

Equipe ADAGIo
Algorithmique Discrète et ses Applications à la Génomique et à l'Imagerie
Applying Discrete Algorithms to Genomics and Imagery

version du 17 novembre 2005

Table des matières

1	Composition de l'équipe	3
2	Présentation générale	3
2.1	Domaine de recherche	3
2.2	Bioinformatique comme domaine d'application	4
2.3	Imagerie comme domaine d'application	4
2.4	Approche méthodologique	4
3	Fondements scientifiques: résultats et perspectives	5
3.1	Géométrie discrète	5
3.1.1	Reconnaissance de structures discrètes	7
3.1.2	Extraction de paramètres géométriques	8
3.1.3	Reconstruction	9
3.1.4	Génération de structures discrètes à partir de contraintes	10
3.2	Algorithmique des séquences	12
4	Logiciels prototypes	13
4.1	SIGffRid	13
4.2	CurvaDis	14
5	Applications bioinformatiques	14
5.1	Problème de phase en cristallographie	15
5.2	Analyse de signaux de régulation dans les séquences nucléotidiques	16
5.3	Courbure de l'ADN	17
6	Applications en imagerie	18
6.1	Approximation polygonale multiordre de courbes discrètes	18
6.2	Détection de cercles et arcs de cercles dans des images techniques	19
6.3	Détection semi-automatique d'ellipses dans des images numériques	19
6.4	Déformation de surfaces	20
6.5	Reconstruction à base de sketching	20

7 Projets, collaborations et contacts	21
7.1 Collaborations internationales	21
7.2 Collaborations nationales	22
7.3 Collaborations et positionnement au sein du LORIA	22
7.4 Actions régionales	23
8 Participation à la vie scientifique	23
8.1 Comités de programme et évaluation d'articles	23
8.2 Jurys de thèse	23
8.3 Divers	24
9 Publications des membres de l'équipe depuis 1995	24
10 Références externes	28

1 Composition de l'équipe

Permanents

- Isabelle Debled-Rennesson (MC IUFM de Lorraine, responsable scientifique)
- Eric Domenjoud (CR CNRS, responsable permanent)
- Philippe Even (Pr UHP, IUT St Dié)
- Bertrand Kerautret (MC UHP, IUT St Dié)
- Jean-Luc Rémy (CR CNRS, 70%)

Assistante

- Céline Simon (TR INRIA)

Non-Permanents

- Fabrice Touzain (Doctorant UHP, depuis le 1/11/2003, co-financement région/Inria, en collaboration avec le Laboratoire de Génétique et Microbiologie)
- Laurent Provot (Doctorant UHP, depuis le 1/10/2005, financement ministère)

2 Présentation générale

2.1 Domaine de recherche

Si l'on voulait définir la thématique de cette équipe par approximations successives, il serait naturel de commencer par la placer dans le domaine de l'*algorithmique discrète*. Notre intérêt pour les modèles discrets est une première caractéristique de notre champ de recherches. Construire un modèle discret d'un problème ou d'un phénomène du monde réel fait appel, sur le plan mathématique, aux *structures discrètes*, telles que graphes, mots, arbres, ensembles de points dans un espace, etc. Ces structures servent à représenter les données du problème (auquel cas elles jouent le rôle de *structures de données*), ou les relations entre les objets, ou encore le fonctionnement du système modélisé.

Pour pouvoir utiliser les modèles discrets, nous sommes donc amenés à étudier les propriétés des structures impliquées. En tant qu'informaticiens, nous nous intéressons tout particulièrement aux *propriétés algorithmiques*, à savoir comment vérifier algorithmiquement si notre objet mathématique possède une caractéristique donnée ou comment calculer des fonctions définies dans notre modèle. Nous nous intéressons à l'*efficacité (complexité)* des calculs impliqués, c'est-à-dire au temps et aux autres ressources utilisés par l'algorithme pour aboutir à un résultat.

Pour pouvoir développer des algorithmes efficaces sur les structures discrètes, ainsi que pour analyser et optimiser ces algorithmes, il faut donc comprendre et maîtriser les propriétés des structures sous-jacentes. Ces propriétés peuvent être de natures différentes : s'il s'agit de propriétés de nature exacte, on parle de *propriétés combinatoires, géométriques ou arithmétiques* selon les cas ; si le modèle en question est défini en termes plus probabilistes, on aura plutôt des *propriétés statistiques*. L'étude de ces diverses propriétés est donc l'objectif principal de cette équipe.

2.2 Bioinformatique comme domaine d'application

Les modèles discrets sont omniprésents. Parmi tous les domaines où ils apparaissent, la génomique, c'est-à-dire l'étude de macromolécules biologiques (ADN, ARN, protéines), joue pour nous un rôle tout à fait particulier. Elle nous fournit d'une part une source d'étude et d'autre part un domaine privilégié pour tester et appliquer nos idées et méthodes. L'irruption de modèles discrets dans ce domaine de recherche est due à la découverte de la structure des molécules constituée d'un enchaînement linéaire d'éléments constitutants qui appartiennent à un petit nombre de types. Cette découverte a débouché sur l'adoption d'un modèle discret : dans leur forme linéaire, ces molécules sont représentées par des chaînes de lettres tirées d'un alphabet de petite taille, nommées *séquences*. Si ce modèle linéaire capture déjà une partie significative et importante des propriétés des molécules biologiques, il va sans dire qu'elles ne s'y limitent pas ; inversement ce n'est qu'une première illustration de l'apparition de modèles discrets dans l'étude des molécules biologiques. Dans ces études, on s'intéresse également à la structure secondaire de ces molécules (leur repliement sur elles-mêmes) et à leur structure tertiaire (leur repliement dans l'espace). On s'intéresse de plus à la position de certains gènes dans ces molécules, à leur identification, leur fonctionnement et leur interaction. On s'intéresse aussi à la classification de ces molécules selon leur rôle fonctionnel et leur structure tridimensionnelle. Toutes ces questions, ainsi que bien d'autres problèmes clefs de la bioinformatique, font appel aux modèles discrets pour leur étude.

Dans le cadre de cette équipe, nous travaillons d'une part sur la recherche de propriétés biologiques, décrites à l'aide de motifs, reflétées au niveau des séquences et d'autre part sur la structure tridimensionnelle des macromolécules ainsi que l'étude de leurs propriétés géométriques.

2.3 Imagerie comme domaine d'application

Dans le domaine de l'imagerie, les systèmes d'acquisition de données comme les appareils photo numériques, les scanners, ou IRM médicaux fournissent des données discrètes. Ils échantillonnent le monde réel sur des grilles régulières en 2 ou 3 dimensions. Ces données doivent ensuite être traitées et interprétées. Des problèmes de détection, d'analyse ou encore de reconnaissance doivent être résolus sur ce type de données. Deux approches sont alors possibles. La première consiste à plonger les données discrètes dans un espace continu, procéder ensuite avec des méthodes d'approximation ou d'interpolation puis utiliser des méthodes paramétriques usuelles. Dans la seconde approche, les notions et objets discrets sous-jacents au problème posé sont définis et les propriétés issues de ces objets permettent la résolution du problème tout en conservant la réversibilité entre le monde discret et le monde analogique.

L'algorithmique discrète intervient naturellement dans les traitements utilisés dans la seconde approche en conservant la nature discrète des données.

Dans le cadre de cette équipe, nous nous intéressons aux propriétés géométriques et topologiques d'ensembles de points discrets en dimension 2 ou 3. Plus particulièrement, nous recherchons des caractéristiques sur les objets discrets analysés comme la courbure, la normale, le périmètre, la surface, l'orientation, le nombre de concavités, ... Nous utilisons en particulier ces caractéristiques pour résoudre des problèmes issus du traitement d'image ou de la modélisation géométrique.

2.4 Approche méthodologique

Notre objectif est de couvrir au maximum toute la chaîne, de la modélisation du problème et l'étude des propriétés mathématiques du modèle, jusqu'à l'implantation d'algorithmes sous la

forme de logiciels expérimentaux et le test de ces logiciels sur des données réelles. Cette démarche scientifique signifie que nous cherchons tout d'abord à développer de nouveaux modèles, de nouvelles méthodes et de nouveaux concepts informatiques pour résoudre nos problèmes. D'autre part, nous ne nous contentons pas de cette étape, mais nous souhaitons mener cette démarche jusqu'à la réalisation logicielle expérimentale de nos développements théoriques, afin de vérifier leur viabilité. L'étape d'expérimentation peut ensuite conduire à ajuster, affiner ou améliorer le modèle ou la méthode sous-jacents, jusqu'à ce que les résultats s'avèrent satisfaisants, auquel cas le logiciel est diffusé à l'extérieur.

3 Fondements scientifiques: résultats et perspectives

3.1 Géométrie discrète

Si les mots sont des structures discrètes générales, ici nous nous intéressons à des objets discrets ayant une interprétation géométrique (planaire ou spatiale) et étudiés dans le cadre de la *géométrie discrète*. Cette branche d'études, apparue dans les années 70 et initiée par des chercheurs connus comme par exemple Azriel Rosenfeld, Gabor T. Herman et Efim Khalimsky, a pour objectif de définir un cadre théorique pour transposer dans $\mathbb{Z}^n$ les bases de la géométrie euclidienne, les notions discrètes définies étant le plus proche possible des notions continues que nous connaissons (telles que distance, longueur, convexité, ...). Plusieurs façons d'aborder cette étude ont été développées [13], comme :

- le point de vue topologique s'intéressant par exemple à l'équivalent discret du théorème de Jordan (toute courbe fermée simple sépare le plan en deux domaines : l'intérieur et l'extérieur de la courbe),
- le point de vue arithmétique, introduit par Jean-Pierre Reveillès en 1989 [52], qui donne une définition en compréhension des droites et plans discrets.

C'est cette dernière approche que nous utilisons. Une droite discrète du plan est ainsi l'ensemble des points de coordonnées (x,y) de $\mathbb{Z}^2$ vérifiant une double inégalité de la forme $\mu \leq ax + by < \mu + \omega$, avec a,b,μ,ω entiers. Les propriétés des droites discrètes définies de la sorte sont en relation étroite avec les propriétés des nombres entiers et nous rapprochent ainsi de la combinatoire des mots (utilisation des mots de Sturm par exemple [8, 9]). Les plans discrets sont définis de manière analogue. Ces définitions analytiques permettent de représenter de manière compacte des objets discrets, d'étudier des objets intrinsèquement discrets (pas uniquement des approximations d'objets continus), et de définir des objets discrets infinis.

Notre approche vise à résoudre *proprement* des problèmes pratiques posés par la non-validité de la plupart des théorèmes euclidiens dans un espace discret (par exemple l'intersection de deux droites discrètes). Elle permet également d'éviter les erreurs engendrées par l'utilisation de nombres réels codés en virgule flottante ; les conséquences pour les calculs géométriques et graphiques sont très importantes en particulier lorsqu'une décision de nature topologique dépend du résultat numérique (par exemple l'appartenance d'un voxel à l'intérieur d'un polyèdre). Notons que cette approche se différencie de la *topologie discrète*, qui considère l'objet discret sous un angle topologique plutôt que par ses propriétés combinatoires.

De nombreux résultats fondés sur cette approche ont vu le jour ces dix dernières années [3, 32, 24, 26, 27, 15, 59] :

- définition et étude de nouvelles classes d'objets discrets (droites 3D, hyperplans, cercles,

sphères, simplexes, ...),

- reconnaissance analytique permettant, non seulement de dire si une suite de points est ou non un objet discret donné (segment de droite, morceau de plan, cercle, ...), mais aussi de fournir les coefficients des inéquations analytiques correspondantes,
- reconstruction analytique visant à passer d'une représentation en compréhension du discret à une représentation en compréhension du continu,
- transformations discrètes: applications quasi-affines, rotations, filtres,
- visualisation, utilisant des propriétés des objets discrets telles que l'épaisseur optimale des objets.

En ce qui nous concerne, nos travaux se rangent principalement parmi les trois premières thématiques de cette liste.

