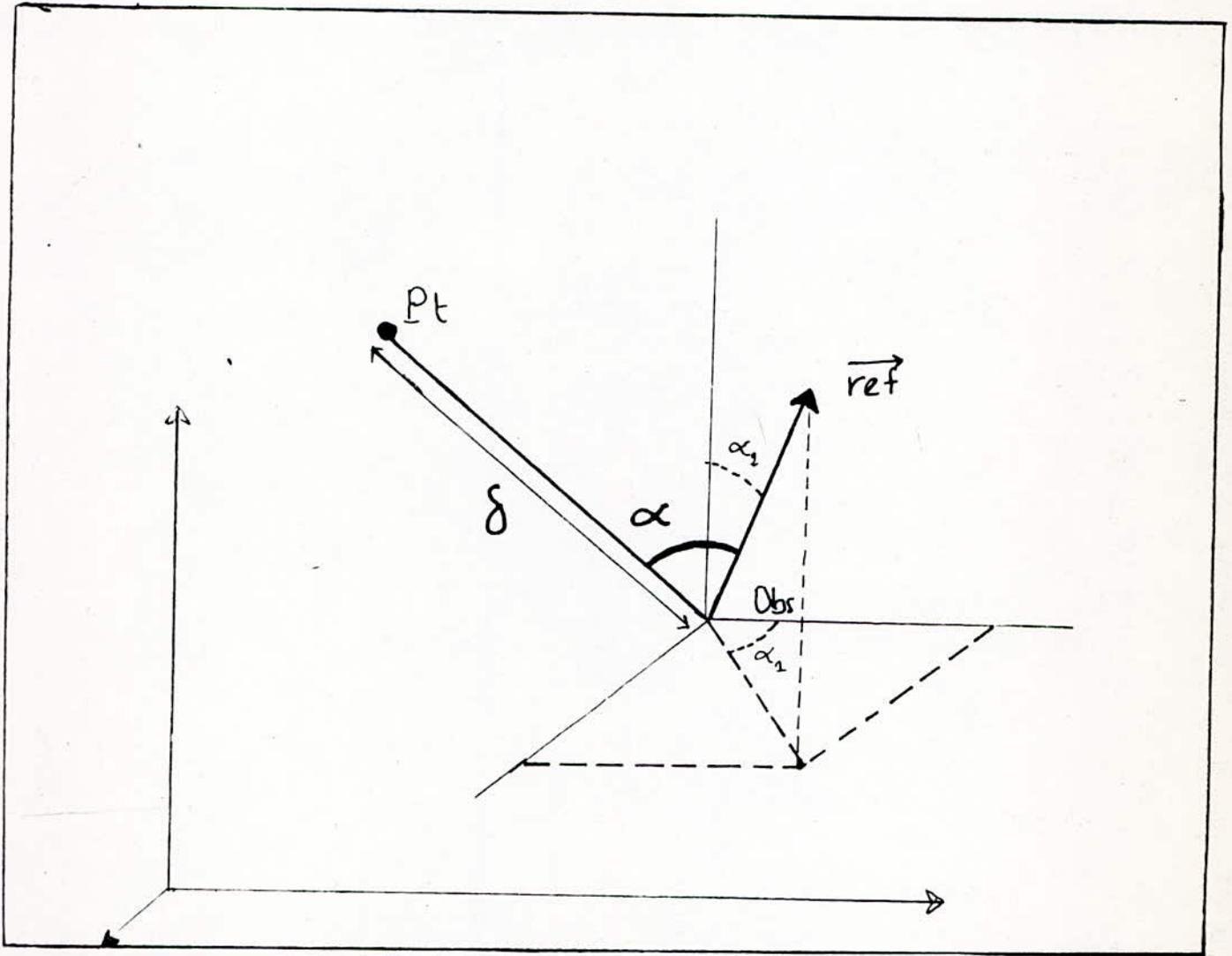


fig. 1 Observateur dans un espace à 3 dimensions

III FORMULATION

Sachant que chaque point est défini par m paramètres, la distance euclidienne entre le point considéré et l'observateur est donnée par la formule (2) suivante :

$$\rho_i^2 = \sum_{j=1}^m (\text{Obs}_j - \text{Pt}_{i,j})^2 \quad (2)$$

où Obs_j et $\text{Pt}_{i,j}$ sont respectivement les $j^{\text{ième}}$ coordonnées de l'observateur et le $j^{\text{ième}}$ point considéré dans l'espace multidimensionnel.

L'angle entre la direction du vecteur de regard de l'observateur et le point est donnée par la formule (3) suivante :

$$\alpha_i = \text{Acos} \left[(1/\rho_i) \cdot \sum_{j=1}^m (\text{Obs}_j - \text{Pt}_{i,j}) * V_j \right] \dots (3)$$

où V_j sont les coordonnées cartésiennes du référentiel données par les formules (1) précédentes.

IV - CONCLUSION

Dans cette méthode , le fait de faire balader l'observateur dans l'espace multidimensionnel, nous permet d'obtenir plusieurs projections qui nous font apparaitre des classes. En même temps que l'observateur se déplace , les classes évoluent . Là , l'intervention et l'expérience de l'opérateur permettent de fixer les résultats de la classification . En effet , vue sous plusieurs directions et de plusieurs points dans l'espace multidimensionnel, on arrive à élucider les points qui se cachent derrière d'autres points . Comme nous le verrons dans le prochain chapitre , cette méthode nous donne des résultats très satisfaisants lorsqu'elle est appliquée à des fichiers de données diverses .

CHAPITRE III

PARTIE LOGICIEL

1-INTRODUCTION

Comme toute autre méthode de classification automatique ,la méthode présentée au deuxième chapitre ne peut être exploitée que si elle est programmée .C'est pourquoi nous avons réalisé un logiciel basé sur cette méthode .Nous présentons dans ce chapitre les fonctions dont il est menu et son utilisation ainsi que les résultats on l'appliquant à des fichiers réels .

2- CHOIX D'IMPLANTATION

Etant donné l'aspect convivialité de la méthode que nous voulons réaliser et sachant la taille élevée des données multidimensionnelles à traiter , nous avons opté pour la station de travail hp 9000 comme support informatique . En effet ,celui-ci dispose d'une souris ,d'un système multifenêtrage " windows 9000 " et d'un écran haute définition (Annexe A) .

Pour la réalisation du logiciel ,nous avons utilisé le langage C et cela pour sa rapidité d'exécution , son aspect modulaire et son universalité .Il présente ainsi l'avantage de compiler séparément les procédures , ce qui facilite la maintenance des programmes . pour notre application particulière il a l'énorme autre avantage de gestion de bases de données à l'aide des pointeurs . /10,11/

3- LES FONCTIONS DU LOGICIEL

a - VISUALISATION DE LA PROJECTION

La procédure d'affichage de la projection du nuage de points est réalisé de façon à tenir compte à chaque fois de la variation des plus grandes valeurs de la projection ,et cela afin d'assurer une visualisation complete de l'image et une large utilisation de l'écran .

