

C-540

Le Traitement Automatisé de l'Image

(DOCUMENTATION ET RECHERCHES)

Textes réunis par Francis LEVY

Avec la collaboration du B.N.I.S.T. (Bureau National de l'Information Scientifique
et Technique) et de la Mission à l'Informatique au Ministère de l'Industrie,
du Commerce et de l'Artisanat.



INTERPHOTOTHEQUE N° SPECIAL 33-34 - Décembre 1977

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

AVERTISSEMENT

Nous tenons à prévenir d'emblée le lecteur qu'il ne trouvera pas ici une étude d'ensemble élaborée par un seul auteur. Le domaine couvert est si vaste, les spécialités si tranchées qu'il n'était pas possible d'envisager sérieusement une telle démarche.

Notre objectif a été de présenter un *panorama raisonné* des études et réalisations françaises en cours dans le domaine de ce que l'on pourrait appeler «l'Informatique et l'Image» ou plutôt «l'Informatique au service de l'Image». Certes la rédaction et la fabrication de l'ouvrage ont imposé des délais et depuis, sans doute, des voies nouvelles ont été explorées. Mais cette publication présente cependant un ensemble de *références* très précieux, puisque jusqu'à ce jour aucune approche globale du sujet n'avait été tentée.

Ce travail devrait être suivi de *mises à jour* régulières portant sur le développement des travaux, l'évolution des équipes et des matériels : informations indispensables au chercheur comme au «décideur» qui s'interroge sur les directions à prendre. De même, il serait souhaitable qu'un ensemble «d'états de la question», de vues synthétiques par domaine soit rédigé par des spécialistes. Une telle politique coordonnée de l'information nécessite *des moyens et du temps*. Le Bureau National de l'Information Scientifique et Technique, la Documentation Française, la Mission à l'Informatique au Ministère de l'Industrie, du Commerce et de l'Artisanat s'orientent dès à présent dans cette voie.

Indiquons dès maintenant que nous recevrons avec reconnaissance toutes nouvelles informations sur les travaux présentés dans cet ouvrage ainsi que toutes indications sur les réalisations qui ont pu nous échapper ou qui ont pris naissance depuis 1976.

Nous tenons à remercier très vivement tous les auteurs ainsi que le B.N.I.S.T., la Mission à l'Informatique, sans lesquels cette publication n'aurait pu se faire.

LA DOCUMENTATION FRANCAISE

BIBLIOTHEQUE DU CERIST

INTRODUCTION : Francis LEVY

7

SCIENCES DE LA NATURE

Disciplines

Astronomie

A. BIJAOUÏ : Le traitement des images dans l'astronomie française

13

A. LABEYRIE : Traitement d'images interférométriques en astronomie

23

Biologie

R.J.P. LE GO : Système d'analyse d'images cytologiques et chromosomiques par ordinateur numérique.

35

Chimie

J.E. DUBOIS : Le système DARC

63

Physique

B. EQUER - G. FONTAINE : Le traitement d'images en physique des hautes énergies

73

H. VIDEAU : Le traitement de l'image en physique nucléaire des hautes énergies

95

Reconnaissance des formes

J. SERRA : Présentation de la morphologie mathématique.

111

Applications

Electronique

B. CARCENAC DE TORNE : Matériels d'aide au tri postal

123

H. NUE - R. BOUILLOT : Système de classement des documents adopté pour les images photographiques.

129

SCIENCES HUMAINES

Disciplines

Anthropologie

A. GUENOCHÉ : Analyse et validation expérimentale du raisonnement en œuvre dans les études sociales (Projet AVEROES).

133

P. RICHARD : Le système descriptif des affiches au musée des Arts et Traditions Populaires.

153

Archéologie

M.S. LAGRANGE : Un aspect du traitement des images : les codes analytiques et répertoires élaborés au Centre d'Analyse Documentaire pour l'Archéologie du C.N.R.S.

173

Applications

Aménagement du territoire et urbanisme

M. GRIFFON - Y.A. JOUAN - D. LOUCHART : Automatisation documentaire et analyse de l'image à DOCAMENOR et au CETE Nord-Picardie.