L'application première de ces algorithmes se trouve en informatique géométrique et graphique, en particulier en synthèse d'image, traitement d'image, analyse de documents et imagerie médicale. Citons par exemple J.-M. Chassery et A. Montanvert qui ont publié un ouvrage [13] sur la géométrie discrète en analyse d'image regroupant les algorithmes connus du domaine, d'autres ouvrages en langue anglaise sont parus depuis relatant les progrès du domaine [38, 30, 35]. Cependant, la géométrie discrète a aussi trouvé des applications dans des domaines où la géométrie intervient moins directement, comme la cristallographie. Cette géométrie à l'usage des ordinateurs est aujourd'hui en plein essor comme en témoigne le colloque international DGCI (Discrete Geometry for Computer Imagery) dont la première édition a eu lieu en 1991 ainsi que le colloque international IWCIA (Workshop on Combinatorial Image Analysis) dont la onzième édition aura lieu en 2006. De plus, un comité technique IAPR (TC18) sur la géométrie discrète a été mis en place avec pour objectif de promouvoir les interactions et les collaborations entre les chercheurs du domaine; une page web (<http://www.cb.uu.se/~tc18/>) regroupe les informations sur les conférences liées à la géométrie discrète et les offres de stage, et aussi met à disposition les codes des algorithmes de base.

Plusieurs questions clés ressortent des dernières conférences du domaine (IWCIA 04 et DGCI 05). L'étude des propriétés des objets discrets comme les droites, cercles, plans, courbes et surfaces discrets reste toujours un thème d'actualité. Dans le thème de la reconstruction, la polyédrisation, c'est à dire la recherche d'une approximation ou d'un codage polyédrique d'un objet 3D discret est un sujet pour lequel de nombreux travaux restent à réaliser.

Ces sujets sont étudiés dans notre équipe et plus particulièrement un des thèmes développés est l'étude des courbes et surfaces discrètes bruitées. Notre objectif est d'élaborer, dans le cadre de la géométrie discrète, un formalisme adapté à ces objets tenant compte du bruit inhérent aux outils et méthodes d'acquisition. Pour étudier les courbes et les surfaces discrètes bruitées, nous *définissons analytiquement* des objets discrets, par exemple en généralisant les notions de segments de droites discrètes et de morceaux de plans discrets. L'étude des propriétés arithmétiques, géométriques et combinatoires de ces objets permet d'obtenir des algorithmes performants (*reconnaissance*, *parcours*, ...) et d'extraire des *paramètres géométriques* (périmètre, courbure, normale, aire, ...) sur les courbes et surfaces bruitées à partir desquels nous pouvons par exemple proposer des algorithmes de *reconstruction* d'objets continus correspondant aux structures discrètes étudiées.

D'autre part, nous souhaitons enrichir la panoplie d'objets discrets non seulement en proposant des définitions analytiques de nouveaux objets mais aussi en développant des techniques de *génération d'objets discrets à partir de leurs propriétés*.

Nous décrivons ci-après les résultats et les perspectives de recherche sur les thèmes de recherche cités ci-dessus.

3.1.1 Reconnaissance de structures discrètes

L'idée générale des travaux présentés dans ce paragraphe consiste à travailler à partir de données discrètes puis, grâce à des définitions analytiques d'objets discrets, consiste à identifier si les données initiales correspondent à des objets discrets connus, c'est ce que nous appelons la *reconnaissance*.

Droites discrètes 2D et 3D Après avoir travaillé sur la reconnaissance exacte de droites discrètes [DRR95], nous nous intéressons à la notion d'objets discrets *floous*, correspondant à des objets discrets bruités, et à leur détection.

Nous avons défini la notion de *segment flou 2D*, en relation avec la définition arithmétique des droites discrètes [52] où l'épaisseur est paramétrable. Un segment flou est une suite 8-connexe de points qui appartiennent à une droite discrète dont l'épaisseur est donnée. Un paramètre, l'ordre du segment flou, permet de contrôler l'amplitude du bruit autorisé en fixant l'épaisseur de la droite englobant le segment flou. Nous avons élaboré un algorithme incrémental et très efficace de reconnaissance de ces structures discrètes à partir duquel un algorithme de découpage de courbes discrètes en segments floous d'ordre fixé a été déduit [DRRRD03, DRRRD05].

Depuis, en collaboration avec Fabien Feschet (LLAIC, Clermont Ferrand), nous avons optimisé cette notion et l'algorithme de segmentation associé [DRFRD05].

Cette approche peut être utilisée dans le milieu de l'analyse de documents et une collaboration avec l'équipe QGAR du LORIA a permis d'appliquer cet algorithme dans ce cadre (cf. 6.1).

Récemment, nous avons étendu la définition de segments floous à la dimension 3 et nous souhaitons développer un algorithme similaire de reconnaissance de segment flou 3D à partir duquel un algorithme de segmentation de courbes discrètes 3D sera déduit.

Plans discrets Un plan discret naïf est défini comme le sous-ensemble de points (x,y,z) de $\mathbb{Z}^3$ vérifiant la double inégalité $h \leq ax + by + cz < h + \max(|a|, |b|, |c|)$, avec $(a,b,c,h) \in \mathbb{Z}^4$. Etant donné un sous-ensemble fini de $\mathbb{Z}^4$, le problème est de déterminer s'il existe ou non un plan discret naïf le contenant. Différentes approches ont été utilisées pour résoudre ce problème [DRR96, 25, 64, 12] et, avec Yan Gérard (LLAIC, Clermont Ferrand) et Paul Zimmermann (équipe SPACES, LORIA), nous avons proposé un nouvel algorithme [GDRZ05] qui utilise une stratégie d'optimisation dans un ensemble de facettes triangulaires ainsi que la notion d'ensemble de cordes introduite dans [27]. Le code de cet algorithme est accessible sur le web (<http://www.loria.fr/~debled/plane>).

Nous nous intéressons maintenant à la reconnaissance de plans discrets bruités et, cette année, un stage de DEA a été effectué sur ce sujet [Pro05] en combinant des techniques de géométrie algorithmique et de géométrie discrète. En effet, des algorithmes de construction d'enveloppes convexes ou des algorithmes géométriques de programmation linéaire apparaissent naturellement comme des solutions possibles à ce problème. Cependant, la spécificité des structures discrètes et les propriétés arithmétiques sous-jacentes impliquent des adaptations et des simplifications de ces algorithmes pour résoudre le problème de la reconnaissance de morceaux de plans bruités.

D'autres algorithmes de reconnaissance existent et nous souhaitons aussi comparer leurs différentes approches en terme de fiabilité, efficacité et rapidité. Une collaboration avec Reinhard Klette, professeur à l'Université d'Auckland, est amorcée sur ce sujet.

3.1.2 Extraction de paramètres géométriques

Nous présentons dans les 2 premiers paragraphes de cette partie la démarche consistant à caractériser les données discrètes par des objets discrets représentés par des définitions analytiques, puis grâce aux propriétés de ceux-ci, à extraire des *paramètres géométriques* comme la longueur, la courbure, la normale, la tangente, l'aire, ... Dans le dernier paragraphe, une autre démarche est adoptée qui consiste à estimer la surface discrète par une surface euclidienne puis en extraire des paramètres géométriques.

Mesure de la longueur d'une courbe discrète 3D

En collaboration avec David Coeurjolly (LIRIS, Lyon) et Olivier Teytaud (LRI, Paris XI), nous avons proposé un nouvel algorithme de calcul de la longueur de courbes discrètes 3D [CDRT01]. Il repose sur une définition arithmétique de droites discrètes 3D et sur l'algorithme linéaire de segmentation de courbes 3D donné dans [DR95]. Le principe de l'algorithme est le suivant : la courbe 3D est découpée en segments de droites de longueurs maximales et la longueur de la courbe est calculée en fonction de la longueur de la ligne polygonale obtenue. La convergence de cette technique d'estimation de longueur a été prouvée. Ce travail démontre la puissance des techniques de segmentation, qui pourront être utilisées pour d'autres problèmes de géométrie discrète. Nous souhaitons adapter ce travail au calcul de la longueur de courbes discrètes bruitées.

Courbure discrète Dans le cadre de notre travail sur les courbes bruitées, nous avons proposé une nouvelle notion de tangente discrète, reposant sur la définition des segments flous (cf. 3.1.1) et adaptée aux courbes discrètes bruitées ce qui n'était pas le cas des définitions antérieures [66, 63]. Un algorithme permet de calculer les paramètres de la tangente en chaque point d'une courbe discrète. A partir de celui-ci, nous pouvons calculer plusieurs paramètres comme le vecteur normal ou la courbure en tout point de la courbe considérée [DR04].

Ce travail a été amélioré en dimension 2 puis étendu à la dimension 3 avec la notion de segment flou 3D dans le cadre de stages de DEA [Rap04, Ngu05].

Nous souhaitons faire une étude théorique des performances de cet estimateur de courbure discrète. Des collaborations autour de ce travail sont engagées avec F. Feschet (LLAIC, Clermont-Ferrand) et M. Tajine (LSIIT, Strasbourg).

En bioinformatique, où nous nous intéressons au calcul de la courbure de l'ADN (cf. 5.3), les modèles discrets de représentation de la structure 3D de l'ADN nécessitent le développement d'algorithmes spécifiques de géométrie discrète pour répondre au problème posé. Nous avons adapté à ce problème les algorithmes présentés précédemment. Un logiciel est en cours de réalisation (cf. 4.2)

D'autre part, en traitement d'image, l'algorithme 2D est utilisé dans le cadre d'une application à la reconnaissance de cercles et d'arcs de cercles dans des documents techniques (cf. 6.2).

Extraction de paramètres par lissage de surfaces discrètes Nous avons introduit une méthode réversible de lissage de surfaces discrètes. Cette méthode se base sur l'estimation statistique du plan discret pour pouvoir transformer la surface discrète en une surface appelée euclidienne. La nouvelle représentation de la surface permet d'extraire directement les paramètres géométriques tels que l'aire, le volume ou les normales de la surface. Ces normales sont utilisées par exemple pour la visualisation. Il est aussi possible d'obtenir des profils 2D de l'objet discret pour en extraire d'autres mesures associées telles que par exemple le périmètre des différentes coupes.

Cette méthode se limite pour l'instant à l'utilisation d'un voisinage symétrique pour l'estimation du plan discret. Cette stratégie peut être améliorée en particulier lorsque l'on se trouve sur le bord d'une région présentant une discontinuité. Dans ce cas, on peut envisager d'utiliser un voisinage non symétrique de façon à prendre en compte plus d'information dans une direction privilégiée. L'estimation du plan discret sur le bord de la région serait alors plus précise, et permettrait d'obtenir une meilleure estimation des paramètres géométriques.

Une autre piste de recherche à travers cette méthode porte sur l'intégration d'une représentation inter-voxel de la surface discrète. Cette stratégie permettrait entre autre d'éviter certains problèmes topologiques que l'on peut rencontrer avec une approche basée sur les centres de voxels. L'utilisation de cette nouvelle représentation requiert l'adaptation du principe de reconnaissance statistique de plans discrets et la définition d'une nouvelle transformation réversible de la surface discrète.

3.1.3 Reconstruction

A partir de données discrètes, nous utilisons des objets discrets et leurs propriétés pour reconstruire des objets du monde continu correspondant aux données initiales.

Décomposition polygonale de courbes discrètes en parties convexes et concaves L'étude de la convexité d'une région discrète du plan se ramenant à celle de figures particulières appelées polyominos hv-convexes, nous avons développé un algorithme incrémental et linéaire de détection de la convexité de tels polyominos [DRRRD03]. Des contacts sur ce sujet ont été établis avec le Professeur Eckhardt de l'Université d'Hambourg à l'occasion de l'invitation d'un membre de l'équipe à un workshop [DR02]. Une de ses étudiantes, Hélène Dörksen-Reiter a fait plusieurs séjours au LORIA et nous avons travaillé ensemble sur des problèmes théoriques liés à son sujet de thèse. Nous avons développé un algorithme linéaire de décomposition polygonale du bord d'un objet discret en parties convexes et concaves reposant sur certaines propriétés établies dans [DRRRD03]. Une telle décomposition est très utile pour décrire un objet discret [DRDR05].

Nous poursuivons ce travail en étudiant les propriétés des algorithmes de décomposition polygonale du bord d'objets discrets.