Les nouvelles coordonnées sont calculées alors par les formules suivantes :

$$X_{\text{aff}} = (X_{\text{proj}} - X_{\text{projmin}}) * \text{SCRNMAX} / (X_{\text{projmax}} - X_{\text{projmin}})$$

$$Y_{\text{aff}} = (Y_{\text{proj}} - Y_{\text{projmin}}) * \text{SCRNMAXY} / (Y_{\text{projmax}} - Y_{\text{projmin}})$$

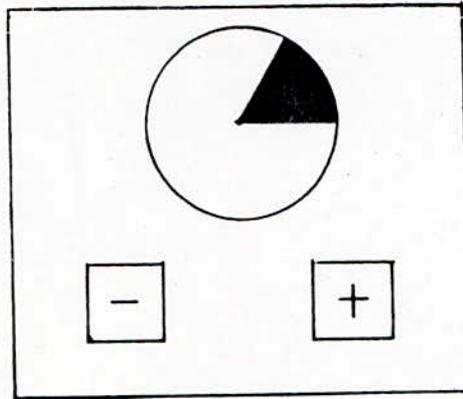
ou SCRMAX et SCRMAXY representent le maximum des coordonnées de l'écran .

$X_{\text{aff}}, Y_{\text{aff}}$: coordonnées des pixels écran .

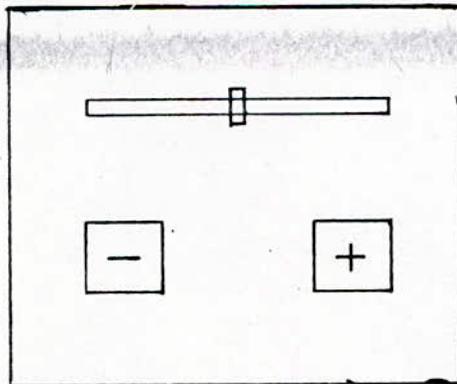
$X_{\text{proj}}, Y_{\text{proj}}$: coordonnées des transformées planes .

b - VISUALISATION DE L'ETAT DU REFERENCIEL :

La position de l'observateur dans l'espace multidimensionnel est représentée par des portions d'arcs de cercles ou camemberts *fig .2* et des curseurs se déplaçant sur des barettes représentant les axes dans leurs parties positives et négatives *fig.3*



*fig x Représentation de la Position Angulaire
Par rapport aux axes*



*fig x Représentation de la Position de l'Observateur
sur un Axe*

En plus de la représentation graphique il est nécessaire d'afficher les coordonnées de l'observateur . La valeur de l'angle de décalage du vecteur référence est affiché en degré , par contre la position est normalisé de sorte que l'observateur peut se déplacer entre $-M_i$ et $+M_i$ avec :

$$M_i = \text{MAX}(| X_{i \text{ max}} |, | X_{i \text{ min}} |)$$

$X_{i \text{ min}}$ et $X_{i \text{ max}}$ sont respectivement les valeurs minimales et maximales de i^{eme} paramètre .

c -SELECTION

On peut sélectionner une classe en l'entourant à l'aide de segments comme il est expliquée dans la partie concernant l'utilisation . On stocke dans un tableau toutes les observations dont la projection *Module/Angle* est détecté à l'intérieur du polygone et on affiche leur moyenne et leur écartype .

On peut étendre cette opération pour isoler et entourer l'ensemble de toutes les classes qui apparaissent après la projection . Cette opération (selection) est décrite par les étapes suivantes :

- Entourer l'ensemble des points par un polygone .
- Dessiner une demi-droite horizontale vers la droite .
- Compter le nombre d'intersections avec le polygone .
- si le nombre d'intersections est impaire alors le point est à l'intérieur . Le point est à l'extérieure du polygone dans l'autre cas ./13/

d - CLASSEMENT

Le but du classement est de décider et cela, en se basant sur les informations obtenues dans l'étape de la classification, à quelle classe on peut affecter un nouveau élément X . Le critère de décision est basé sur le calcul des distances euclidiennes entre le point P_t et les différentes classes. L'élément X est alors affecté à la classe C_i , si sa distance par rapport au centre de gravité de cette classe est inférieure à celles des autres classes. On a ainsi :

$$X \in C_i \quad \text{Si} \quad d(X, C_i) < d(X, C_j) \quad \forall j \neq i$$

Le logiciel que nous avons réalisé offre la possibilité de classer soit un ou plusieurs nouveaux éléments rangés dans un fichier.

La détermination du centre de gravité et le calcul de l'écart type d'une classe sont faits à chaque fois qu'un nouveau élément s'ajoute à cette classe.

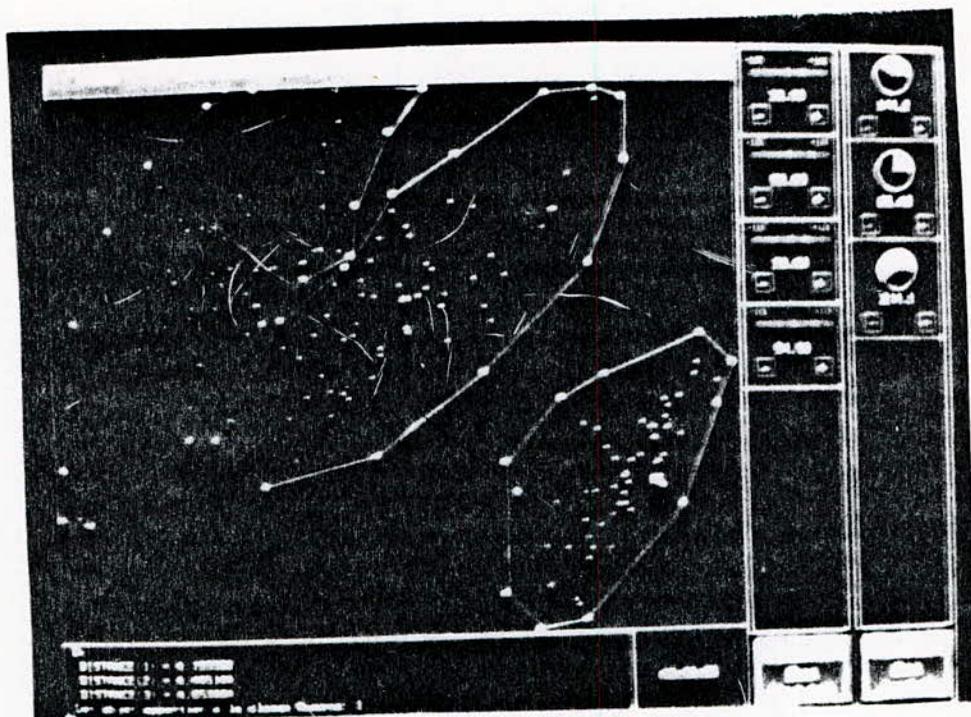


fig .4 Classement d'un objet

4- FONCTIONNEMENT DU LOGICIEL

1- Lancer le system multifenêtrage de la machine ,au moyen de la commande "Go" .Ce systeme multifenetrage à été adapté pour notre application ,sinon ce systeme s'exécute par la commande "wmstart".

