UNIVERSITÉ PARIS VIII - VINCENNES-SAINT DENIS
ÉCOLE DOCTORALE CLI
COGNITION LANGAGE INTERACTION

T H È S E

pour obtenir le titre de

Docteur en Sciences
de l'Université Paris VIII - Vincennes-Saint Denis
Mention : Informatique
Présentée et soutenue par

Yong CHEN

Analyse et interprétation d'images à l'usage des personnes non-voyantes Application à la génération automatique d'images en relief à partir d'équipements banalisés

Thèse dirigée par Jaime Lopez-Krahe
soutenue le 27/11/2015

Rapporteurs :	Damien Olivier	Pr. Université du Havre
	Nadine Baptiste Jessel	MdC HDR Irit Toulouse
Directeur :	Jaime Lopez-Krahe	Pr.Université Paris 8-Vincennes-Saint-Denis
Président :	Dominique Archambault	Pr. Université Paris 8
Examinateurs :	Dominique Archambault	Pr. Université Paris 8
	Franck Lebourgeois	MdC Insa de Lyon
	Gérard Uzan	IGR THIM, Université Paris8

© Yong CHEN, 2015

Dédicaces

À ma très chère Yan pour son soutien quotidien indéfectible et pour tous les sentiments d'affection et d'amour qui représentent pour moi le pilier de tous mes efforts.

À ma très chère mère Xiuzhi et à mon très cher père Yushu qui n'ont cessé de me combler par leur amour et leur tendresse.

À tous les membres de ma famille sans aucune exception.

Et à tous ceux que ma réussite leur tient à cœur.

Remerciements

En premier lieu, je tiens à remercier mon directeur de thèse, monsieur Jaime López Krahe, pour la confiance qu'il m'a accordée en acceptant d'encadrer ce travail doctoral, pour ses multiples conseils et pour toutes les heures qu'il a consacrées à diriger cette recherche.

Merci à Zehira Haddad et Edna Louboungou pour la relecture enrichissante de la partie méthodologie et leurs aides le long du travail.

Merci à Céline Jost pour la relecture globale de la thèse et tous ses conseils et commentaires consciencieux.

Mes remerciements vont également à Gérard Uzan pour le support et les conseils.

Aux personnes qui ont participé dans les expériences le long du travail de recherche. Leur contribution a enrichi énormément ce travail et a apporté des idées pour ceux à venir, et je leur en suis très reconnaissant.

À l'École Doctorale Cognition Langage Interaction, pour avoir supporté les activités de divulgation.

Je tiens également à remercier le laboratoire THIM pour l'accueil et les conditions de travail privilégiées qui m'ont été offertes.

Je remercie profondément les chercheurs qui m'ont fait l'honneur de participer au Jury de soutenance.

Table des matières

Dédicaces iii
Remerciements v
Table des matières vii
Liste des tableaux xi
Liste des figures xiii
Introduction xix

CHAPITRE 1
État de l'art sur les technologies des représentations graphiques pour les malvoyants et les non-voyants 1
1.1 Introduction et définition 2
1.2 La perception 5
1.2.1 La vision 5
1.2.2 Le toucher 7
1.2.3 L'audition 13
1.3 La représentation mentale des non-voyants 14
1.4 Les utilisations des représentations graphiques en tactile 19
1.4.1 Image en relief dans le domaine de l'éducation 19
1.4.2 Cartes tactiles pour l'accessibilité des déficients visuels dans les lieux publics 20
1.5 Les autres technologies d'assistance destinées aux malvoyants et aux non-voyants 25
1.5.1 Lecteur d'écran 26
1.5.2 Reconnaissance automatique de la parole 26
1.5.3 Plage Braille 27

CHAPITRE 2
La conception d'une image en relief 29
2.1 Introduction 30
2.2 Représentation des éléments en relief 35
2.2.1 Les couleurs et la lumière 35
2.2.2 La forme et les contours 36
2.2.3 Le texte 38