Polyédrisation A partir des algorithmes de reconnaissance de plans bruités présentés en 3.1.1, nous étudions la polyédrisation d'objets discrets bruités tridimensionnels en contrôlant l'approximation effectuée. Notre objectif est d'obtenir des algorithmes performants en combinant les approches de géométrie discrète et de géométrie algorithmique. Ce travail fait l'objet d'une thèse, il permettra de représenter les surfaces discrètes bruitées avec un nombre minimal d'approximations de facettes planes discrètes en utilisant par exemple les stratégies proposées dans [59]. Cette représentation paramétrique pourra être utilisée pour extraire rapidement des propriétés géométriques des surfaces discrètes ou encore pour lisser *sous contrôle* la surface des objets discrets bruités et ensuite la transformer en surface euclidienne permettant ainsi l'utilisation de calculs rapides de la géométrie euclidienne sur celle-ci (transformations géométriques, tests de visibilité, d'intersection, ...).

Reconstruction de surfaces à partir d'images Dans des travaux antérieurs [BK05], nous avons proposé une méthode discrète de reconstruction de surfaces de type lambertien. Elle est basée sur l'estimation discrète de la normale sur le bord d'une région de la surface. La reconstruction s'effectue à partir de contraintes géométriques et de l'information photométrique contenues dans

une ou plusieurs images. L'approche adoptée permet d'exprimer des contraintes explicites pour lever l'ambiguïté concave/convexe liée à l'utilisation d'une seule source lumineuse frontale.

Cette méthode a montré une grande robustesse par rapport au bruit des images et par rapport au modèle de réflectance pouvant prendre en compte une composante spéculaire [Ker04]. Nous travaillons actuellement sur l'intégration d'une source lumineuse ponctuelle non distante dans le modèle. Cette extension vise une meilleure reconstruction, en particulier dans le cas d'objets de grande taille pour lesquels il est expérimentalement très difficile de satisfaire l'hypothèse d'une source lumineuse directionnelle sans matériel spécifique. En effet les méthodes développées jusqu'à maintenant montrent une grande complexité de calcul. Par exemple celle de Kozera [40] n'a été validée que sur des images synthétiques et nécessite la parallélisation de l'algorithme. Par ailleurs la résolution des images utilisées pour la reconstruction reste très faible, de l'ordre de 64x64 pixels.

Une autre amélioration envisagée est de travailler sur les images discrètes de départ pour obtenir une reconstruction à résolution adaptative. L'idée principale est d'utiliser les informations photométriques d'une ou plusieurs images et d'obtenir des régions homogènes pour simplifier à la fois le processus de reconstruction et la surface résultante. Cette approche rejoindrait en partie la solution proposée par Kovalevsky [39] basée sur la segmentation de l'image source. Cependant certaines hypothèses émises par l'auteur ne se vérifient pas systématiquement pour tous types de surfaces. Aussi la stratégie que nous pensons adopter s'appuie sur l'utilisation de plusieurs sources lumineuses pour définir des critères de segmentation valides et adaptés à la forme à reconstruire.

3.1.4 Génération de structures discrètes à partir de contraintes

Jusqu'à présent, nous nous sommes principalement intéressés à une définition analytique des objets discrets par des équations ou inéquations sur les coordonnées des points qui composent l'objet. Si une telle définition fournit un moyen rapide de tester l'appartenance d'un point à l'objet discret considéré, en revanche elle ne permet en général pas directement d'engendrer efficacement les points de l'objet. Notre objectif est de développer des techniques de génération d'objets discrets à partir de leurs propriétés.

Extension du formalisme En plus de ne pas fournir de méthode de génération des objets, le formalisme utilisé s'avère parfois insuffisant pour exprimer certaines propriétés. On ne peut par exemple pas définir les pentominos avec une équation analytique ou encore définir la connexité d'un objet. Nous proposons une autre démarche permettant de considérer des problèmes où les objets ne s'adaptent pas au formalisme proposé précédemment : une information plus globale est donnée sur l'objet avec pour objectif de reconstruire géométriquement un ou plusieurs objets solutions. Cette démarche ne remplace pas la précédente mais l'enrichit de manière à pouvoir considérer un éventail plus large de propriétés des objets dont nous traitons.

En particulier, nous nous intéressons au formalisme qui consiste à définir un objet par sa fonction caractéristique, c'est à dire une fonction qui en chaque point vaut 1 ou 0 selon que le point appartient ou n'appartient pas à l'objet. Ce type de définition permet d'exprimer d'autres propriétés d'un objet à travers des propriétés de sa fonction caractéristique. Par exemple, pour un pentomino, la somme de la fonction caractéristique vaut cinq. Ces propriétés peuvent être plus complexes et s'exprimer sur des transformées de la fonction caractéristique telles que la transformée de Fourier.

Génération d'objets discrets En termes de problématiques, nous nous sommes jusqu'à présent intéressés à la recherche d'une définition en compréhension adaptée à la manipulation des objets

discrets ou au raisonnement sur ceux-ci. A l'inverse, dans certaines situations, on dispose d'une telle définition ou plus généralement de certaines propriétés de l'objet, et on s'intéresse à sa génération pour le visualiser ou travailler sur sa représentation géométrique. Dans ce cas, on souhaite obtenir des algorithmes efficaces en particulier dans des domaines critiques comme la visualisation en temps réel où encore dans les cas où un grand nombre de solutions doivent être engendrées.

Bien que l'objectif soit toujours d'aboutir à des algorithmes efficaces, les techniques de résolution peuvent être très variées. Ainsi, dans le cas de la droite discrète définie par une double inégalité sur les coordonnées, des adaptations de l'algorithme de Bresenham [11, 52] fournissent des méthodes très performantes permettant une génération directe des points de la droite. A l'inverse, dans le cadre de l'application au problème de cristallographie moléculaire décrit en 5.1, l'objet discret est décrit par la solution d'un problème de minimisation de la valeur d'une fonction, dite fonction objectif, mettant en œuvre la transformée de Fourier de la fonction caractéristique. On fait alors appel à des techniques d'optimisation par recherche locale afin de converger progressivement vers l'objet recherché. A mi-chemin de ces extrêmes, on trouve par exemple des structures telles que les courbes algébriques discrètes décrites par des équations algébriques pour lesquelles une solution ne peut pas toujours être obtenue facilement ou encore des polyèdres définis par un ensemble d'inégalités. Dans ce dernier cas, la génération passe par une recherche des points extrêmes par des techniques d'optimisation comme la méthode du simplexe, suivie d'une génération de morceaux de plans discrets pour les facettes.

Nous envisageons de développer des techniques hybrides de génération d'objets discrets en combinant selon la nature et la complexité des problèmes posés des techniques issues de l'arithmétique, de l'algèbre et de l'optimisation discrète.

Les deux domaines d'applications dans lesquels nous souhaitons concentrer nos efforts dans un premier temps sont d'une part le tracé de courbes algébriques discrètes et d'autre part l'application à la détermination de structures de macromolécules organiques. Bien qu'en apparence assez éloignés, ces deux domaines se rejoignent dans la nécessité d'obtenir des algorithmes de génération efficaces à partir des contraintes décrivant l'objet à engendrer.

Dans le cas des courbes algébriques, l'efficacité des algorithmes est cruciale pour leur exploitation par exemple dans des applications de visualisation en temps réel, en particulier lors des déformations des objets affichés. Pour cette application, un premier travail de formalisation est nécessaire car l'approche adoptée dans le cas de la droite discrète ne s'adapte pas immédiatement. En effet, la droite discrète est définie au moyen de deux droites réelles parallèles par la double inéquation $0 \leq ax + by - c \leq \omega$. La transposition naïve au cas plus général des courbes algébriques en définissant une courbe algébrique discrète par $0 \leq f(x, y) \leq \omega$ n'est pas possible. On ne peut par exemple pas définir une parabole discrète par $0 \leq y - x^2 \leq \omega$ car alors l'épaisseur n'est pas constante.

Dans le domaine de la détermination de structures de macromolécules, le souci d'efficacité des algorithmes provient d'une part de la plus grande complexité des objets étudiés, et d'autre part du fait que les contraintes sur l'objet à engendrer sont souvent beaucoup moins précises. On a alors potentiellement un grand nombre de solutions sur lesquelles il faudra encore effectuer un travail de post-traitement souvent lourd. Il est donc crucial d'engendrer les solutions le plus efficacement possible, mais également de limiter a priori le nombre de solutions engendrées en incorporant le plus tôt possible dans le processus de génération des contraintes non algébriques visant à éliminer a priori des solutions qui seraient éliminées lors des post-traitements. Une telle contrainte est par exemple la connexité des objets engendrés.

3.2 Algorithmique des séquences

L'*algorithmique des séquences* (ou l'*algorithmique des mots*) est un domaine qui a vu des progrès considérables ces dernières années, comme en témoigne la parution récente de quelques monographies sur ce sujet [18, 28, 19]¹. Le premier ouvrage français couvrant le domaine d'une façon systématique est paru récemment [17]. Tout en étant une partie intégrante de l'algorithmique discrète en général, l'algorithmique des mots forme aujourd'hui un domaine de recherche en tant que tel. Le progrès dans le développement de ce domaine a été largement alimenté par ses nombreux domaines d'application, dont deux – la bioinformatique et la recherche d'information sur l'Internet – sont particulièrement d'actualité aujourd'hui².

En outre, l'algorithmique des mots a développé des structures de données très puissantes, telles que l'arbre des suffixes (*suffix tree*) ou le DAWG (*Directed Acyclic Word Graph*). Le premier objectif de ces structures est de servir d'outils d'*indexation* de textes, c'est-à-dire de fournir une représentation spéciale de textes permettant d'exécuter efficacement des requêtes diverses. De plus, la transformation d'un texte en cette représentation se fait aussi très efficacement, à savoir en temps linéaire. Une fois cette représentation obtenue, de nombreuses tâches peuvent être accomplies très efficacement. Sans essayer de les énumérer, nous renvoyons au livre [28] dont une grande partie est consacrée aux diverses utilisations de l'arbre des suffixes.

Sur ces algorithmes et structures de données de base repose une riche variété d'autres algorithmes et techniques, qui forment le domaine de l'algorithmique des mots.

Jusqu'à présent l'un des domaines de recherche les plus actifs de l'équipe ADAGE, il trouve d'avantage sa place au sein de l'équipe ADAGIo d'un point de vue applicatif. La facette théorique perdurera cependant via les collaborations entretenues avec Grégory Kucherov (LIFL, Lille).

Dans les applications de l'algorithmique des mots, la bioinformatique joue pour nous un rôle particulier (cf. section 5). En effet parmi les problèmes posés par la bioinformatique, la séquence de génomes reste un objet d'étude primordial. Notre intérêt pour les algorithmes textuels est donc en rapport direct avec l'analyse de séquences génomiques.

Nous allons maintenant évoquer deux thèmes de recherche de l'algorithmique des mots dont nous exploitons les résultats. Le troisième thème concerne nos préoccupations actuelles dans ce domaine.

Recherche de motifs approchés Dans de nombreuses applications, il est capital de pouvoir rechercher les motifs d'une façon approchée. Or si l'on autorise la possibilité d'*erreurs* dans les occurrences de motifs, les méthodes pour la recherche exacte ne s'appliquent plus en général. La recherche de motifs approchés a fait l'objet de recherches actives ces dernières années (voir la monographie [45]), mais les algorithmes existants restent insuffisants pour les applications pratiques de grande taille.

Une idée centrale pour rechercher un motif en admettant des erreurs consiste à rechercher, d'une façon exacte, ses fragments, appelés *graines*, qui servent de *témoins* d'occurrences potentielles du motif. En d'autres termes, on élimine des parties du texte qui ne contiennent aucun fragment du motif et par conséquent ne peuvent pas contenir, même de façon approchée, le motif entier. Cette technique, appelée la technique du *filtrage*, s'avère fondamentale dans les applications en

1. La page <http://www.cs.ucr.edu/~stelo/pattern.html> est une bonne collection de liens relatifs au domaine et permet d'en avoir un premier aperçu.

2. Pour observer l'interaction entre l'algorithmique des mots et ces domaines d'application, il suffit de regarder les actes de la conférence *Combinatorial Pattern Matching*.

particulier en bioinformatique. C'est également cette technique qui est employée dans l'alignement des séquences – opération de base dans le traitement des séquences biologiques.

Indexation de données de séquences Rechercher une séquence de caractères dans un corpus textuel exige un temps linéaire en la taille du corpus. Dans le cas de multiples recherches, il est fort avantageux de construire une structure de données, nommée index, capable d'effectuer la recherche de n'importe quel motif en un temps sous-linéaire en la taille du texte. Cela suppose que les recherches sont plus fréquentes que les modifications du corpus (recherche de mots dans un dictionnaire).