199

M. HENRY - M. HERCELIN - J. CHAUMIER : Le système Urbathèque de l'(I.A.U.R.I.F.) (Institut d'Aménagement et d'Urbanisme de la Région Ile de France).	217
B. LECOUR - GRANDMAISON : La diathèque du Secrétariat des Missions d'Urbanisme et d'Habitat. (S.M.U.H.).	255
Architecture et construction	
J.M. BROUARD - J.L. SAPINIA : La recherche des documents par mini-ordinateur applications à la photographie.	259
Archives historiques	
J. LE SEIGNEUR : Cinémathèque et Documentation	277
J.M. PEROUSE DE MONTCLOS : Les centres de documentation de l'Inventaire général des monuments et richesses artistiques de la France.	287
Cartographie	
F. BRUN - J. DENEGRE - B. PASQUIER : Recherches en cartographie et photo-interprétation	293
L. CAIRO - Y. VUILLAUME : Traitement des images de télédétection multispectrale en hydrogéologie.	308
F. DELACROIX - Y.F. THOMAS - F. VERGER : Classification des paysages à partir des données multispectrales de télédétection.	329
Médecine	
P. GOLDBERG - M. CHAVANCE - D. SAMSON-DOLLFUS : Analyse temporelle de l'électroencéphalogramme.	339
J. INGRAND, C. SÉPÉ, F. RAVARD, J. FONDARAI : Description du système de recherches documentaires de la diathèque de l'Institut PAOLI-CALMETTES (Centre régional de lutte contre le cancer). Marseille.	349
Pédagogie	
C. BERTHET : Une expérience d'application des principes de l'automatique documentaire au traitement de l'image fixe.	357
Procédés audiovisuels	
R. MAES : Le système IMAGO : Système documentaire de l'Institut National de l'Audio-visuel.	375
Recherches criminelles	
G. PROUTEAU : Système PROSAM	389
Technique documentaire	
H. HUDRISIER - P. SANSON : Projet pour l'utilisation de l'analyse multidimensionnelle dans une iconothèque automatisée.	401
autres applications	
R. DARRIGO : La visualisation automatique des données statistiques en vue du traitement et de la communication de l'information.	423
M. REBUFFET : Le traitement d'images à l'E.T.C.A. (Etablissement Technique Central de l'Armement). - Conception de système de traitement d'images.	451

INTRODUCTION

Cet ouvrage manifeste un des traits supplémentaires de ce que l'on appelle la «civilisation de l'image». Il n'est pas de domaine en effet où l'image ne joue un rôle qui dépasse de loin l'illustration ou la distraction : aujourd'hui elle informe, elle est à la fois la matière et le produit du développement scientifique et technique.

Dès 1976, la Documentation Française se proposait de publier un numéro spécial de la revue «Interphotothèque» portant sur la classification et l'exploitation de corpus de photos et de films. Il s'agissait alors de problèmes documentaires au sens strict. Il apparut que ce sujet n'était qu'une partie d'un domaine plus vaste, celui des questions soulevées par le classement, documentaire ou non, des images. D'accord avec le Bureau National de l'Information Scientifique et Technique (B.N.I.S.T.) on entreprit de dresser un inventaire, aussi complet que possible, des études et réalisations françaises.

L'ensemble est fait de diverses contributions, rédigées par les spécialistes dans leurs branches respectives, sous leur responsabilité.

Les «compilateurs» se sont efforcés, pour leur part, de circonscrire les limites du sujet et de suggérer un mode de présentation.

Compte-tenu de la très générale bonne volonté des auteurs, à laquelle l'équipe de la Documentation Française tient à rendre hommage, la disparité dans la taille et le vocabulaire des contributions tient à la variété des problèmes rencontrés.

On souhaite que le lecteur surmonte ces difficultés : il y trouvera, pense-t-on, matière à d'inédites et fructueuses comparaisons.

Pour les illustrer on s'efforcera ici d'exposer les grands traits d'études et réalisations en quatre domaines. Elles ont été choisies pour leur valeur exemplaire, en admettant toutefois qu'elles ne représentent pas la totalité des problèmes posés. Les domaines d'application relèvent respectivement des «sciences humaines» (Urbanisme, Archéologie et Géographie) et des «sciences exactes» (Physique).

I – Urbanisme : systèmes documentaires

Très typiquement les matériaux considérés comme utiles aux urbanistes et planificateurs comportent photographies et films. Le nombre des documents «représentatifs» s'accroît constamment, et dépasserait vite le million si, fort justement, le sujet n'en était circonscrit.