2.2.4 La légende 39
2.2.5 Conclusion 40
2.3 Les contraintes de la conception d'image en relief 40
2.3.1 Le trait 40
2.3.2 La surface 41
2.3.3 L'espace entre les éléments 43
2.3.4 La mise en page 44
2.4 Conclusion et proposition de nouvelle méthode de conception 46

CHAPITRE 3
La création automatique d'images en relief avec l'aide des technologies de traitement d'image--traitements bas niveaux 49
3.1 Introduction 50
3.2 La segmentation de l'image 54
3.2.1 Introduction de la segmentation de l'image 55
3.2.2 Méthodes testées 62
3.2.3 Conclusion 70
3.3 Traitement de contours 70
3.3.1 Introduction de la détection de contours 70
3.3.2 Méthodes testées 73
3.3.3 Conclusion 81
3.4 Conclusion 82

CHAPITRE 4
La création automatique d'images en relief avec l'aide des technologies de traitement d'image--traitements haut niveaux 83
4.1 Introduction 84
4.2 Étape de la reconnaissance de formes 84
4.2.1 Introduction de la reconnaissance de formes 84
4.2.2 Méthodes testées 88
4.2.3 Conclusion 104
4.3 Étape de détection de flèches 105
4.3.1 Introduction et problématique 105

4.3.2 Méthodes testées 107
4.4 Étape de la reconnaissance de texte 117
4.4.1 Introduction de la ROC 117
4.4.2 Méthodes testées 119
4.4.3 Conclusion 132
4.5 Conclusion 132

CHAPITRE 5
Développement de l'interface et analyses expérimentales 133
5.1 Introduction 134
5.2 Interface réalisée 134
5.2.1 Étape du prétraitement d'image 135
5.2.2 Étape de validation 137
5.2.3 Étape de segmentation 137
5.2.4 Étape de reconnaissance de formes 140
5.2.5 Étape de la transcription du texte en Braille 140
5.3 Image en relief imprimée 149
5.4 Test avec les utilisateurs finaux et résultats expérimentaux 150
5.4.1 Protocole 150
5.4.2 Procédure expérimentale 155
5.4.3 Résultats et perspectives 169

CHAPITRE 6
Discussion et conclusion 173
6.1 Avantages de la méthode proposée 174
6.2 Du point de vue d'accessibilité du document numérique 176

Bibliographie 179
Liste des publications 191
Annexe A
Codage Braille 193
Annexe B
Images testées en relief 195

Annexe C
Image astronomique réalisée en relief 225

Liste des tableaux

Tableau 2.1. La valeur minimale de chaque cas de distance de séparation. 44
Tableau 4.1. Les différentes formes et leur transformation en Hough 93
Tableau 4.2. Les moments invariants des ellipses différentes 97
Tableau 4.3. Règle de conduite automobile à l'approche d'un carrefour contrôlé par des feux tricolores. 98
Tableau 4.4. Caractéristiques utilisées par les formes. 101
Tableau 4.5. Valeurs des paramètres du vecteur caractéristique correspondant à différentes formes. 102
Tableau 4.6. Comparaison des invariances de certaines mesures de corrélation..... 114
Tableau 5.1. Comparaison des moteurs ROC 147
Tableau 5.2. Tableau de la distribution..... 151
Tableau 5.3. Informations personnelles 156
Tableau 5.4. Carte a (astronomique)..... 157
Tableau 5.5. Carte b (géographique)..... 157
Tableau 5.6. Carte c (histogramme)..... 158
Tableau 5.7. Carte d (camembert)..... 159
Tableau 5.8. Carte e (astronomique)..... 160
Tableau 5.9. Carte f (géographique) 161
Tableau 5.10. Carte g (histogramme) 162
Tableau 5.11. Carte h (camembert)..... 163
Tableau 5.12. Notes et temps d'exploration de chaque image selon les participants. 164
Tableau 5.13. Intervalle de confiance estimé du temps d'exploration de chaque image testée 166
Tableau 5.14. Intervalle de confiance estimé de la note de chaque image testée 167
Tableau A1. Les chiffres..... 193
Tableau A2. Symboles braille codés..... 194