La recherche dans les dictionnaires et les bases de données, la compression de fichiers, l'annotation de séquences génomiques et la fouille de données tirent leur efficacité de l'indexation des textes. Ces applications ont nécessité le développement d'index tels que le trie, l'arbre ou la table des suffixes, le DAWG. Ces structures furent prévues pour un texte fixe, un type limité de recherches et une utilisation en mémoire vive. Un objectif toujours d'actualité est d'améliorer leur capacité évolutive, leur versatilité et leur adaptabilité à la mémoire secondaire.

En pratique, l'application choisie va contraindre le type de recherche effectuée et les structures utilisables par la nécessité de mettre à jour la structure ou par la taille du corpus à indexer. Un de nos objectifs est d'étudier l'efficacité de ces structures en pratique, notamment pour les applications en génomique.

Recherche de motifs composites Dans le cadre de la recherche de sites de fixation de facteur de transcription, le programme développé au sein de l'équipe ADAGIo, SIGffRid (voir section 5.2), fait massivement appel à la recherche de couples de mots dans des ensembles de données de grande taille (de l'ordre de 8 Mb). Nous souhaitons optimiser les algorithmes de recherche utilisés par l'emploi ou l'adaptation de certaines des techniques sus-citées. Plus précisément, nous nous intéressons aux problèmes appliqués de recherche combinatoire présentés ci dessous, soient w_1 et w_2 deux mots sur l'alphabet $\mathcal{A} = \{a, c, g, t\}$:

- le dénombrement optimal d'un couple de mots vérifiant $w_1(x,y)w_2$, avec x et y des entiers bornant l'espacement, dans une séquence de grande taille,
- la recherche de couples de mots (à gaps fixes) $w_1(x,y)w_2$ où w_1 et w_2 appartiennent à un dictionnaire $\mathcal{D}$ de très grande taille déterminé initialement (constitué de plusieurs milliers de mots de 3 à 7 lettres de $\mathcal{A}$ répondant à des critères dépendant du génome étudié),
- l'utilisation d'un ensemble de graines (de 3 à 6 lettres sur $\mathcal{A}' = \mathcal{A} + \{-\}$, où $-$ est la lettre joker) présentant certaines contraintes structurales, pour simultanément déterminer des couples de mots contenant des graines identiques dans deux dictionnaires de génomes de bactéries phylogénétiquement proches.

4 Logiciels prototypes

4.1 SIGffRid

SIGffRid est en cours de réalisation; il permet de rechercher des motifs potentiellement promoteurs (cf 5.2) à partir d'un génome bactérien donné, et d'un panel de séquences localisées en amont

de gènes orthologues putatifs entre ce génôme et celui d'une autre espèce. Les motifs résultats sont composés de couples de mots issus des paires de séquences amonts d'orthologues :

- chacun des mots étant sur-représenté dans le génome (utilisation du programme **R'MES** [57] pour définir un dictionnaire de mots de 3 à 7 lettres),
- présents dans les deux séquences avec un espacement voisin (paramètre: ± 1 nucléotide),
- ayant un écart typique d'un site de régulation de la transcription (paramètres: 15 à 20 lettres),
- statistiquement sur-représentés dans l'ensemble des séquences intergéniques amont du génome.

A termes, les couples de mots définis pourront posséder des gaps, s'adaptant donc à la forme de motif recherchée, sous réserve d'une bonne connaissance structurale de ce dernier. **SIGffRid** se devra d'être le plus convivial possible, permettant l'interprétation des résultats sans ambiguïté. Ainsi donnera-t-il l'ensemble des sites de fixation de facteurs de transcriptions potentiels liés à un motif donné en fournissant les annotations des gènes concernés, informations indispensables pour les prémices d'une interprétation biologique. Acceptant des fichiers issus de Genbank ou Embl, il devrait à terme se baser sur les versions XML de ces derniers, garantissant ainsi la possibilité de traitement de n'importe quel couple de génomes bactériens.

4.2 CurvaDis

CurvaDis est un programme qui permet de calculer la courbure en tout point d'une courbe discrète. CurvaDis repose sur les résultats de géométrie discrète présentés en 3.1.2 concernant les courbes discrètes 2D et 3D bruitées. Nous développons plusieurs versions de ce logiciel en fonction des applications envisagées :

- **CurvaDis_2D**: ce logiciel permet de calculer la courbure en chaque point d'une courbe discrète de dimension 2. Cette version est disponible et est utilisée dans le cadre de l'application à la détection de cercles et d'arcs de cercles dans des images techniques décrite dans la section 6.2.
- **CurvaDis_3D**: ce logiciel permet de calculer la courbure en chaque point d'une courbe discrète de dimension 3.
- **CurvaDis_ADN**: ce logiciel est en cours de réalisation. Il repose sur le calcul de la courbure en chaque point d'une courbe discrète 3D. De plus il tient compte des spécificités de la représentation tridimensionnelle de l'ADN et des différents modèles existant. Il permet à l'utilisateur de calculer la courbure de l'ADN en chaque point de la représentation 3D de l'ADN sans utiliser de notion de fenêtre glissante ce qui le démarque des autres logiciels comme **curvature**, **diamod** ou **bend** [58, 22]. Nous souhaitons mettre en place une interface conviviale, incluant les principaux modèles tridimensionnels de l'ADN, permettant une utilisation simple du logiciel par les biologistes. Une visualisation graphique des résultats obtenus sera proposée pour en faciliter l'interprétation.

5 Applications bioinformatiques

Sur les trois applications présentées dans cette section, les deux premières se situent dans le cadre de deux opérations (opérations 1 et 3) du thème *Bioinformatique et Applications à la Génomique* du PRST (Pôle de Recherche Scientifique et Technologique) Intelligence Logicielle. Elles ont été initiées depuis plus de quatre années en collaboration avec des chercheurs en biologie nancéiens. La troisième application a été engagée plus récemment et les résultats obtenus seront utilisés dans le cadre de l'opération 3.

5.1 Problème de phase en cristallographie

La cristallographie par rayons X constitue un outil essentiel pour la détermination de la structure des macromolécules organiques (ADN, ARN, protéine, etc.). En effet, si d'autres méthodes telles que la résonance magnétique nucléaire (RMN) fournissent généralement de bons résultats, celles-ci ne sont pas toujours applicables.

L'analyse de structures par rayons X comporte 3 étapes :

1. purification et cristallisation de la molécule étudiée ;
2. expérimentation aux rayons X dans un accélérateur de particules ;
3. résolution du problème inverse de la théorie de la diffraction pour déterminer la fonction de répartition de la densité électronique dans le cristal et interprétation de cette fonction en termes d'atomes.

La densité électronique dans un cristal est une fonction réelle positive périodique. Elle peut donc être exprimée comme la somme d'une série de Fourier. Les coefficients de Fourier, aussi appelés *facteurs de structure* sont des nombres complexes que l'expérimentation aux rayons X permet théoriquement de mesurer. Malheureusement, seuls les modules de ces coefficients sont mesurés. Pour obtenir la densité électronique, il faut donc déterminer les phases par d'autres techniques. Ceci constitue le *problème de phasage* en cristallographie [33, 29].

Le problème consiste à déterminer une fonction réelle positive dont les modules des coefficients de la transformée de Fourier approximent au mieux les valeurs mesurées.

Dans un travail pionnier, V. Lunin, A. Urshumtsev et A. Bockmayr [44] ont montré que le problème de phasage peut être modélisé et résolu par la programmation entière binaire. La méthode consiste à rechercher une fonction à valeurs dans $\{0,1\}$ aux noeuds d'un maillage de la cellule élémentaire du cristal. Les équations liant la fonction recherchée aux coefficients de la transformée de Fourier sont linéarisées puis résolues par des techniques d'optimisation entière. On obtient ainsi un *masque* de la molécule étudiée, c'est à dire une information sur les zones où la densité électronique est la plus élevée. La méthode produit de nombreuses solutions qui doivent d'abord être filtrées par des critères de vraisemblance puis les solutions restantes sont classées après alignement et des moyennes sont calculées dans chaque classe.

Cette méthode a produit des premiers résultats encourageants mais souffre de quelques inconvénients. (i) L'introduction des variables de phase se fait de manière *artificielle* car les données issues de l'expérimentation ne contiennent de fait aucune information sur la valeur de ces variables. (ii) La complexité des algorithmes de résolution limite fortement la taille du maillage que l'on peut considérer.

Nous avons donc, en collaboration avec V. Lunin, A. Urshumtsev et A. Bockmayr, exploré une autre voie qui exploite les équations de Patterson [46]. Ces équations relient directement les modules des coefficients de la transformée de Fourier à la densité électronique et permettent donc de s'affranchir totalement des phases. Là encore, nous recherchons une fonction à valeurs discrètes mais cette fois dans $\{0, \dots, M\}$ où M est un paramètre qui détermine la finesse de la fonction recherchée. Les équations obtenues ne sont plus linéaires mais leur forme particulière permet une résolution efficace par minimisation d'une fonction objectif par des techniques de recherche locale. La méthode consiste à rechercher itérativement dans un voisinage de la fonction courante une fonction qui améliore la fonction objectif. Le voisinage est défini par des combinaisons d'opérations élémentaires consistant à incrémenter ou décrémenter la valeur de la fonction en un point.

Cette nouvelle approche produit de bons résultats et a notamment permis d'accroître de manière significative la taille du maillage utilisé. En effet la technique de recherche locale permet de réduire

fortement l'impact de la taille du maillage sur les temps de calcul. Mais cette méthode fournit encore de nombreuses solutions dont beaucoup ne satisfont pas aux critères de vraisemblance. Notre but est donc de l'améliorer dans plusieurs directions dont l'une des plus importantes est la connexité de la molécule recherchée. Nous travaillons actuellement à l'intégration en amont de ce critère, c'est-à-dire que nous tentons de modifier la méthode de manière à n'engendrer que des solutions connexes. La notion de connexité doit cependant être définie soigneusement car d'une part nous travaillons dans le domaine discret et d'autre part, nous nous intéressons à une fonction périodique. Il s'agit donc de connexité sur un tore en dimension 3. De plus, la connexité n'est pas suffisante en pratique. En effet, une molécule composée de deux parties denses qui ne sont reliées que par des zones où la densité électronique est très faible ne nous intéresse pas a priori. Nous considérons donc plutôt une notion de densité locale de la molécule. Nous étudions diverses heuristiques favorisant la génération de solutions *connexes* selon cette notion.

Un autre critère que nous tentons d'introduire en amont dans la méthode de résolution est l'*atomicité* qui exprime que la fonction de répartition de la densité électronique dans la molécule est la somme des fonctions correspondant aux atomes individuels. Ce critère, qui est intrinsèquement linéaire, s'avère difficile à intégrer à nos équations non linéaires. Ce problème est à l'étude actuellement.

Enfin nous travaillons à l'optimisation de la méthode de résolution elle-même afin de pouvoir considérer un maillage plus fin de la cellule élémentaire du cristal. La finesse de ce maillage est en effet un paramètre essentiel de la viabilité de la méthode. Ces optimisations sont de deux ordres. D'une part, nous tentons d'accélérer les calculs par des techniques de tabulation des calculs intermédiaires. D'autre part, nous tentons de réduire la quantité même de ces calculs par une définition plus fine des voisinages considérés lors de la recherche locale.

5.2 Analyse de signaux de régulation dans les séquences nucléotidiques

L'expression du génome est un processus biologique extrêmement complexe. Elle se passe à plusieurs stades – d'abord au stade de la *transcription* de la molécule d'ADN, ensuite au stade de la *maturation* et de l'*épissage* de l'ARN messager (chez les organismes supérieurs), et enfin au stade de la *traduction* de l'ARN messager en protéine. À chacun de ces stades, un complexe protéique entre en interaction avec la molécule nucléotidique pour initier et piloter l'étape correspondante d'expression. La biologie de cette interaction entre la protéine et l'ARN/ADN n'est pas encore bien comprise. Cependant, la présence d'un *signal* dans la molécule d'ARN/ADN reconnu par la protéine est une condition nécessaire (mais pas suffisante) pour que cette interaction ait lieu. En terme de séquence, il s'agit donc d'un *motif* qui code ce signal. Identifier et décrire ces motifs est un problème d'une très grande importance pour les biologistes, mais il est aussi d'une difficulté notoire. L'approche courante à ce problème consiste à se donner un échantillon de séquences censées contenir le signal en question, et ensuite chercher un motif significatif commun à ces séquences. Plusieurs méthodes et logiciels existent pour cette tâche. Certains sont comparés dans un article récent [61] qui montre la diversité et le nombre de solutions proposées pour la résolution de ce problème fondamental en bioinformatique. Cependant la plupart de ces logiciels (par exemple MEME [5]) ne peuvent pas utiliser d'espacement variable entre deux mots (forme majoritairement prise par les motifs recherchés) ou font appel à des méthodes pour lesquelles le bruit de fond peut interférer avec le motif biologique (BioProspector [42]).