Une des contributions, par exemple, limite la masse documentaire aux images concernant la seule région parisienne et les grandes agglomérations.

Le «stock» en est réduit et varie généralement entre quelques dizaines de milliers et quelques centaines de milliers d'unités documentaires.

Les deux premiers problèmes auxquels se trouve confronté le documentaliste portent sur la définition du domaine et sur les modalités de la collecte.

Tous deux sont résolus de manière empirique. C'est par approximation successive, compte-tenu d'un «pari» initial et du degré ultérieur de satisfaction des usagers que le sujet est correctement défini. Quant à la collecte, elle dépend des institutions maîtresses de l'œuvre, de leurs moyens et des rapports qu'elles entretiennent avec les «producteurs» d'images. Remarquons seulement que la très grande disparité des supports (photographies sur papiers, diapositives, microfiches, etc.) et des formats empêche pour l'instant, de donner un sens général aux pratiques de fait.

Il n'en est pas de même pour ce problème central qui consiste, après archivage et stockage, à ranger les images pour retrouver celles qui intéressent l'utilisateur. Les images qui composent le fonds documentaire ont été retenues par appréciation intuitive du sens qu'elles sont censées receler. Ce premier tri était global : celui qui est nécessaire pour répondre à une «question documentaire» est de même nature mais nécessairement plus fin. C'est dans le stock même qu'il faut extraire le ou les documents pertinents.

La difficulté n'est pas d'ordre matériel. A supposer même que les images soient réduites à un format unique le désordre de la collection n'en subsisterait pas moins. La solution classique et fort ancienne qui consiste à donner des noms à ces objets est alors adoptée.

L'ensemble des vocables choisis, autrefois appelé «classification» s'intitule souvent aujourd'hui «liste de mots-vedettes, de mots-clés, de descripteurs», et plus généralement «thesaurus» ou «langage documentaire». Cette terminologie a ses raisons. Il fut un temps où la classification pouvait prétendre à une vocation universelle. La rapidité et l'accroissement des spécialisations la rendent inopérante dans la plupart des cas. D'où la nécessité de créer des «outils intellectuels» de classement plus fins, plus souples, plus complexes. C'est là encore une activité toute empirique : des «mots» sont choisis et définis de manière relativement conventionnelle ; ils serviront à «représenter» les images en leur associant une signification utile.

Les exemples fournis manifestent quelques unes des difficultés rencontrées dans la constitution des définitions : les termes des langages documentaires sont en effet présentés dans un ordre conceptuel qui les relie, accompagnés parfois de «règles d'usage» qui en précisent le sens. La lexicographie, tout comme la traduction, sont en effet des savoir-faire.

Aucune des opérations qui précèdent ne sont effectuées à l'aide de machines. Leur usage est en revanche largement répandu pour le stockage des représentations, la recherche des documents pertinents et la production de «réponses» sous forme de listes de références.

Lorsqu'un certain volume de données est atteint l'utilisation d'un ordinateur se révèle économique. Les programmes ne sont pas spécialisés dans le traitement de l'image car les données en mémoire aussi bien que les produits de sortie sont des représentations à forme alphanumérique. Ils sont donc applicables, et appliqués, à l'exploitation documentaire de textes.

En définitive la machine «assiste» le documentaliste sans qu'il apparaisse d'autre substitution des tâches que celle d'édition.

Pour faire mieux et plus il faut connaître les propriétés des opérations empiriquement réalisées et des langages documentaires.

II – Archéologie et «codes analytiques»

Aux vocables précédents (classifications, thesaurus, etc.) il convient d'ajouter celui de «code analytique» dans les domaines de l'archéologie ou de l'histoire de l'art. Deux faits semblent être à l'origine de cette nouvelle dénomination : d'une part, les données sont relativement mieux circonscrites (dans certains cas la collection des objets est en effet pratiquement close) ; d'autre part, un effort particulier a été fait pour déterminer les règles de construction des langages documentaires (Il se peut que l'aspect traditionnel de l'archéologie ait là joué un rôle : un vocabulaire est en principe fixé; son usage ne s'est pas révélé satisfaisant). De toute manière les termes et définitions servent, tout

comme dans le cas de l'urbanisme, à des fins documentaires. D'emblée pourtant le rôle de la machine n'y apparaît qu'au sens général, presque abstrait du terme : dans la plupart des applications citées les fichiers sont constitués et exploités manuellement, même si les principes de fonctionnement en sont de type «mécanique» (fichier à tringle ou à coïncidence visuelle type «peek-a-boo»). Les règles d'emploi sont donc transposables sur ordinateur.