Liste des figures

Figure 1.1. Illusion de couleurs: Tous les cylindres sont de la même couleur! 6
Figure 1.2. Les mouvements de procédure exploratoire manuelle 8
Figure 1.3. Réponse phasique (a) et tonique (b) 9
Figure 1.4. Schéma des récepteurs de la peau 10
Figure 1.5. Homonculus sensitif (à gauche) et Homonculus volontaire (à droite) 11
Figure 1.6. Anatomie de l'oreille 13
Figure 1.7. Le four à thermorelief et la carte en relief 17
Figure 1.8. Exemples d'images tactiles pédagogiques 20
Figure 1.9. Le plan tactile des monuments du patrimoine mondial de l'Unesco de Lyon 21
Figure 1.10. Panneau tactile de La Tour Eiffel 22
Figure 1.11. Tables d'orientation dans les stations de métro parisien 22
Figure 1.12. Carte en relief des réseaux de transports en commun de l'Île-de-France 23
Figure 1.13. Une image thermogonflée avec un système « Tact2Voice » 24
Figure 1.14. Écran tactile avec Feedback-électro-vibration 25
Figure 1.15. Une plage Braille 27
Figure 2.1. Procédure manuelle de réalisation d'une image tactile. 30
Figure 2.2. Carte gaufrée des îles Britanniques 31
Figure 2.3. Déroulement d'un cycle de thermoformage 32
Figure 2.4. Un dessin thermogonflé « Saint Sébastien soigné par Irène » 33
Figure 2.5. Les couleurs sont représentées par les trames 36
Figure 2.6. Perspective à 2 points de fuite 37
Figure 2.7. Une simplification de contour. 38
Figure 2.8. Étiquette en Braille d'une phrase et son explication en Braille dans la légende. 39
Figure 2.9. Un exemple de légende 39
Figure 2.10. Profil caractéristique d'un trait. 41
Figure 2.11. Les paramètres de la texture 42
Figure 2.12. Exemple de dessin avec les différents cas de blanc de séparation. 43
Figure 2.13. Évaluation globale du temps et de la complexité de lecture à l'aide de la valeur de l'empan 45

Figure 3.1. Carte en relief créée par Wang 51
Figure 3.2. Carte en relief créée par Kostopoulos 51
Figure 3.3. Image segmentée et phrase segmentée 52
Figure 3.4. Description globale de l'image 53
Figure 3.5. Description partielle de l'image 53
Figure 3.6. L'étape de décomposition 57
Figure 3.7. L'étape de fusion 58
Figure 3.8. Les points de départ. 59
Figure 3.9. Croissance progressive d'une région. 60
Figure 3.10. Image en binaire conservée par un seuillage de l'histogramme. 61
Figure 3.11. Les principales étapes de K-Moyennes. 64
Figure 3.12. Image testée (Photo paysage) 65
Figure 3.13. Image testée (carte fuseau horaire) 65
Figure 3.14. Résultats de segmentation par l'algorithme K-moyennes (Photo paysage) 66
Figure 3.15. Résultats de segmentation par l'algorithme K-moyennes (carte fuseau horaire) 66
Figure 3.16. Les principales étapes de *C-Moyens flou.* 68
Figure 3.17. Résultats de segmentation par l'algorithme FCM (Photo paysage) 69
Figure 3.18. Résultats de segmentation par l'algorithme FCM (carte fuseau horaire) 69
Figure 3.19. Quelques modèles de contours 71
Figure 3.20. Le détecteur Laplacien 71
Figure 3.21. Image testée «Lena» 74
Figure 3.22. Image testée (carte géographique-fuseau horaire) 75
Figure 3.23. Contours détectés par l'opérateur de Roberts «Lena» 75
Figure 3.24. Contours détectés par l'opérateur de Roberts (carte géographique-fuseau horaire) 76
Figure 3.25. Contours détectés par l'opérateur de Sobel «Lena» 77
Figure 3.26. Contours détectés par l'opérateur de Sobel (carte géographique-fuseau horaire) 77
Figure 3.27. Contours détectés par l'opérateur de Canny «Lena» 80
Figure 3.28. Contours détectés par l'opérateur de Canny (carte géographique-fuseau horaire) 81