En résumé, ce problème comporte des aspects à la fois biologiques et informatiques qui interagissent étroitement. Du côté biologique, il est capital de mieux comprendre le mécanisme d'interac-

tion protéine-ADN/ARN pour spécifier la forme de motifs que l'on recherche. Du côté informatique, de nouvelles méthodes doivent être mises en place qui, compte tenu de ces réalités biologiques, pourront identifier des motifs communs de façon souple et efficace.

Nous travaillons sur ce thème en collaboration avec des chercheurs du Laboratoire de Génétique et de Microbiologie de l'Université Henri Poincaré de Nancy (P. Leblond, B. Aigle) depuis 2001 et plusieurs étudiants y ont participé, une thèse se déroule actuellement sur ce sujet. Ce travail a pour objectif de décrire les sites de fixation dans les zones promotrices du génome de la bactérie *Streptomyces coelicolor*, dont la famille produit approximativement 70% des antibiotiques connus.

Après une étude de la *sur-représentation* statistique supposée des sites de fixation des facteurs de transcription (utilisation du programme R'MES³ [57]) afin de les identifier [LDRA⁺03], nous avons utilisé une *approche comparative*, c'est à dire fondée sur la comparaison de génomes de plusieurs bactéries relativement proches sur le plan phylogénétique. Cette approche repose sur l'hypothèse que des protéines de deux bactéries issues d'une bactérie ancestrale doivent avoir gardé les mêmes mécanismes (et donc les mêmes sites) de régulation. Des résultats intéressants ont été obtenus, notamment avec deux espèces (*Streptomyces coelicolor* et *Streptomyces avermitilis*) et trois espèces (*Streptomyces coelicolor*, *Streptomyces avermitilis* et *Mycobacterium tuberculosis*) [TLDR⁺03]. Néanmoins, l'approche, basée sur des modèles de Markov, s'est avérée inadaptée à la recherche de *boîtes* à espacement variable, soulevant la nécessité du développement d'une méthode apte à prendre en compte toutes les caractéristiques connues de ces motifs, pour mieux cibler nos recherches et choisir des modes de sélection statistiques à orientation biologique.

Le programme SIGffRid (cf. 4.1) a donc été développé au cours de l'année passée. Définissant les boîtes via le programme R'MES [57], il les recherche dans un premier temps par couples conservés dans les séquences intergéniques amonts de paires d'orthologues en autorisant une légère variation de l'espace les séparant. Il tient donc compte à la fois des propriétés structurales des motifs (espace entre les boîtes, boîtes, variations), et de la conservation des motifs régulateurs entre bactéries phylogénétiquement proches (orthologues). Ce programme a été utilisé avec les génomes de *Streptomyces coelicolor* et *Streptomyces avermitilis*, nous avons pu retrouver directement deux motifs de sites de fixation de facteurs σ ou des variations de sites connus mais jusqu'à présent mal caractérisés (dont sigR: ggaat-n(18)-gttg, BldN: cgtaac-n(16)-cgttga, HrdB...) avec de nombreux nouveaux gènes potentiellement régulés par ces facteurs (un seul site est connu biologiquement pour BldN pour le moment). Au moins trois sites de fixation de facteurs sigma pourraient être mieux définis en plus de ceux déjà cités, et un site inédit a pu être déterminé (à confirmer par amélioration de la méthode et éventuellement expérience biologique). Ce travail a donné lieu à une communication orale à la dernière conférence JOBIM à Lyon en juillet 2005 [TSDR⁺05].

Plusieurs améliorations doivent encore être apportées, notamment:

- l'utilisation de regroupement par doublet de trinuécléotides en autorisant l'un d'eux à présenter un trou.
- l'association possible d'un motif avec une zone intrinsèquement courbée de l'ADN (utilisation du logiciel CurvaDis développé dans l'équipe (cf. 5.3).

5.3 Courbure de l'ADN

Dans le cadre du thème de recherche de notre équipe sur la régulation des gènes, nous étudions la courbure de l'ADN [10]. En effet, pour certains gènes, elle constitue un facteur important dans

3. <http://www.inra.fr/bia/J/AB/genome/RMES2>

de nombreux processus biologiques : transcription, initiation, réplication de l'ADN, etc. L'objectif général de ce travail est d'étudier le rôle de la courbure de l'ADN dans la régulation des gènes.

Les différentes notions de courbure de l'ADN reposent sur la représentation de la structure 3D de l'ADN; divers modèles figurent dans la littérature mais l'idée générale consiste à représenter l'ADN comme un tube de diamètre constant. Plusieurs logiciels de modélisation de la courbure de l'ADN existent (**curvature**, **diamod**, **bend**) et utilisent des approches diverses [58, 22] mais, pour tous, le calcul de la courbure dépend d'un paramètre, fixé par l'utilisateur, correspondant à la largeur d'une fenêtre "glissant" le long de la séquence étudiée. La modification de ce paramètre implique de grandes variations dans les résultats obtenus. Des résultats récents de géométrie discrète nous ont permis de proposer un nouvel algorithme pour calculer la courbure en chaque point d'une courbe discrète 3D bruitée sans utiliser l'approche "fenêtre glissante" [Rap04]. Cet algorithme a été optimisé [Ngu05] et implémenté dans un logiciel prototype nommé **CurvaDis** (cf. 4.2). Les premiers résultats obtenus sur certaines régions promotrices connues pour leurs fortes courbures sont encourageants et d'autres tests sont en cours.

Le logiciel résultant sera mis à la disposition de la communauté scientifique par l'intermédiaire d'une plate-forme Web. Les principaux modèles de structure 3D y seront proposés ainsi que les calculs de la courbure qui y sont associés. Par la suite, dans la continuité des travaux déjà engagés sur les protéines régulatrices H-NS/FIS présentes dans les bactéries, nous souhaitons appliquer les nouveaux modèles de courbure que nous aurons développés pour étudier d'autres mécanismes de régulation. Nous testerons aussi ce logiciel dans le cadre de l'identification et la classification des promoteurs dans les génomes procaryotes (cf. 5.2), sujet de collaboration entre notre équipe et le Laboratoire de Génétique et de Microbiologie de l'université Henri Poincaré de Nancy.

6 Applications en imagerie

Nous proposons dans cette section cinq applications de nos travaux en imagerie, les deux premières sont en cours de réalisation, la première ayant déjà donné lieu à plusieurs publications, et les trois dernières seront engagées prochainement.

6.1 Approximation polygonale multiordre de courbes discrètes

Généralement, l'objectif d'une méthode d'approximation polygonale est de fournir un découpage d'une chaîne de points en segments de façon à minimiser un critère d'erreur global ou à ne pas dépasser localement une erreur. L'intérêt est de donner une description qui soit plus compacte et plus adaptée pour des traitements ultérieurs d'interprétation ou de reconnaissance de formes. Les chaînes de points sont principalement issues d'un détecteur de contours dans des images à niveaux de gris ou d'un squelette dans des documents graphiques.

De nombreuses méthodes concernant l'approximation polygonale ont été proposées ces dernières années [20, 54, 68]. Dans certains cas [49, 60, 65] on examine successivement les points pour déterminer le plus long segment qui vérifie un seuil de tolérance prédéfini. Perez et Vidal [47] ont donné un algorithme qui minimise une erreur globale en se basant sur la programmation dynamique. L'approche souffre néanmoins de sa complexité élevée car toutes les configurations sont examinées et, pour chacune d'elles, une erreur globale d'approximation est calculée. Cet algorithme a suscité beaucoup d'intérêt et de récents travaux [37, 56] ont été proposés pour réduire la complexité de l'approche. Par ailleurs, des méthodes de division/fusion [50, 53] se focalisent sur la division récursive

d'une chaîne de points en segments si le découpage excède un seuil d'erreur prédéfini et le cas échéant fusionnent les segments. Dans d'autres approches [4, 53, 67] des points d'accroche correspondant à des points de forte courbure sont injectés dans le processus de segmentation.

Malgré le nombre important de méthodes d'approximation, il subsiste des problèmes majeurs de robustesse, de stabilité aux transformations géométriques et de complexité. De plus, les algorithmes reposent sur des seuils de tolérance d'erreur qui sont définis manuellement sans aucune connaissance sur la valeur du seuil le plus pertinent. Ceux-ci peuvent aussi être différents d'une chaîne de points à l'autre.

Dans cette perspective, en collaboration avec Antoine Tabbone et Laurent Wendling de l'équipe QGAR, nous avons proposé une analyse multi-niveaux qui permet de définir une méthode exempte de seuil et de décrire une chaîne de points en ses segments flous (cf. 3.1.1) les plus significatifs. Pour ce faire une chaîne de points est décomposée dans un espace multi-niveaux. Le premier niveau correspond à l'épaisseur la plus fine dans le cas discret. Le haut niveau est défini suivant un critère de stabilité entre deux niveaux successifs. Ensuite les extrémités des segments flous du haut niveau sont repositionnées à des niveaux inférieurs.

Contrairement aux autres approches où l'erreur d'approximation est définie manuellement, ici la méthode est exempte de seuils et produit le meilleur compromis entre le nombre de segments et leur localisation. L'approche propose un découpage automatique de la chaîne de points en ses segments principaux. Les résultats obtenus sont prometteurs [DRTW05, DRTW04].

Nous souhaitons poursuivre dans cette voie en améliorant la technique utilisée et en l'appliquant à une classification automatique de formes.

6.2 Détection de cercles et arcs de cercles dans des images techniques

Les verrous scientifiques permettant une recherche précise de symboles dans des documents techniques sont loin d'être levés : évaluation de la qualité des documents et des descripteurs utilisés lors de la reconnaissance, gestion des occultations, passage à l'échelle, etc.

Une collaboration a été engagée avec Laurent Wendling et Jean-Pierre Salmon de l'équipe QGAR au sujet de l'étude des descripteurs les plus adéquats pour la recherche de symboles complexes dans des documents techniques. Nous étudions les descripteurs obtenus en utilisant les graphes de courbures (cf. 3.1.2), fournis par le logiciel *Curvadis_2D* (cf. 4.2), des symboles des documents étudiés. Plusieurs techniques similaires ont déjà été proposées, cependant, notre approche a la spécificité de pouvoir s'adapter au bruit et à la taille des éléments à détecter dans les documents techniques, cela en fait son originalité et son efficacité.

Les opérateurs fiables de reconnaissance des formes doivent être invariants à, ou permettre de retrouver facilement, des transformations géométriques simples comme la rotation, le changement d'échelle et la translation. Certaines familles de symboles peuvent aussi être caractérisées par une opération d'étirement. Les descripteurs sélectionnés permettront d'évaluer de manière précise une telle transformation et de l'intégrer dans le mécanisme de reconnaissance.

Ce travail se situe dans le cadre du consortium européen FRESH auquel participe l'équipe QGAR et qui porte sur l'analyse et l'interprétation du document technique de type avionique.

6.3 Détection semi-automatique d'ellipses dans des images numériques

Les contours circulaires sont fréquents dans les environnements structurés par l'activité humaine, tels que les milieux industriels, les sites architecturaux, etc. Sous l'hypothèse de distorsions d'optique

négligeables, leur projection perspective dans les images de caméras ou d'appareils photographiques numériques les transforme en contours elliptiques. Inversement, l'interprétation tridimensionnelle de contours elliptiques en tant que projection de cercles sur le plan image peut contribuer à résoudre de multiples problèmes d'imagerie tridimensionnelle : étalonnage ou localisation des capteurs, suivi de cibles, reconstruction d'objets [48].