2- Dès que la fenêtre systeme "wconsole" apparait, on exécute notre logiciel avec la commande "LOCIM" .Les cinq fenetres suivantes apparaissent .

- " Visualisation de la projection "
- " Visualisation de l'etat de la position de l'oservateur "
- " Visualisation de l'etat du référenciel "
- " Saisie d'information à partir du clavier "
- " Menu principal "

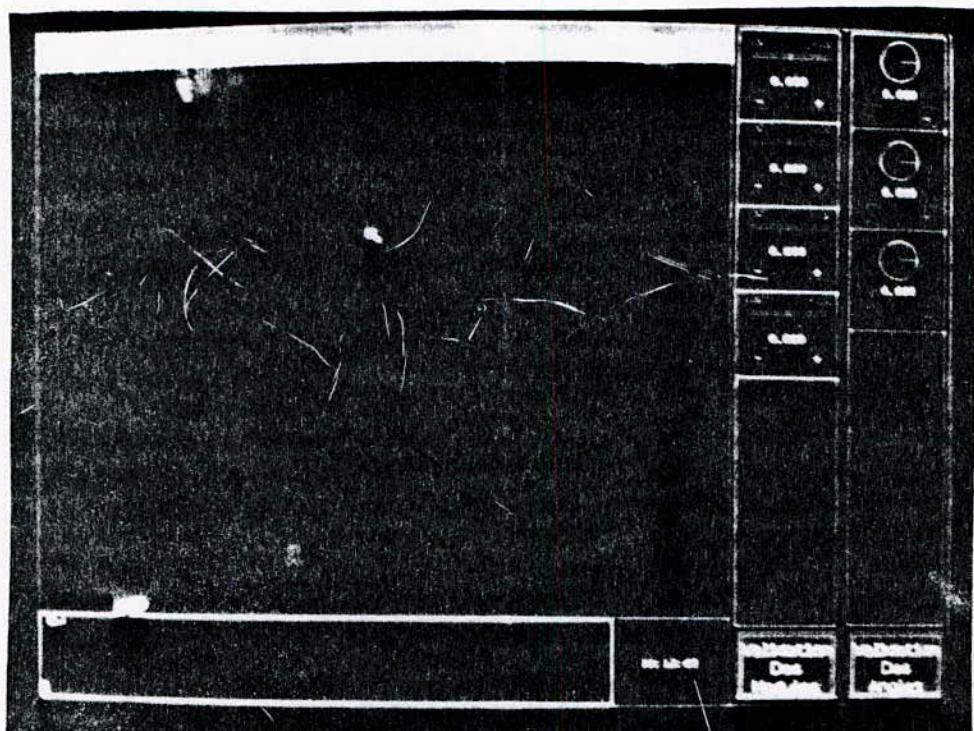


fig .5 Configuration initiale

Après la création de cette configuration ,le message " LOCIM Ready " apparait .Le logiciel est prêt à être utilisé.

3- Menu Principal

Le menu principal se présente comme suit :

Fichier	Visualisation	Configuration	Saisie	Classement	Quit
---------	---------------	---------------	--------	------------	------

Validation de la souris :

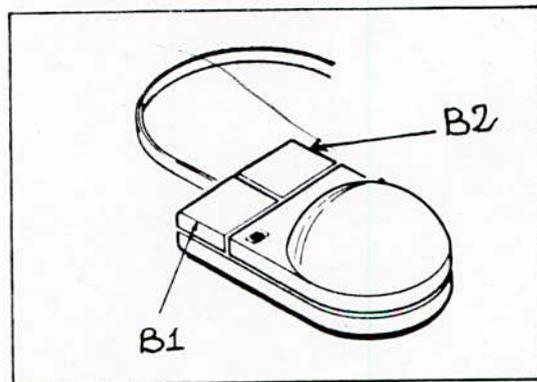


fig. 6 Présentation des boutons de la souris

Toutes les opérations sont sélectionnables à l'aide du bouton B1 de la souris .A chaque fois qu'une option est choisie un " Pop-menu " ou sous menu apparait .Il comprend à son tour plusieurs options qui sont ainsi sélectionnables à l'aide du bouton B1 .Ce " Pop-menu " se présente pendant une minute .Pour abandonner et revenir au menu principal, on utilise le bouton B2 de la souris .

4-1 OPTION FICHIER

FICHIER
Lire Fichier
Caractéristique Fichier
Caractéristique Classes
Sauvegarde Fichier Classé
Help

4-1-1 OPTION " Lire Fichier "

Cette option sert à introduire le fichier contenant les données à traiter . Une fois cette option validée , le message " *Nom. de fichier de données :*" apparait sur la fenêtre texte.

La dimension de l'espace des données se trouvant dans le fichier de données étant lue ,la représentation de cet espace ,par des segments représentant les axes et par des cercles en camemberts représentant les angles , est affichée (*fig 3*). Notons qu'initialement l'observateur se trouve à l'origine de l'espace avec le premier axe comme direction du regard.

4-1-2 Option " Caractéristique fichier "

Cette option permet d'avoir les informations suivantes sur le fichier de données :

- Nom de fichier de données
- Nombre d'observations
- Nombre de paramètres

Ces informations sont visualisées sur la fenêtre texte .

4-1-3 Option " Caractéristiques Classes "

Après avoir introduit au clavier le numéro de la classe à laquelle on s'intéresse , les informations suivantes apparaissent :

- Numéro de la classe
- Nombre d'observations dans cette classe
- Nombre de paramètres
- Vecteur moyen de cette classe

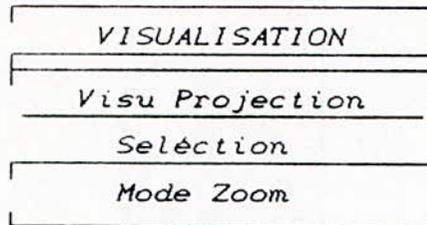
4-1-4 Option " Sauvegarde du fichier classée "

Une fois la classification décidée , l'opérateur peut sauvegarder le résultat de la classification dans un fichier dont la structure est décrite dans le paragraphe § 5 .

4-1-5 Option " Help "

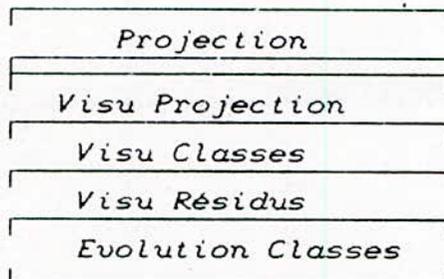
Cette option permet à l'utilisateur d'avoir une aide sur l'utilisation globale du logiciel .

4-2 OPTION VISUALISATION



4-2-1 Option " Visu Projection "

Validée ,cette option conduit au sous menu suivant :



L'option " Visu projection " permet de visualiser le nuage de points projetés dans le plan *Module/Angle* .La fin de cette visualisation est signalée par le message " Fin Projection " .