L'insistance mise sur les propriétés des langages documentaires manifeste deux traits presque paradoxaux : les codes élaborés sont, dans leur nature (choix des termes, modalités de leurs définitions, règles de combinaison) étroitement dépendants des données décrites (outils, vases, décors abstraits ou non, etc.) et des objectifs ; mais par ailleurs les principes pour leur élaboration semblent avoir une portée générale qui en étend l'application, de l'archéologie à l'ethnologie (et, pourquoi pas, à l'urbanisme).

On s'efforce là d'expliciter certaines des opérations appliquées de manière pragmatique par les documentalistes. Dès lors que les données sont faites d'objets, ou d'images de ces objets, il apparaît spontanément nécessaire d'en donner une «représentation» significative. De manière moins évidente qu'en d'autres domaines il apparaît aussi que les descriptions en langue naturelle, même spécialisée, n'ont pas la précision ni l'économie satisfaisant aux objectifs documentaires. D'où la nécessité de créer des «systèmes d'expression artificiels» aux propriétés quelque peu ambiguës.

La définition des termes s'y veut plus rigoureuse que ceux qu'utilisent les spécialistes : le «rangement» efficace l'exige en effet. De nouvelles conventions sont établies. Elles déterminent la finesse des descriptions, la validité des combinaisons, l'étendue des confusions (classes d'objets) admises. L'usage interdit pourtant d'aller trop loin dans cette voie. D'abord parce que les regroupements documentaires ont un but pratique où le «sens» (datation, aire culturelle, géographie, etc.) doit apparaître au spécialiste.

Ensuite et surtout peut être parce que la discipline a encore des fondements intuitifs. On remarquera par exemple que la mesure, donc la quantification, garante de la précision, reste du domaine de la recherche.

III – Géographie : établissement de cartes

Les mesures sont déjà bien moins étrangères aux préoccupations des géographes. Faits physiques ou faits humains, la matière et le sens qu'on peut lui accorder, en sont prédéterminés : reliefs, réseaux hydrologiques, caractéristiques culturelles sont autant de données sélectionnées que les cartes ont pour but de représenter. Les données ont des propriétés physiques mesurables qui les distinguent. Des capteurs placés dans des avions ou des satellites en enregistrent les caractéristiques. D'emblée on saisit l'importance des matériels dont les conditions d'emploi ne sauraient être parfaites. D'où la nécessité de corriger les fonctionnements perturbateurs (déplacement des capteurs, angles de prise de vue etc.). Celle aussi de tenir compte des évolutions chronologiques (modification climatologiques ou saisonnières).

Nombre de ces conditions apparaissent comme connues, dans la mesure au moins où elles sont liées au fonctionnement des machines. Les données recueillies doivent cependant être encore épurées («lissages» et corrections diverses) pour apparaître sous une forme fidèle et homogène.

Il s'en faut, semble-t-il, que la carte soit dressée : il suffit, pour s'en assurer tout intuitivement, de comparer le plan d'une ville à la photo aérienne du même lieu. Là encore le classement prédomine. Il assure seul que le produit final puisse être appréhendé de manière significative.



Et c'est ce rapport déjà significatif entre instruments de mesure et données ultimes qui autorise le géographe, avec des succès divers, à utiliser tout l'arsenal des méthodes de classification automatique. Aux deux bouts de la « chaîne » les faits sont connus : il s'agit du « terrain », il s'agit de carte. Les difficultés de réalisation sont d'ordre plus pratique que théorique ; elles sont d'hommes, de temps et d'argent. Machines et méthodes y suppléent.

4 – Physique nucléaire : des millions de clichés

Aucune trace de « fait humain » dans les images scrutées par les physiciens.