Ces contours ne sont pas toujours très nets, ni très contrastés et ne sont parfois que partiellement visibles. Leur extraction automatique n'est donc généralement pas aisée. Par ailleurs, la présence de l'homme est souvent utile pour les détecter rapidement dans l'image et les interpréter correctement en fonction du contexte. Nous nous proposons de concevoir et valider un mode semi-automatique pour la détection de contours elliptiques, mettant en œuvre une coopération entre l'homme et la machine, dans laquelle l'homme a la charge d'effectuer une première détection et de valider les solutions plus fines calculées par la machine. Le principe de ce mode d'assistance s'appuie sur une étude analogue [EM00] concernant l'extraction et l'interprétation de contours droits à partir d'une technique de vote (transformée de Hough). Nous envisageons dans le cas présent de nous appuyer sur l'efficacité et l'exactitude des algorithmes de géométrie discrète, et plus particulièrement sur une extension des travaux en cours dans l'équipe sur la détection de cercles, pour répondre aux questions suivantes :

- Existe-t-il un ou plusieurs contours elliptiques à proximité de la solution initiale?
- S'il y en a plusieurs, peuvent-ils être la projection de cercles concentriques?
- Quelles sont leurs caractéristiques géométriques?
- Quelle est la précision des paramètres extraits?

Cette problématique s'inscrit dans le cadre plus général de l'étude d'assistances semi-automatiques basées sur l'algorithmique discrète pour la reconstruction tridimensionnelle de scènes à partir d'images.

6.4 Déformation de surfaces

Dans le cadre d'une collaboration avec les archéologues de l'Institut de Préhistoire et de Géologie du Quaternaire de Bordeaux, nous avons pu nous intéresser au problème d'un des premiers systèmes de notation présentant des cavités sur sa surface. Les archéologues s'intéressent à l'interprétation de ces marques accumulées au cours du temps. L'analyse morphologique de ces marques permet d'émettre certaines hypothèses par rapport à leur origine au cours du temps. Plus précisément, le geste et l'outil ayant permis de créer ces marques jouent aussi un rôle important dans leur interprétation. Un apport important pour les archéologues serait de pouvoir simuler la déformation de la surface à partir de la modélisation de l'outil et du geste ayant permis de faire la marque.

A travers ce problème archéologique, nous pensons définir une méthode discrète de déformation de surfaces basée sur la modélisation discrète de l'objet d'origine. L'avantage d'utiliser un modèle discret se situe dans l'expression des contraintes définies à travers l'interaction de l'outil et de la surface. Afin d'obtenir une interaction rapide, il serait intéressant d'intégrer une représentation à différents niveaux de résolution de la surface discrète.

6.5 Reconstruction à base de sketching

Dans des travaux antérieurs, nous avons introduit une méthode de création de formes en trois dimensions à partir de dessins. L'idée principale est de dessiner de manière intuitive la surface vue

sous différentes directions d'éclairage. La première étape consiste dans un premier temps à dessiner les contours et caractéristiques de la forme que l'on souhaite reconstruire. Il s'agit ensuite pour chaque image de dessiner intuitivement les nuances d'intensité lumineuse pour donner du relief aux différents dessins en fonction de la direction d'éclairage. La reconstruction s'effectue ensuite à partir de la méthode discrète décrite dans la section 3.1.3 [BK05].

Cette méthode permet l'édition et la création de formes qui se limitent pour l'instant à une carte de hauteurs. Nous souhaitons par la suite étendre ce concept pour la création et l'édition d'objets entièrement en trois dimensions. Nous envisageons pour cela d'adapter la méthode de reconstruction pour l'appliquer directement sur un volume initial. Ce dernier pourrait par exemple être déduit à partir de contours associés au profil dessiné par l'utilisateur.

D'autres perspectives de recherche connexes à ce problème méritent d'être explorées. Dans un premier temps, nous voulons définir des outils pour obtenir une certaine interaction par rapport à la création des différents dessins avec la surface que l'utilisateur souhaite reconstruire. Cette interaction concerne tout d'abord l'aide à l'utilisateur pour la création des dessins en utilisant par exemple des techniques de dégradés à partir de segments discrets. Ensuite, il s'agit de pouvoir donner des critères pour utiliser le modèle de réflectance de la lumière le mieux adapté pour l'édition ou la création des différents dessins en fonction des détails de la surface à reconstruire ou à éditer.

7 Projets, collaborations et contacts

7.1 Collaborations internationales

Dans le cadre de l'opération *Application de méthodes d'optimisation discrète à la résolution du problème de phases en cristallographie macromoléculaire* du projet *Bioinformatique et application à la génomique* du PRST Intelligence logicielle, Eric Domenjoud collabore avec Vladimir Lunin de l'Institut pour les problèmes mathématiques en biologie (IMPB) de l'Académie des Sciences de Russie à Puschino et avec Alexander Bockmayr du laboratoire Matheon de l'Université libre de Berlin.

Dans le cadre d'une collaboration avec Y.-F.S. Pétermann, de l'Université de Genève, Jean-Luc Rémy poursuit une recherche en théorie des nombres sur les suites auto-décrites. Ces préoccupations sont reliées à l'intérêt de l'équipe pour les définitions arithmétiques d'objets discrets. La collaboration, qui dure depuis 1996, s'est traduite en 2004 par deux séjours à Genève de Jean-Luc Rémy, respectivement de 8 et 10 jours.

Bertrand Kerautret et Isabelle Debled-Rennesson sont membres du comité technique IAPR sur la géométrie discrète (TC18)⁴.

Isabelle Debled-Rennesson collabore dans le cadre de la géométrie discrète avec le professeur Ulrich Eckhardt et une de ses étudiantes Hélène Reiter-Dorksen de l'Université d'Hambourg, Allemagne, ainsi qu'avec Reinhardt Klette de l'Université d'Auckland, Nouvelle-Zélande et Valentin Brimkov de l'Université de Fredonia, USA.

4. <http://www.cb.uu.se/TC18/>

7.2 Collaborations nationales

Bertrand Kerautret et Isabelle Debled-Rennesson ont participé à l'Action Spécifique *Géométrie Discrète et Géométrie Algorithmique* pour l'année 2004. Faisant suite à cette action spécifique, une journée est organisée cette année, soutenue par le GDR ALP et le CNRS. Elle a pour but de rapprocher les communautés de géométrie algorithmique et de géométrie discrète et la majorité des membres de l'équipe ADAGIo y ont participé.

Fabrice Touzain et Isabelle Debled-Rennesson participent à l'ACI IMPBio dans le cadre du projet FouDAnGA (Fouille de données pour l'annotation de génomes d'actinomycètes), accepté en 2004.

Philippe Even est membre du GDR-PRC ISIS et est rattaché au thème "Forme, Mouvement, Structure de scènes". Il participe également au Groupe de Travail inter-GDR sur la Réalité Virtuelle (GT/RV).

Bertrand Kerautret a mené un projet de collaboration avec Francesco d'Errico, directeur de recherche au CNRS à l'Institut de Préhistoire et de Géologie du Quaternaire. Il a également entrepris un travail en collaboration avec Xavier Granier de l'équipe INRIA à travers le projet IPARLA [KGB05].

Dans le cadre de la géométrie discrète, plusieurs collaborations nationales sont engagées : avec Yan Gérard et Fabien Feschet (LLAIC, Clermont-Ferrand), Isabelle Sivignon et David Coeurjolly (LIRIS, Lyon), Jaques Olivier Lachaud et Achille Braquelaire (LaBRI, Bordeaux), Mohamed Tajine (LSIIT, Strasbourg).

Dans le cadre des applications en bioinformatique, une collaboration avec Sophie Schbath (INRA, Jouy) est en cours ainsi qu'avec Alexander Urshumtsev du laboratoire LCM3B de l'UHP et avec Pierre Leblond et Bertrand Aigle du laboratoire LGM INRA-UHP.

7.3 Collaborations et positionnement au sein du LORIA

Nous avons des collaborations scientifiques avec les équipes et les personnes suivantes :

- ORPAILLEUR (J.-F. Mari) sur le thème des promoteurs dans les génomes procaryotes (section 5.2),
- QGAR (A. Tabbone, L. Wendling, X. Hilaire, J.-P. Salmon) sur l'utilisation d'algorithmes de géométrie discrète pour l'analyse de documents (sections 6.1 et 6.2),
- SPACES (P. Zimmermann) sur la reconnaissance des morceaux de plans discrets (section 3.1.1).

Nous avons également des liens scientifiques avec l'équipe VEGAS, et nous avons participé à une action spécifique commune en 2003-2004 ayant pour but de rapprocher les communautés de géométrie algorithmique et de géométrie discrète.

Au niveau des applications en bioinformatique, l'équipe MODBIO s'intéresse également à la structure des macromolécules organiques. La problématique et les techniques utilisées sont néanmoins totalement différentes. Nous nous intéressons à la génération de la géométrie globale des molécules à partir de données expérimentales tandis que l'équipe MODBIO prédit des propriétés locales de cette structure par des techniques d'apprentissage statistique.

7.4 Actions régionales

Plusieurs membres de l'équipe sont impliqués dans le projet *Bioinformatique et application à la génomique* du PRST Intelligence Logicielle :

- Isabelle Debled-Rennesson et Fabrice Touzain participent à l'opération *Fouille de données pour l'annotation de génomes d'actinomycètes*.
- Eric Domenjoud est coordonnateur de l'opération *Application de méthodes d'optimisation discrète à la résolution du problème de phases en cristallographie macromoléculaire*.

8 Participation à la vie scientifique

8.1 Comités de programme et évaluation d'articles

Isabelle Debled-Rennesson a été :

- Membre du *comité de programme* des conférences : 10ème et 11ème International Workshop on Combinatorial Image Analysis (IWCIA 2004 et IWCIA 2006) ainsi que 12ème et 13ème International Conference on Discrete Geometry for Computer Imagery (DGCI 2005 et DGCI 2006).
- Relectrice d'articles pour un numéro spécial de LNCS sur "Digital and Image Geometry" ainsi que pour les revues Computer Graphics Forum, Computer & Graphics, Theoretical Computer Science et Journal of Discrete Algorithm.

Bertrand Kerautret a été relecteur d'article pour un numéro spécial de la revue Computer & Graphics.

8.2 Jurys de thèse

Isabelle Debled-Rennesson a participé ou participera aux *jurys de thèse* suivants :

- Membre du jury de thèse de Xavier Hilaire, de l'équipe QGAR du Loria, soutenue le 15 janvier 2004.
- Membre du jury de thèse d'Hélène Dörksen-Reiter, de l'Université de Hambourg, soutenue le 13 novembre 2004.
- Membre du jury de thèse de Damien Jamet, de l'Université de Montpellier qui sera soutenue le 13 décembre 2005.

Jean-Luc Rémy a participé au *jury de la thèse* d'Olivier Binda, à l'UHP, en novembre 2004.

Philippe Even a participé à quatre *jurys de thèse* dont deux en tant que rapporteur et à un *jury d'Habilitation à Diriger des Recherches* en tant que rapporteur :

- Membre du jury de thèse de Fabrice Harrouet, de l'Ecole Nationale d'Ingénieurs de Brest, soutenue le 8 décembre 2000.
- Membre du jury de thèse de Samir Otmane, au CEMIF d'Evry, soutenue le 13 décembre 2000.
- Rapporteur de la thèse d'Omar Al Khalil, à l'ENSAIS de Strasbourg, soutenue le 12 septembre 2002.

- Rapporteur de la thèse de Fakhreddine Ababsa, au CEMIF d'Evry, soutenue le 16 décembre 2002.
- Rapporteur du mémoire de HDR de Laurent Nana à l'Université de Bretagne Occidentale, soutenue le 9 septembre 2004.

8.3 Divers

Au niveau national, depuis octobre 2003, Isabelle Debled-Rennesson est membre élue du *conseil national des universités (CNU)* pour la section 27 et, depuis septembre 2004, *membre du bureau* et seconde vice-présidente du CNU de la section 27. De plus elle a été membre titulaire de la *commission de spécialistes* scientifique de l'IUFM de Lorraine de septembre 2000 à août 2003.

Jean-Luc Rémy est membre élu du *conseil de laboratoire* du Loria. Il est également membre du département Informatique de l'Ecole doctorale IAEM de Nancy.

Philippe Even est membre suppléant élu de la *commission de spécialistes* en 27^e section de l'université Henri Poincaré / Nancy 1 depuis janvier 2004.

Eric Domenjoud est membre de la *commission de spécialistes* en 27^e section de l'INPL.

Bertrand Kerautret a été membre du *comité local d'organisation* de la conférence internationale: 10th International Conference on Discrete Geometry for Computer Imagery (DGCi 2002).

Fabrice Touzain est responsable de l'organisation des séminaires en bioinformatique du Loria.