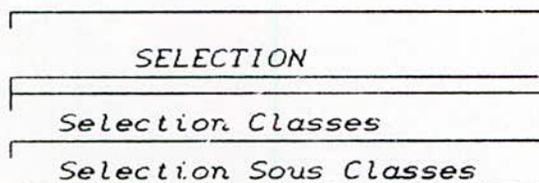
L'option " Visu Classes " n'est activée que si une classe est sélectionnée au préalable .Le numéro de la classe est alors demandé .

L'option " Visu Résidus " permet de visualiser le reste des points qui n'appartiennent pas à la classe sélectionné.

L'option "Evolution Classes " offre la possibilité de suivre l'évolution d'une classe spécifiée ,en fonction de la variation de l'observateur et du référentiel .Cette classe est dans une couleur differente de la couleur initiale .

4-2-2 Option " Sélection "

Validée cette option conduit au sous menu suivant :



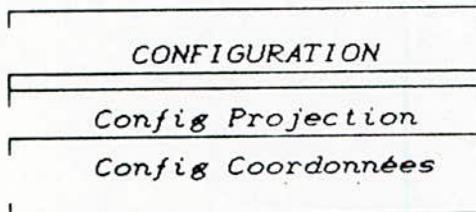
L'option " Selection Classes " permet de délimiter un groupe de points susceptibles de former une classe et de les sauvegarder dans un tableau .

La classe est délimitée par un polygone formé par la jonction des points choisis à l'aide de la souris ,le polygone est fermé en appuyant sur le bouton B2 puis sur le bouton B1 lorsque on est au dernier point .

L'option " Selection Sous Classes " permet d'affiner l'analyse ,en sélectionnant des sous ensembles dans une classe et ceci en passant par le Mode Zoom .

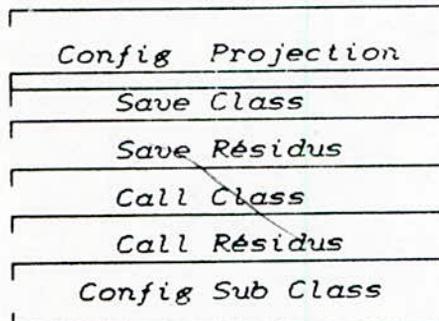
4-3 OPTION CONFIGURATION

Une fois choisie, le sous menu suivant apparait.



4-3-1 Option " Config Projection "

Validée, cette option conduit au sous menu suivant:



L'option "Save class " permet de sauvegarder la classe que l'on spécifie par son numéro dont on précisera le nom .

L'option " Save Residus " sert à sauvegarder le reste du nuage sans la classe spécifiée , comme à l'étape précédente.

L'option "Call Class " assure le chargement d'une classe spécifier par son nom déjà existant dans le disque dur.

L'option " Call Residus " joue le même rôle que la précédente mais appliquer au residus de la classe en question .

L'option " Config Sub Class " interesse les sous-classes et offre le menu suivant:

<i>Config Sub Class</i>
<i>Save Sub Class</i>
<i>Save Sub Residus</i>
<i>Call Sub Class</i>
<i>Call Sub Residus</i>

Ces options jouent le même rôle que les précédentes ,mais appliquées à des sous-classes.

4-3-2 option " Config-Coordonnes ".

Validée ,cette option conduit au sous-menu suivant :

<i>Config Coordonnes</i>
<i>Save Angle</i>
<i>Save Module</i>
<i>Save Ag & Module</i>
<i>Call Angle</i>
<i>Call Module</i>
<i>Call Ag & Module</i>

Ces fonctions sont très utilisées dans la classification interactive ,elles permettent de sauvegarder dans la memoire de masse la position et la direction de vecteur regard .

L'option " Save Angle " permet de sauvegarder les angles qui specifient la direction du vecteur regard .

L'option " Save Module " permet la sauvegarde de la position du vecteur regard .

L'option " Save Angle & Module " realise les deux fonctions précédentes .

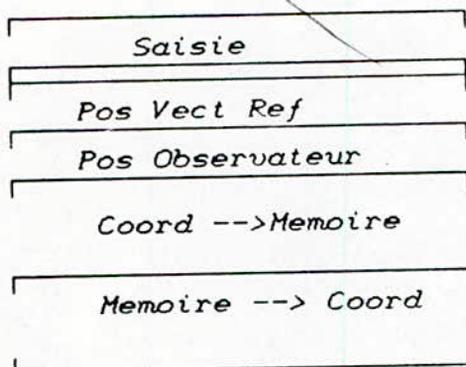
L'option " Call Angle " permet de ramener du disque dur les valeurs des angles qui specifient la direction du vecteur regard

L'option " Call Module " permet de ramener du disque dur la position du vecteur regard.

L'option " Call Ag & Module" realise les deux fonctions précédentes .

4-4 SAISIE

C'est une option specialisé dans la modification de la position et la direction du regard de l'observateur .Elle offre le sous-menu suivant:



L'option " Pos Vect Ref " permet de changer la direction du regard de l'observateur. Une fois sélectionnée la souris est placé dans la fenetre de visualisation de l'état du referentiel .

On peut soit incrementer ou decrementer l'angle que fait l'observateur par rapport à chaque axe de l'espace multidimensionnel et cela en plaçant la souris dans le carré contenant soit le signe + ou - et en appuyant sur le bouton B1 .La variation de l'angle est visualisée. Une fois la saisie faite on rentre dans la case de validation pour valider (B1)

L'option " Pos Observateur " permet de choisir la position du vecteur regard .La saisie se fait comme pour les angles .Les variations de la position de l'observateur est visualisée graphiquement à l'aide d'un cavalier qui se déplace sur une barrette . La valeur numerique de cette position est affiche.

L'option " Coord-->Memoire " permet de sauvegarder en memoire vive les coordonnées cartésiennes de l'observateur ainsi que les coordonnées angulaires du vecteur regard .

L'option "Memoire-->Coord " permet de faire le chemin inverse de l'option précédente .

4-5 CLASSEMENT

Une fois validée, cette option conduit au sous-menu suivant:

Classement
Classement d'un Objet
Classement d'un Fichier
Visualisation

L'option "Classement d'un Objet " permet de classer un nouveau objet dans la classe correspondante et cela apres l'introduction du vecteur de forme de cet objet au clavier . Cet objet est visualise d'une couleur differante dans la projection Module/Angle.

L'option "Classement d'un fichier " permet de classer un ensemble d'objets existant dans un fichier ,en se basant sur la classification qui a été faite .

L'option " Visualisation " permet de visualiser l'ensemble des objets dans la projection Module/Angle .