Là où le géographe sélectionnait des « données naturelles » (hydrologiques, minières etc.) et de civilisation (peuplement, agriculture) le physicien reconnaît des trajets de particules. Les images, clichés photographiques et films, sont le produit de machines conçues et réalisées aux fins propres de la discipline (chambre à bulle notamment). Ce n'était certes pas le cas de l'urbanisme ni de l'archéologie ; cette spécificité commençait d'apparaître avec les télécapteurs aéroportés des cartographes. Un lien essentiel, « symbiotique » pourrait-on dire, entre machines et processus intellectuels caractérise en fait tout le traitement de l'image en physique nucléaire.

Il interdit d'abord toute sélection : la nature et le nombre des images sont déterminés par les appareils. Le volume des clichés a d'ailleurs un sens précis et calculé ; en deçà de certains seuils les phénomènes représentés ne peuvent être interprétés. Il est donc nécessaire de scruter jusqu'à plusieurs millions de clichés. En tout autre domaine cette difficulté serait pratiquement insurmontable. Les physiciens traitent heureusement d'images particulières (photos représentant les trajectoires de particules) qui se prêtent, dans une certaine mesure, au calcul.

Il est possible d'effectuer des tris préalables par des moyens automatiques ou semi-automatiques qui augmentent les chances d'une observation significative. L'opération exige de transformer, et plus exactement de transposer, les clichés photographiques sous une forme qui permette les calculs (digitalisation). Une telle transposition de nature « analogique », ne dénature pas l'information dans le cas de la physique. Cette première sélection, à fondements statistiques, laisse subsister une masse considérable d'images potentiellement intéressantes.

Le lecteur trouvera dans le présent ouvrage quelques informations sur l'histoire des recherches qui permettent aujourd'hui de réaliser le classement ultime. On a visé quelque temps une automatisation intégrale du processus. Puis une association entre spécialiste et machine travaillant en séquence pour aboutir finalement à une opération simultanée. La « symbiose » est provisoirement atteinte, en précisant toutefois l'extrême spécialisation des « acteurs », humains ou mécaniques.

La chaîne complexe du traitement intègre les spécialistes, les machines et les méthodes : c'est là une condition indispensable à la gestion d'un volume considérable de données « en temps réel », sur le moment en quelque sorte, sans délais économiquement ni techniquement impraticables.

Pour conclure cette introduction

Les exemples cités ci-dessus manifestent une espèce de progression.

A première vue ni les matériaux, ni les équipements, les méthodes ou les objectifs ne peuvent se comparer. Il existe pourtant un point commun : la quasi-totalité des recherches et réalisations réunies dans cet ouvrage visent à classer correctement des images (quelles qu'en soient les formes pourvu qu'elles ne relèvent pas de la langue).

Dès lors les points de comparaison peuvent être multipliés en fonction des intérêts propres au lecteur. On se limitera ici à quelques considérations d'ordre général. La progression déjà mentionnée porte sur la **dépendance croissante** des données initiales à l'égard de la discipline. Il est loisible aux documentalistes de réunir un corpus d'informations par des procédés empiriques. C'est que rien ne définit avec précision l'usage qui sera fait des données stockées puis sélectionnées. L'utilité d'un centre de documentation, pas plus que celle d'une bibliothèque n'est ici en cause bien au contraire : ce n'est pas la diminuer que de reconnaître la difficulté à en mesurer la « productivité ». Ce rapport incertain explique peut-être le caractère empirique du traitement, les délais entre la réunion des informations et leur éventuelle distribution, l'intrusion quelque peu aléatoire des machines.

Les contraintes s'affirment déjà beaucoup plus nettement lorsqu'à partir de données initiales on vise à donner une représentation **directement utilisable** pour la recherche. Tel est le cas d'une carte géographique. Enfin la spécialisation presque absolue et l'intégration de tous les éléments semblent nécessaire quand on cherche des **résultats**.

On observe un fait presque paradoxal : certaines applications ou disciplines ne sont pas en mesure de définir avec précision les propriétés des phénomènes (ou des images) qui leur sont les plus utiles ; d'autres l'ont si bien fait qu'elles asservissent ou construisent les machines et méthodes. On ignore pourtant les critères qui permettraient de décider de « l'utilité » relative (sociale, politique peut être) de la physique ou de l'archéologie...

Il appartient au lecteur de décider quelles transpositions sont concevables.

Francis LEVY