9 Publications des membres de l'équipe depuis 1995

Revue internationale et chapitres de livres

- [ARS97a] L. Alonso, J.-L. Rémy, and R. Schott. A linear-time algorithm for the generation of trees. *Algorithmica*, 17(2):162–182, February 1997.
- [ARS97b] L. Alonso, J.-L. Rémy, and R. Schott. Uniform generation of a Schröder tree. *Information Processing Letters*, 64:305–308, 1997.
- [BK05] A. Braquelaire and B. Kerautret. Reconstruction of lambertian surfaces by discrete equal height contours and regions propagation. *Image and Vision Computing*, 23(2):177–189, February 2005.
- [DRDR05] H. Dörksen-Reiter and I. Debled-Rennesson. Convex and concave parts of digital curves. In R. Klette, R. Kozera, L. Noakes, and J. Weickert, editors, *Geometric Properties from Incomplete Data*, volume 31 of *Computational Imaging and Vision*. Springer-Verlag, January 2005.
- [DRFRD06] I. Debled-Rennesson, F. Feschet, and J. Rouyer-Degli. Optimal blurred segments decomposition of noisy shapes in linear time. *Computers & Graphics*, 30(1), 2006. (in press).

- [DRR95] I. Debled-Rennesson and J.-P. Reveillès. A linear algorithm for segmentation of digital curves. *International Journal of Pattern Recognition and Artificial Intelligence*, 9(4):635–662, December 1995.
- [DRRRD03] I. Debled-Rennesson, J.-L. Rémy, and J. Rouyer-Degli. Detection of the discrete convexity of polyominoes. *Discrete Applied Mathematics*, 125(1):115–133, January 2003.
- [DRRRD05] I. Debled-Rennesson, J.-L. Rémy, and J. Rouyer-Degli. Linear segmentation of discrete curves into fuzzy segments. *Discrete Applied Mathematics*, 151:122–137, October 2005.
- [DRTW05] I. Debled-Rennesson, S. Tabbone, and L. Wendling. Multiorder polygonal approximation of digital curves. *Electronic Letters on Computer Vision and Image Analysis*, 5(2):98–110, August 2005. Special Issue on Document Analysis.
- [EM00] P. Even and A. Malavaud. Semi-automated edge segment specification for an interactive modelling system of robot environments. *International Archives of Photogrammetry and Remote Sensing*, 33(B5):222–229, July 2000.
- [Eve03] P. Even. Modélisation interactive pour la TAO. In A. Kheddar and P. Coiffet, editors, *Traité IC2: Téléopération et réalité virtuelle*, chapter 7, pages 193–216. Editions Hermès, Paris, 2003.
- [Eve04] P. Even. A generic procedure for interactive 3-D model registration on images. *International Archives of Photogrammetry and Remote Sensing*, 35(B5):204–209, July 2004.
- [GDRZ05] Y. Gérard, I. Debled-Rennesson, and P. Zimmermann. An elementary digital plane recognition algorithm. *Discrete Applied Mathematics*, 151:169–183, October 2005.
- [KB06] B. Kerautret and A. Braquelaire. A reversible and statistical method for discrete surfaces smoothing. *Computer & Graphics*, 30(1), 2006. (in press).
- [LEL⁺02] C. Leroux, P. Even, A. Lottin, R. Gelin, J.-M. Idasiak, J.-F. Boissonneau, and M. Jean-jacques. Flexible 3-D modelling of hostile environments with constrained viewpoint and multiple unknowns. *International Archives of Photogrammetry and Remote Sensing*, 34(5):175–180, September 2002.
- [PR98] Y.-F. Pétermann and J.-L. Rémy. Golomb’s self described sequence and functional differential equations. *Illinois Journal of Mathematics*, 42(3):420–440, 1998.
- [PR04] Y.-F.S. Pétermann and J.-L. Rémy. Increased self-described sequences. *Annals of Combinatorics*, 8:325–346, 2004.
- [PR06] Y.-F.S. Pétermann and J.-L. Rémy. On the cohen-olivier algorithm for computing zeta(s): Error analysis in the real case for an arbitrary precision. *Advances in Applied Mathematics*, 2006.
- [PRV99] Y.-F.S. Pétermann, J.-L. Rémy, and I. Vardi. On a functional-differential equation related to golomb’s self-described sequence. *Journal de théorie des Nombres de Bordeaux*, 11:211–230, 1999.
- [PRV01] Y.-F.S. Pétermann, J.-L. Rémy, and I. Vardi. Discrete derivatives of sequences. *Advances in Applied Mathematics*, 27:562–584, 2001.
- [Rém97] J.-L. Rémy. Sur la suite autodécrite de Golomb. *Journal of Number Theory*, 66(1):1–28, September 1997.

Actes de conférences internationales

- [BK03] A. Braquelaire and B. Kerautret. Reconstruction of discrete surfaces from shading images by propagation of geometric features. In I. Nystrom, G. Sanniti di Baja, and S. Svensson, editors, *Proceedings of the International Conference on Discrete Geometry for Computer Imagery*, volume 2886 of *LNCS*, pages 257–266, Naples, Italy, November 2003. Springer-Verlag.
- [CDRT01] D. Coeurjolly, I. Debled-Rennesson, and O. Teytaud. Segmentation and length estimation of 3D discrete curves. In G. Bertrand, A. Imiya, and R. Klette, editors, *Digital and Image Geometry*, volume 2243 of *LNCS*, pages 295–313. Springer-Verlag, 2001.
- [DR04] I. Debled-Rennesson. Estimation of tangents to a noisy discrete curve. In Longin Jan Latecki, David M. Mount, and Angela Y. Wu, editors, *Vision Geometry XII, Electronic Imaging*, volume 5300 of *Proceedings of the SPIE*, pages 117–126, San Jose, California, USA, January 2004.
- [DRFRD05] I. Debled-Rennesson, F. Feschet, and J. Rouyer-Degli. Blurred segments decomposition in linear time. In E. Andres, G. Damiand, and P. Lienhardt, editors, *Proceedings of the 12th International Conference on Discrete Geometry for Computer Imagery*, volume 3429 of *LNCS*, pages 371–382, Poitiers, France, April 2005. Springer-Verlag.
- [DRR96] I. Debled-Rennesson and J.-P. Reveillès. Incremental algorithm for recognizing pieces of digital planes. In *Proceedings of Spie's International Symposium on Optical Science, Engineering, and Instrumentation, Technical Conference, Vision Geometry 5*, Denver, USA, August 1996.
- [DRRRD00] I. Debled-Rennesson, J.-L. Rémy, and J. Rouyer-Degli. Detection of the discrete convexity of polyominoes. In G. Borgefors, I. Nyström, and G. Sanniti di Baja, editors, *Proceedings of the 9th International Conference on Discrete Geometry for Computer Imagery*, volume 1953 of *LNCS*, Uppsala, Sweden, December 2000. Springer-Verlag.
- [DRRRD03] I. Debled-Rennesson, J.-L. Rémy, and J. Rouyer-Degli. Segmentation of discrete curves into fuzzy segments. In *Proceedings of the 9th International Workshop on Combinatorial Image Analysis*, volume 12 of *Electronic Notes in Discrete Mathematics*, Palermo, Italy, May 2003.
- [DRTW04] I. Debled-Rennesson, S. Tabbone, and L. Wendling. Fast polygonal approximation of digital curves. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 1, pages 465–468, Cambridge, United Kingdom, August 2004.
- [EFG00] P. Even, R. Fournier, and R. Gelin. Using structural knowledge for interactive 3-D modeling of piping environments. In *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, pages 2013–2018, San Francisco, CA, April 2000.
- [EG05] P. Even and S. Gobron. Interactive three-dimensional reconstruction and weathering simulations on buildings. In *Proceedings of the 20th International Symposium of CIPA*, Torino, Italy, September 2005.
- [EGMF99] P. Even, P. Gravez, E. Maillard, and R. Fournier. Acquisition and exploitation of a 3D environment model for computer aided teleoperation. In *Proceedings of the 8th International Workshop on Robot and Human Interaction*, pages 261–266, Pisa, Italy, September 1999.

- [Eve01] P. Even. A top-down approach to interactive 3-D modelling of architectural sites. In *Proceedings of the 18th International Symposium of CIPA*, pages 585–592, Potsdam, Germany, September 2001.
- [Eve04] P. Even. Task-oriented 3D reconstruction. In *3D Modelling*, Paris, France, April 2004.
- [KB05] B. Kerautret and A. Braquelaire. A statistical approach for geometric smoothing of discrete surfaces. In E. Andres, G. Damiand, and P. Lienhardt, editors, *Proceeding of the International Conference on Discrete Geometry for Computer Imagery*, volume 3429 of *LNCS*, pages 405–414, Poitiers, France, April 2005. Springer-Verlag.
- [Ker04] B. Kerautret. A robust discrete approach for shape from shading and photometric stereo. In K. Wojciechowski, B. Smolka, H. Palus, R.S. Kozera, W. Skarbek, and L. Noakes, editors, *Proceedings of the International Conference on Computer Vision and Graphics*, volume 32 of *Computational Imaging and Vision*, pages 581–587, Warszawa, Poland, September 2004. Springer-Verlag.
- [KGB05] B. Kerautret, X. Granier, and A. Braquelaire. Intuitive shape modeling by shading design. In *Proceedings of the 5th International Symposium on Smart Graphics*, number 3638 in *LNCS*, pages 163–174, Frauenwoerth Cloister, Germany, August 2005. Springer-Verlag.
- [TDR99] P. Tellier and I. Debled-Rennesson. 3D discrete normal vectors. In G. Bertrand, M. Couprie, and L. Perroton, editors, *Proceedings of the 8th International Conference on Discrete Geometry for Computer Imagery*, volume 1568 of *LNCS*, pages 447–458, Marne-la-Vallée, France, March 1999. Springer-Verlag.
- [TLDR⁺03] F. Touzain, P. Lavigne, I. Debled-Rennesson, B. Aigle, P. Leblond, and G. Kucherov. Identification of transcription factor binding sites in *Streptomyces coelicolor* A3(2) by phylogenetic comparison. In *Proceedings of the European Conference for Computer Biology*, Paris, France, September 2003.

Actes de conférences nationales

- [DRTW04] I. Debled-Rennesson, S. Tabbone, and L. Wendling. Approximation polygonale à partir d’une analyse multi-ordres des points de contour. In *Colloque International Francophone sur l’Ecrit et le Document*, La Rochelle, France, June 2004.
- [Eve02] P. Even. Modélisation 3D interactive orientée-tâche de sites architecturaux. In *Journée Thématique : Maquette numérique et Patrimoine*, pages 27–34, Chalon sur Saône, France, March 2002.
- [KBdV03] B. Kerautret, A. Braquelaire, F. d’Errico, and M. Vanhaeren. Discrete geometry for parameters extraction and shape reconstruction in archeology. In *Colloque GMPCA, Archéométrie*, page 164, 2003.
- [LDRA⁺03] P. Lavigne, I. Debled-Rennesson, B. Aigle, P. Leblond, and G. Kucherov. Identification of transcription factor binding sites in the genome of *streptomyces coelicolor* a3(2). In *Journées Post Génomique de la Doua Lyon, France*, May 2003.
- [TSDR⁺05] F. Touzain, S. Schbath, I. Debled-Rennesson, B. Aigle, P. Leblond, and G. Kucherov. Sigffrid : Programme de recherche des sites de fixation des facteurs de transcription par approche comparative. In *Journées Ouvertes Biologie Informatique Mathématiques*, Lyon, France, July 2005.