4-6 QUIT

Cette option permet de quitter le logiciel par une mise en garde de l'utilisateur et ceci on confirmant son choix par le biais du menu suivant:

Quitter Le Logiciel
Oui
Non

5- STRUCTURE DES FICHIERS UTILISES

On dispose de deux types de fichiers, les fichiers de données brutes et les fichiers résultats. Les fichiers de données contiennent des données quantitatives, ceux sont des tableaux (N,P) où N est le nombre d'individus sur les quels on a mesuré P paramètres. La valeur de chaque variable est donnée par :

$$X_i(j), i=1,2,3,\dots,N \text{ et } j=1,2,3,\dots,P.$$

La structure de ces fichiers est la suivante :
On donne dès le début le nombre d'observations et le nombre de paramètres. On aura alors la forme suivante :

N , P

$X_1(1), X_2(1), X_3(1) \dots X_p(1)$

$X_1(2), X_2(2), X_3(2) \dots X_p(2)$

.. .. "

.. .. "

.. .. "

$X_1(N), X_2(N), X_3(N) \dots X_p(N)$

les fichiers résultats sont présentés sous la structure suivante :

N, P, Nbc

$N_1, N_2, N_3, \dots, N_{Nbc}$

$X_1(1), \dots, X_p(1)$

" " " "

" " " "

$X_1(N_1), \dots, X_p(N_1)$

" " " "

" " " "

$X_1(N_1+N_2), \dots, X_p(N_1+N_2)$

" " " "

" " " "

$X_1(N), \dots, X_p(N)$

où N et P sont les mêmes paramètres que le tableau de données, Nbc est le nombre de classes obtenues. $N_1, N_2, N_3, \dots, N_{Nbc}$ sont les nombres d'individus dans chaque classe et $X_i(j)$ sont les éléments de chacune des classes.

VI- APPLICATION ET RESULTATS

1- TESTE DE LA VALIDITE DE LA METHODE

Dans le but de tester la méthode présentée, nous avons utilisé un fichier appelé fichier *IRIS /12/*. C'est un fichier de données classique utilisé pour la première fois par Fisher. Il est ensuite devenu un fichier de référence que d'autres chercheurs utilisent pour tester leurs propres méthodes de classification. Ce fichier contient des données relatives aux trois espèces d'*IRIS* suivantes : *L'iris setosa*, *l'iris versicolor* et *l'iris virginica*.

Ce fichier contient un échantillon de cent cinquante fleurs d'iris de façons que chacune des trois espèces d'iris ai 50 observations, une observation correspond à la mesure des quatres paramètres suivants : la longueur des pétales, la largeur des pétales, la longueur des sépales et la largeur des sépales.

Le nuage des cent cinquante points est projeté sur le plan *Module/Angle* tel qu'il apparait à la figure 7.

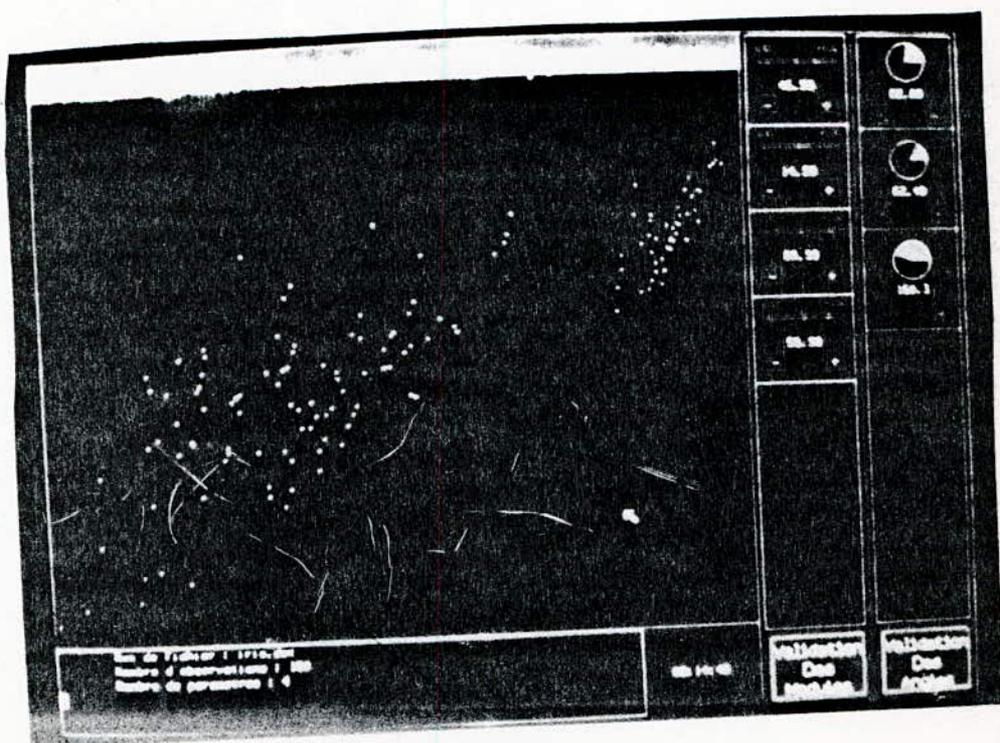


fig .7 Projection initiale

Sur la figure (7) deux classes apparaissent distinctement , la première est formée de 50 iris appartenant à l'espèce *sétosa* , la deuxième est composée de 100 iris regroupant les deux races restantes . Cette séparation paraît évidente lorsqu'on inspecte les données originales . Nous avons sélectionné cette classe dans le but de séparer les deux classes restantes et nous obtenant en fin de classification la projection représentée par la figure (8). Les deux classes relatives aux iris *virginica* et *versicolor* sont très semblables et difficiles à distinguer, car les deux races sont très proches .

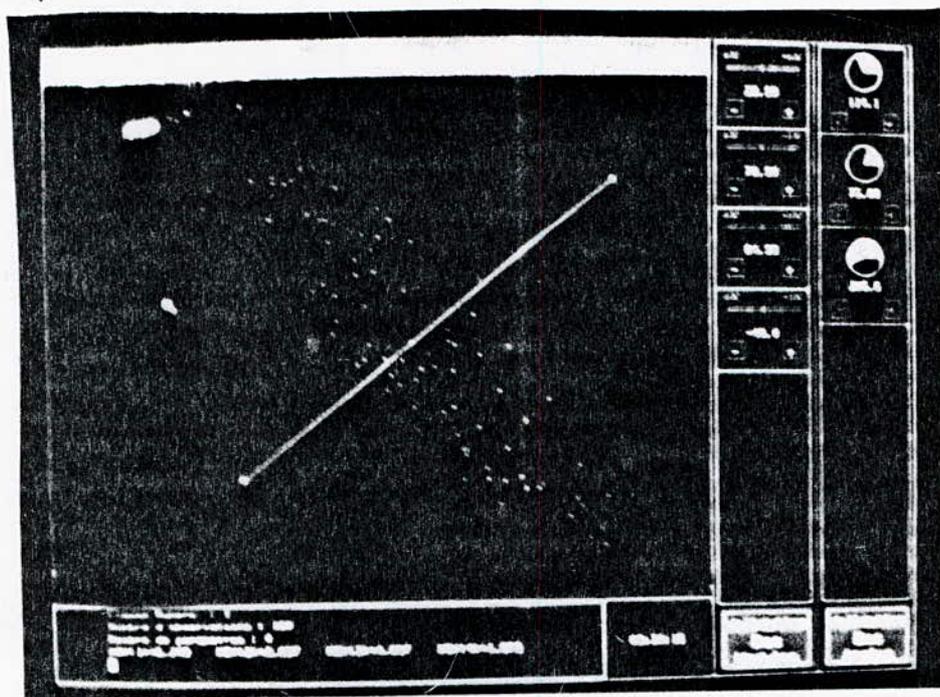


fig. 8 Projection des 100 points

En intervenant en tant qu'opérateur , on arrive à peine à séparer les deux classes.