Autres

- [DR95] I. Debled-Rennesson. *Etude et reconnaissance des droites et des plans discrets*. PhD thesis, Université Louis Pasteur, Strasbourg I, December 1995.
- [DR02] I. Debled-Rennesson. Discrete lines and segmentation of 2d an 3d curves. In *Workshop on Mathematical Aspects of Visual Cognition*, Hamburg, Germany, November 2002.
- [Eve00] P. Even. Modélisation 3D interactive de l'environnement de robots téléopérés. Mémoire d'Habilitation à Diriger des Recherches, Université de Versailles Saint-Quentin-en-Yvelines, Vélizy, December 2000.
- [Ker01] B. Kerautret. Extraction de paramètres d'objets préhistoriques par des méthodes d'analyses d'images. Mémoire de DEA, Université de Bordeaux 1 (LaBRI), June 2001.
- [Ker04] B. Kerautret. *Reconstruction en lissage de surfaces discrètes*. PhD thesis, Université de Bordeaux 1, December 2004.
- [Ngu05] T.P. Nguyen. Optimisation du calcul de la courbure de l'ADN. Mémoire de DEA, IFI, October 2005.
- [Pro05] L. Provot. Reconnaissance de morceaux de plans discrets bruités. Mémoire de DEA, LORIA, June 2005.
- [Rap04] F. Rapaport. Calcul de la courbure de l'ADN. Mémoire de DEA, Supelec, September 2004.
- [RdV⁺05] F.V. Ramirez Rozzi, F. d'Errico, M. Vanhaeren, P.M. Grootes, B. Kerautret, and V. Dujardin. Modern human and cutmarked neanderthal remains associated to the late auriacian. *Proceedings of the National Academics of Science*, 2005. (submitted).

10 Références externes

- [1] S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215:403–410, 1990.
- [2] A. Amir, M. Farach, Z. Galil, R. Giancarlo, and P. Kunssoo. Dynamic dictionary matching. *Journal of Computer and System Sciences*, 49(2):208–222, October 1994.
- [3] E. Andres. Le plan discret. In *Colloque de géométrie discrète en imagerie: fondements et applications*, Strasbourg, France, September 1993.
- [4] N. Ansari and E. J. Delp. On detecting dominant points. *Pattern Recognition*, 24(5):441–451, 1991.
- [5] T.L. Bailey and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent System for Molecular Biology*, pages 28–36, Stanford, California, USA, August 1994.
- [6] E. Barcucci, A. Del Longo, M. Nivat, and R. Pinzani. Reconstructing convex polyominoes from horizontal and vertical projections. *Theoretical Computer Science*, 155(2):321–347, 1996.
- [7] E. Barcucci, A. Del Longo, M. Nivat, and R. Pinzani. Medians of polyominoes: A property for the reconstruction. *International Journal of Imaging Systems and Technology*, 9:69–77, 1998.
- [8] J. Berstel. Tracé de droites, fractions continues et morphismes itérés. In *Mots: mélanges offerts à M.P. Schützenberger*, pages 298–309. Editions Hermès, 1990.
- [9] J. Berstel and P. Séébold. *Algebraic Combinatorics on Words*, chapter 2. Cambridge University Press, 2000. available at <http://www-igm.univ-mlv.fr/~berstel/Lothaire/>.

- [10] A. Bolshoy, P. McNamara, P. Harrington, and E.N. Trifonov. Curved DNA without a-a: experimental estimation of all 16 DNA wedge angles. *Proceedings of the National Academy of Science*, 88:2312–2316, 1991.
- [11] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.
- [12] L. Buzer. An incremental linear time algorithm for digital line and plane recognition using a linear incremental feasibility problem. In *Proceedings of the 10th International Conference on Discrete Geometry for Computer Imagery*, volume 2301 of *LNCS*, pages 372–381. Springer-Verlag, 2002.
- [13] J.-M. Chassery and A. Montanvert. *Géométrie discrète en analyse d’images*. Hermès, 1991.
- [14] M. Chrobak and C. Dürr. Reconstructing hv-convex polyominoes from orthogonal projections. *Information Processing Letters*, 69(6):283–291, March 1999.
- [15] D. Coeurjolly. *Algorithmes et géométrie discrète pour la caractérisation des courbes et des surfaces*. PhD thesis, Université Lumière, Lyon, 2002.
- [16] D. Coeurjolly and S. Svensson. Estimation of curvature along curves with application to fibres in 3d images of paper. *LNCS*, 2749:247–254, January 2003.
- [17] M. Crochemore, C. Hancart, and T. Lecroq. *Algorithmique du texte*. Vuibert Informatique, 2001.
- [18] M. Crochemore and W. Rytter. *Text algorithms*. Oxford University Press, 1994.
- [19] M. Crochemore and W. Rytter. *Jewels of Stringology*. World Scientific, 2002.
- [20] T.J. Davis. Fast decomposition of digital curves into polygons using the haar transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):786–790, 1999.
- [21] A. Del Lungo, M. Nivat, and R. Pinzani. The number of convex polyominoes reconstructible from their orthogonal projections. *Discrete Mathematics*, 157:65–78, 1996.
- [22] M. Dlakic and R.E. Harrington. Diamod: display and modeling of DNA bending. *Bioinformatics*, 14:326–331, 1998.
- [23] F. Feschet and L. Tougne. Optimal time computation of the tangent of a discrete curve: application to the curvature. *LNCS*, 1568:31–40, 1999.
- [24] O. Figueiredo and J.-P. Reveillès. A contribution to 3D digital lines. In *Proceedings 5th International Conference on Discrete Geometry for Computer Imagery*, pages 187–198, Clermont-Ferrand, September 1995.
- [25] J. Françon, J.M. Schramm, and M. Tajine. Recognizing arithmetic straight lines and planes. In *Proceedings of the 6th International Conference on Discrete Geometry for Computer Imagery*, volume 1176 of *LNCS*, pages 141–150, Lyon, France, 1996. Springer-Verlag.
- [26] J. Françon. Discrete combinatorial surfaces. *Graphical Models and Image Processing*, 57(1):20–26, January 1995.
- [27] Y. Gérard. *Contribution à la Géométrie Discrète*. PhD thesis, Université d’Auvergne, Clermont-Ferrand, 2000.
- [28] D. Gusfield. *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press, 1997.
- [29] H. A. Hauptman. Direct methods and anomalous dispersion. In Bo G. Malmström, editor, *Nobel Lectures, Chemistry 1981-1990*. World Scientific Publishing Co., Singapore, 1992, December 1985.
- [30] G.T. Herman. *Geometry of Digital Spaces*. Birkhauser, 1998.

- [31] R. M. Idury and A. A. Schäffer. Dynamic dictionary matching with failure functions. *Theoretical Computer Science*, 131(2):295–310, September 1994.
- [32] M.A. Jacob. *Les applications quasi-affines*. PhD thesis, Université Louis Pasteur, Strasbourg, 1993.
- [33] J. Karle. Recovering phase information from intensity data. In Bo G. Malmström, editor, *Nobel Lectures, Chemistry 1981-1990*. World Scientific Publishing Co., Singapore, 1992, December 1985.
- [34] W. J. Kent. BLAT—the BLAST-like alignment tool. *Genome Research*, 12:656–664, 2002.
- [35] R. Klette and A. Rosenfeld. *Digital Geometry*. Morgan Kaufmann publishers, Elsevier, 2004.
- [36] D. Knuth, J. Morris, and V. Pratt. Fast pattern matching in strings. *SIAM Journal on Scientific Computing*, 6:323–350, June 1977.
- [37] A. Kolesnikov and P. Fränti. Reduced-search dynamic programming for approximation of polygonal curves. *Pattern Recognition Letters*, 24(14):2243–2254, 2003.
- [38] T.Y. Kong and A. Rosenfeld. *Topological Algorithms for Digital Image Processing*. North Holland, 1996.
- [39] V. A. Kovalevsky. A new approach to shape from shading. In F. Solina, W.G. Kropatsch, R. Klette, and R. Bajcsy, editors, *Advances in Computer Vision*. Springer-Verlag, 1997.
- [40] R.S. Kozera and L.Noakes. Noise reduction in photometric stereo with non-distant light sources. In K. Wojciechowski, B. Smolka, H. Palus, R.S. Kozera, W. Skarbek, and L. Noakes, editors, *International Conference on Computer Vision and Graphics*, volume 32, pages 103–111, Warsaw, Poland, September 2004. Springer-Verlag.
- [41] C. Laurence, S. Altschul, M. Boguski, J. Liu, A. Neuwald, and J. Wooton. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignments. *Science*, 262:208–214, 1993.
- [42] X. Liu, D.L. Brutlag, and J.S. Liu. Bioprospector: discovering conserved dna motifs in upstream regulatory regions of co-expressed genes. *Pacific Symposium on Biocomputing*, pages 127–138, 2001.
- [43] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002. available at <http://www-igm.univ-mlv.fr/~berstel/Lothaire/index.html>.
- [44] V. Y. Lunin, A. Urzhumtsev, and A. Bockmayr. Direct phasing by binary integer programming. *Acta Crystallographica*, A58:283–291, 2002.
- [45] G. Navarro and M. Raffinot. *Flexible Pattern Matching in Strings*. Cambridge University Press, 2002.
- [46] A.L. Patterson. A fourier series method for the determination of the components of inter-atomic distances in crystals. *Physical Review*, 46:372–376, 1934.
- [47] J.C Perez and E. Vidal. Optimum polygonal approximation of digitized curves. *Pattern Recognition Letters*, 15:743–750, 1994.
- [48] L. Quan. Conic reconstruction and correspondence from two views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):151–160, February 1996.
- [49] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–256, 1972.
- [50] B.K. Ray and K.S. Ray. A new split-and-merge technique for polygonal approximation of chain coded curves. *Pattern Recognition Letters*, 16, February 1995.

- [51] J.-P. Reveillès. Structure des droites discrètes. In *Journées mathématique et informatique*, Marseille-Luminy, October 1989.
- [52] J.-P. Reveillès. *Géométrie discrète, calculs en nombre entiers et algorithmique*. Thèse d'état, Université Louis Pasteur, Strasbourg, 1991.
- [53] P. L. Rosin and G. A. West. Segmentation of edges into lines and arcs. *Image and Vision Computing*, 7(2):109–114, May 1989.
- [54] Paul L. Rosin. Techniques for assessing polygonal approximations of curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):659–666, June 1997.
- [55] H. Rushmeir, J. Gomes, L. Balmelli, F. Bernardi, and G. Taubin. Image-based object editing. In *Processing of the Fourth International Conference on 3-D Digital Imaging and Modeling*, pages 20–28, October 2003.
- [56] M. Salotti. An efficient algorithm for the optimal polygonal approximation of digitized curves. *Pattern Recognition Letters*, 22(2):215–221, February 2001.
- [57] S. Schbath. An efficient statistic software to detect over- and under-represented words in DNA sequences. *Journal Computer Biology*, 4:189–192, 1997.
- [58] A.S. Shpigelman, E.N. Trifonov, and A. Bolshoy. Curvature: Software for the analysis of curved DNA. *Computer Applications in the Biosciences*, 9(4):435–440, 1993.
- [59] I. Sivignon. *De la caractérisation des primitives à la reconstruction polyédrique de surfaces en géométrie discrète*. PhD thesis, Institut National Polytechnique de Grenoble, 2004.
- [60] J. Sklansky and V. Gonzalez. Fast polygonal approximation of digitized curves. *Pattern Recognition*, 12(5):327–331, 1980.
- [61] M. Tompa, N. Li, T.L. Bailey, G.M. Church, B. De Moor, E. Eskin, A.V. Favorov, M.C. Frith, Y. Fu, W.J. Kent, V.J. Makeev, A.A. Mironov, W.S. Noble, G. Pavesi, G. Pesole, M. Régnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, and Z. Zhu. Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnology*, 23(1):137–144, 2005.
- [62] C.W.A.M. van Overveld. Painting gradients: free-form surface design using shading patterns. In *Proceedings of Graphics interface*, pages 151–158. Canadian Information Processing Society, 1996.
- [63] A. Vialard. Geometrical parameters extraction from discrete path. In *Proceedings of the 6th International Conference on Discrete Geometry for Computer Imagery*, LNCS, Lyon, France, 1996. Springer-Verlag.
- [64] J. Vittone and J.M. Chassery. Recognition of digital naive planes and polyhedrization. In *Proceedings of the 9th International Conference on Discrete Geometry for Computer Imagery*, volume 1953 of *LNCS*, pages 296–307, Uppsala, Suède, 2000. Springer-Verlag.
- [65] K. Wall and P. Danielsson. A fast sequential method for polygonal approximation of digitized curves. *Computer Vision, Graphics and Image Processing*, 28(2):220–227, 1984.
- [66] M. Worring and W.M. Smeulders. Digital curvature estimation. *CVGIP: Image Understanding*, 58(3):366–382, 1993.
- [67] W.Y. Wu and M.J.J. Wang. Detecting the dominant points by the curvature-based polygonal approximation. *Graphical Models and Image Processing*, 55(2):79–88, March 1993.
- [68] P.Y. Yin. A tabu search approach to polygonal approximation of digital curves. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(2):243–255, March 2000.