Des résultats analogues ont été obtenus par d'autres chercheurs /6,12,14,15/ .

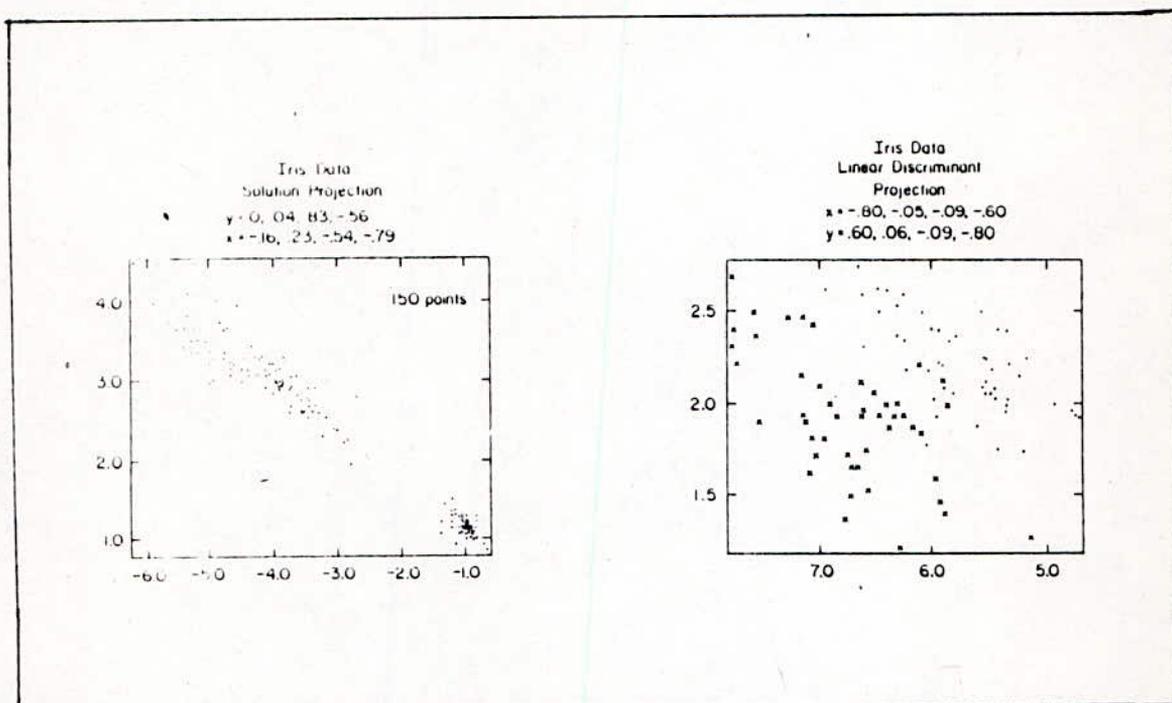


fig.9 nuage des 150 points

fig.10 nuage des 100 points

Ces figures sont semblables à celles que nous avons obtenu à la figure (7) et (8).

2- ETUDE D'UN GROUPE D'ABEILLES

C'est un fichier d'observations biometriques faites sur des abeilles .Les parametres mesurés sont , la coloration ,la pilosité ,le tomentum,la longueur de la langue et l'index cubitale .

A la première projection (*Module/Angle*) ,on observe quatre classes d'observations distinctes (*fig .11*) .

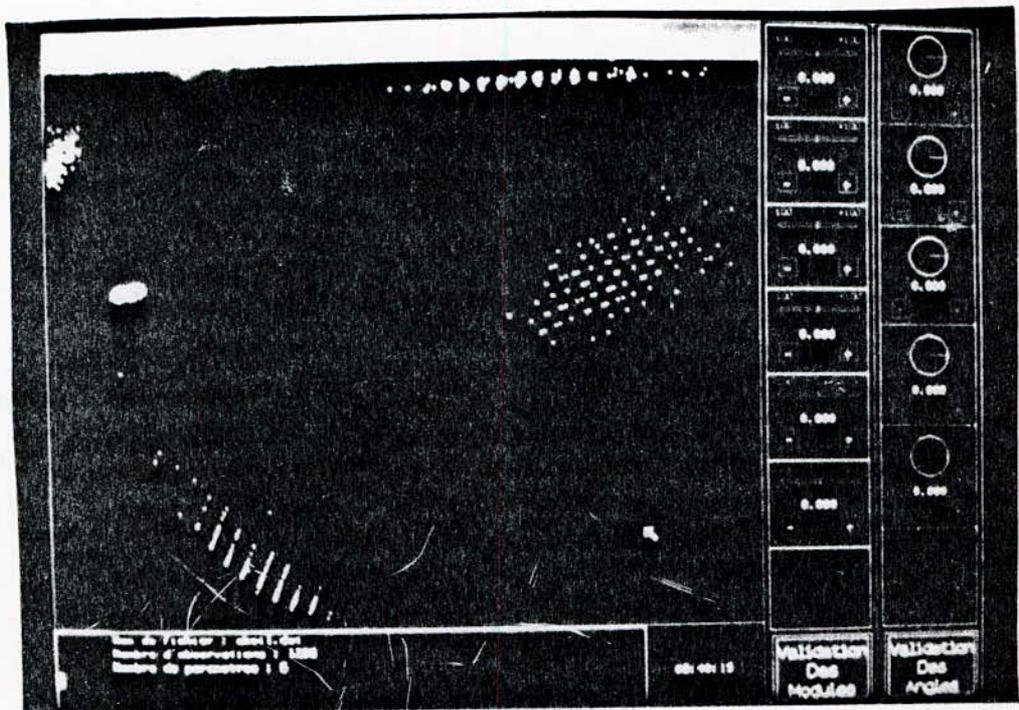


fig. 11 Configuration initiale du référentiel

Après une modification de l'état de l'observateur la classe la plus dense se divise en deux classes proches l'une de l'autre (fig. 12) .

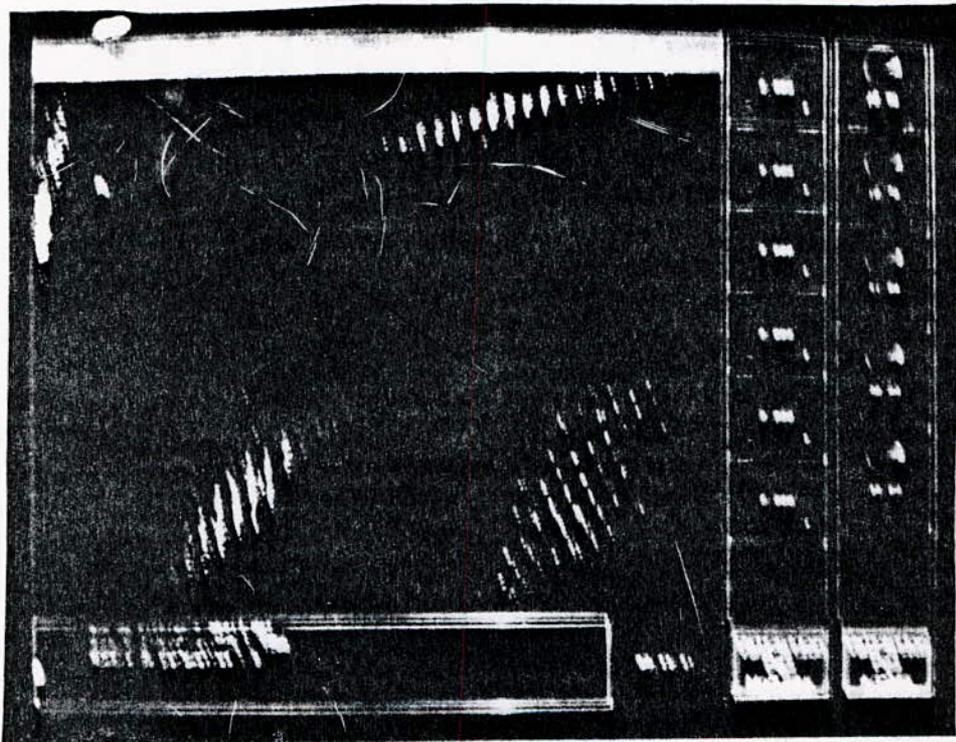


fig.12 Nouvelle configuration du référentiel

Le déplacement de l'origine à proximité de cette classe augmente le pouvoir de séparation angulaire. L'application d'un zoom permet de focaliser ainsi une classe supplémentaire qui apparaît nettement suivant plusieurs configurations du référentiel (fig. 13)



fig. 13 zoom sur la classe

En conclusion ,on peut affirmer que le fichier contient cinq espèces d'abeilles , dont les caractéristiques sont résumées dans les tableaux suivants :

	Moyen	Ecartype	Nombre d'obs
Classe 1	13.76	14.02	308
Classe 2	9.91	12.15	278
Classe 3	11.45	13.44	300
Classe 4	13.85	13.30	220
Classe 5	34.52	10.54	160

Caracteristiques des classes vue par le premier parametre

	Moyen	Ecartype	Nombre d'obs
Classe 1	16.75	08.81	308
Classe 2	16.72	02.57	278
Classe 3	16.80	04.30	300
Classe 4	16.22	03.31	220
Classe 5	14.26	02.25	160

Caracteristiques des classes vue par le deuxieme parametre

	<i>Moyen</i>	<i>Ecartype</i>	<i>Nombre d'obs</i>
<i>Classe 1</i>	20.17	04.27	308
<i>Classe 2</i>	19.26	01.57	278
<i>Classe 3</i>	19.34	01.63	300
<i>Classe 4</i>	19.57	01.79	220
<i>Classe 5</i>	20.61	01.90	160

*Caracteristiques des classes vue par
le troisieme parametre*

	<i>Moyen</i>	<i>Ecartype</i>	<i>Nombre d'obs</i>
<i>Classe 1</i>	73.62	02.16	308
<i>Classe 2</i>	73.26	02.12	278
<i>Classe 3</i>	73.43	02.22	300
<i>Classe 4</i>	73.80	02.06	220
<i>Classe 5</i>	76.48	01.61	160

*Caracteristiques des classes vue par
le quatrieme parametre*

	<i>Moyen</i>	<i>Ecartype</i>	<i>Nombre d'obs</i>
<i>Classe 1</i>	333.74	29.65	308
<i>Classe 2</i>	32.66	03.26	278
<i>Classe 3</i>	32.78	03.27	300
<i>Classe 4</i>	335.13	29.14	220
<i>Classe 5</i>	34.52	03.31	160

*Caracteristiques des classes vue par
le cinquieme parametre*

	<i>Moyen</i>	<i>Ecartype</i>	<i>Nombre d'obs</i>
<i>Classe 1</i>	18.22	12.04	308
<i>Classe 2</i>	176.18	16.47	278
<i>Classe 3</i>	172.09	21.23	300
<i>Classe 4</i>	174.02	18.65	220
<i>Classe 5</i>	135.87	17.92	160

*Caracteristiques des classes vue par
le sixieme parametre*

CONCLUSION GENERALE

Après avoir vu les généralités sur les méthodes de classifications automatiques et interactives , nous avons programmé en langage C la méthode *Module/Angle* qui correspond à la description de l'observateur qui se déplace dans un espace multidimensionnel .Par sa simplicité cette méthode nous permet de visualiser les classes sans choisir le plan de projection comme on a l'habitude de le faire dans les autres méthodes .

En déplaçant l'observateur les classes formées évoluent et la décision de garder telle ou telle configuration est remise à l'utilisateur .

Tester sur différents types de fichiers la méthode a donné des résultats satisfaisants . Appliquée au fichier des fleurs d'iris qui constitue un test pour toutes les méthodes de classification , la méthode donne des résultats comparables à ceux fournis par d'autres considérées très performantes . Cette méthode fonctionne aussi efficacement lorsqu'il s'agit de fichiers contenant un nombre très élevé de données tel que le fichier de 1336 abeilles à six paramètres auquel nous l'avons appliquée .

Une procédure de classement d'une nouvelle donnée à été prévue . La visualisation graphique offerte par l'écran couleur à haute définition du hp 9000 utilisé convient parfaitement à la méthode .

En perspective ,il serait intéressant de remplacer l'interactivité de la méthode par un calcul optimal de séparation des classes .

BIBLIOGRAPHIE

- [1] K.Diday , J.Lemaire , J.Pouget , F.Tester
" *Element d'Analyse De Données* "
Edition Dunod -1982-
- [2] I.C.Lerman
" *Classification Et Analyse Ordinale Des Données* "
Edition Dunod -1982-
- [3] Bertier P ,Bouroche J.H
" *Analyse Des Données Multidimensionnelles* "
P.U.F Paris -1977-
- [4] Jaumbu M.
" *Classification Automatique Pour l'Analyse Des Données* "
1 -Methodes & Algorithmes
Edition Dunod -1980-
- [5]
" *A Comparison Of Seven Techniques For Choosing Subsets Of* "
Pattern Recognition Properties
IEEE Trans On Computer Vol C-20 N=9 Sep 71
- [6]
" *Some Studies In The Interactive Design Of Character* "
Recognition Systems
IEEE Trans On Computer Vol C-20 Sep 71
- [7] Aderito Pastor
" *Methode Pratiques Pour La Reconnaissance Des Formes* "
Université De Clermons II
- [8] Caillez
" *Intoduction à l'Analyse De Données* "
Edition Dunod

- [9] Jean-marie Pinon
 " *Reconnaissance De Formes Dans l'Electroencephalogramme* "
 (These De Docteur Ingenieur 29-8-1980)
 Institut National Des Sciences Appliquées De Lyon
- [10] Al .Kelly , Iraponl
 " *A Book On C* "
 The Benjamin / Cummings .
- [11] Kernigham.W ,Ritchie.D
 " *Langage C* "
 Paris Masson 1988
- [12] Chien
 " *Pattern Recognition problem* "
- [13] Philippe Schweizer
 " *Infographie Tome II* "
 Presses Polytechniques ROMANDES
- [14] J.H.Friedman & J.W.Tukey
 " *A Projection Pursuit Algorithm For Exploratory* "
Data Analysis
 IEEE Trans Comput, Vol C-23 Sept 1974
- [15] K.Fukunaga,D.R.Olsen
 " *A Two Dimentionnel Display For The Classification Of* "
Multivariate Data
 IEEE Trans Comput ,Vol C-20 August 1971
- [16] J.P et F.Benzécri
 " *T1-Analyse des correspondances et classification* "
 Edition Dunod
- [17] M.Roux
 " *Algorithmes de classifications* "
méthodes + programmes

ANNEXE A

1- PRESENTATION DE LA STATION hp 9000

a- Unite centrale

- Monoprocasseur 32 bits
- Coprocasseur 32 bits
- Memoire centrale 16 MO
- Mot memoire 64 bits
- Carte multiplexeur X2 (6 terminaux par carte)

b- Peripherique

- Unite disque :
 - Drive 0 : HP 7933 (404 MO)
 - Drive 1 : HP 7933 (404 MO)
- Imprimante Graphique
 - HP 2563 line printer (300 ligne/mn)
- Carte graphique : CGT (Color Graphic terminal
 - Accelérateur graphique
 - Hight resolution 1280 1024
 - Medium resolution 1024 768

c- Logiciels

- Systeme d'exploitation HP-UX
- Structure arborescente (notion de chemin et sous-repertoire)
- windows 9000 (logiciel de gestion de fenêtres)
- Fichiers (3 types) :
 - * Fichiers systemes
 - * Fichiers ASCII
 - * Fichiers specialises

- Compilateurs :

- * Compilateur C
- * Compilateur Fortran 77
- * Compilateur Pascal

- Utilitaires :

- * Allbase (SGBD SQL, IMAGE)
- * Starbase (Graphisme)

2- CARACTERISTIQUE DU SYSTEME UNIX

- C'est un systeme multi-utilisateurs et multi-taches .
- C'est un systeme modulaire .
- E/S generalisés (ils sont integrees au systeme de fichiers) .
- Systeme de fichiers hierarchiques .

Il possede aussi un interpreteur de commandes qui est une interface entre l'utilisateur et le systeme d'exploitation " SHELL " .

Les commandes SHELL concernent principalement la gestion des terminaux , le controle des spools , la gestion du systeme , et la manipulation des fichiers et des repertoires .

UNIX comprends aussi une bibliotheque riche en utilitaires .

3- PRESENTATION DU " WINDOWS 9000 "

Nous avons utilise pour la mise en oeuvre de l'interface utilisateur le logiciel utilitaire de programmation " windows 9000 ". En effet ,celui-ci fournit une base de procedures permettant a la fois le multifenêtrage l'utilisation de la souris .

```

/* Test d'inclusion des points dans un polygone */

void Test (int NIN, Xin, Yin, Nout, Xout, Yout)
int     Nin, Nout;
float   Xin[500], Yin[500], Xout[500], Yout[500];

/* Nin      Nombre de points du nuage initiale          *
 * Xin, Yin  Coordonnes du nuage de point initiale      *
 * Nout      Nombre de point a l'interieur du polygone  *
 * Xout, Yout Coordonnes des points al'interier du polygone */

int     KC, i, j, h, gh, ContSel;
float   a[20], b[20], Xp[20], Yp[20], gard[20];
float   MinS, MaxS, Dem, test;

/* KC      Nombre de sommets du polygone                *
 * a, b    parametre de l'equation des segment du polygone *
 * Xp, Yp  Coordonnes des sommets du polygone          */

/* Calcul des coefficients a(i), b(i) des segments de droites *
 * du polygone  /=a(i)*x+b(i)                               */

for (i=1; i<=KC; i++) {
    if ( i==KC ) j=1;
    else j=i+1;
    if ( Xp[i]==Xp[j] ) {
        a[i]=30;
        gh=1;
        gard[i]=Xp[i];
    }
    else {
        a[i]=(Yp[j]-Yp[i])/(Xp[j]-Xp[i]);
        b[i]=Yp[i]-a[i]*Xp[i];
    }
}

/* Selection des points qui sont a l'interieur du polygone */

for (i=1; i<=Nin; i++) {
    ContSel=0;
    for (j=1; j<=KC; j++) {
        if ( j==KC ) h=1;
        else h=j+1;
        if ( Yp[i]>Yp[j] ) MaxS=Yp[i], MinS=Yp[j];
        else MaxS=Yp[j], MinS=Yp[i];
        if ( a[j]==30 ) Dem=gard[gh];
        else {
            if ( a[j]!=0 ) {
                Dem=(Yin[j]-b[j])/a[j];
            }
            if ( Yin[j]>=MinS && Yin[j]<=MaxS && Dem>=Xin[j] )
                ContSel++;
        }
    }
}

```

```

    * Calcul des coordonnées de la projection Module/Angle */
void calc_coor (Nin, Din, Nout, Dout)
int Nin, Nout;
float Din[8][1000], Dout[8][1000];

/* Nin      Nombre d'observations      *
 * Din      tableau de données          *
 * Xproj, Yproj  Coordonnées de la projection M/A */

float me_deg *3.1416/180
float me_rad *180/3.1416
int i, j, Npar;
float pob[8], ref[8], alf[8];
float t1[8][8], s1, s2;

/* Npar      Nombre de parametre d'un individu      *
 * pob      Coordonnées cartesiens de l'observateur      *
 * alf      Coordonnées angulaires du vecteur référence *
 * ref      Coordonnées cartesiens du vecteur référence */

    * Calcul des coordonnées cartesiens du vecteur référence */
t1[i][1]=cos(alf[i][2] deg);
t1[i][2]=sin(alf[i][2] deg);
for (j=1; j<=Npar; j++) {
    t1[i][j]=cos(alf[i][j+1] deg);
    for (k=1; k>=0; k--) t1[i][k]=t1[i][k-1]*sin(alf[i][j+1] deg);
}
for (i=1; i<=Npar; i++) ref[i]=t1[i-1][Npar-1];
    * Calcul des coordonnées de la projection */
for (i=1; i<=Nin; i++) {
    s1=s2=0;
    for (j=1; j<=Npar; j++) {
        s1+=(Din[i][j]-pob[j])**2;
        s2+=(Din[i][j]-pob[j])*ref[j];
    }
    Xproj[i]=sqrt(s1);
    Yproj[i]=acos(s2/Xproj[i]) rad;
}

```