Figure 4.1. Deux classes de rectangle : rectangle carré et non-carré avec deux caractéristiques. 85
Figure 4.2. L'espace de caractéristique de rectangle 86
Figure 4.3. La procédure générale de la reconnaissance de forme statistique. 86
Figure 4.4. Une chaise est décrite par l'approche structurelle 87
Figure 4.5. La procédure de la reconnaissance de formes structurelles. 87
Figure 4.6. Une ligne droite dans le plan d'image 88
Figure 4.7. Transformation de Hough 89
Figure 4.8. Paramétrage polaire d'une droite 90
Figure 4.9. Transformation de Hough (paramétrage polaire) 91
Figure 4.10. Les formes géométriques testées 92
Figure 4.11. Les 3 ellipses testées 96
Figure 4.12. Fonctions d'appartenance. 99
Figure 4.13. Modèle polygone pour estimer les caractéristiques 101
Figure 4.14. Représentation des sommets de différentes formes. 103
Figure 4.15. Différence entre les surfaces concaves et convexes de différentes formes. 103
Figure 4.16. Représentation de l'enveloppe convexe et de l'enveloppe englobante. 103
Figure 4.17. Les résultats de détection des formes avec la méthode de la logique floue. 104
Figure 4.18. Images comprenant des petits symboles de flèches 106
Figure 4.19. Exemples de types de points d'intérêt. 107
Figure 4.20. Image de template utilisée pour l'image '4 saisons' 108
Figure 4.21. Image de template de vérification. 109
Figure 4.22. Points d'intérêt trouvés sur l'image originale et celles des templates. 109
Figure 4.23. Exemple 1: Résultat de l'appariement 110
Figure 4.24. Exemple 2: Résultat de l'appariement 111
Figure 4.25. Corrélation entre le template $I_1(m \times n)$ et l'image originale $I_2(M \times N)$. 112
Figure 4.26. Les flèches détectées avec la corrélation. 115
Figure 4.27. Autre exemple de détection de flèches avec la corrélation. 115
Figure 4.28. Exemple 1 d'une image tactile contenant des flèches accessibles. 116
Figure 4.29. Exemple 2 d'une image tactile contenant des flèches accessibles. 116
Figure 4.30. Compression d'image par la transformée en ondelette. 120

Figure 4.31. Débruitage d'une image par la transformée en ondelette 121
Figure 4.32. Histogramme de la zone de texte et la zone de non texte 122
Figure 4.33. Carte de l'Asie 123
Figure 4.34. Explication des phases de la lune 123
Figure 4.35. Les résultats de la localisation du texte par l'ondelette 124
Figure 4.36. Détection du texte par le RLSA 126
Figure 4.37. Les résultats de la localisation du texte par RLSA 127
Figure 4.38. Segmentation d'image avec SWT 130
Figure 4.39. La construction de la zone du texte. 130
Figure 4.40. Les résultats de la localisation du texte par SWT 131
Figure 5.1. Conception d'une interface simple 134
Figure 5.2. Ajustement des images 135
Figure 5.3. Binarisation de l'image 136
Figure 5.4. Segmentation de l'image 138
Figure 5.5. Image avec les textures adaptées 139
Figure 5.6. Carte de légende. 139
Figure 5.7. Les formes dans une classe 140
Figure 5.8. Localisation des zones du texte 141
Figure 5.9. Résultats de la ROC 141
Figure 5.10. Étape de correction 142
Figure 5.11. Correction d'orthographe 143
Figure 5.12. Correction directe manuelle 143
Figure 5.13. Différentes fonctions proposées sur le clic droit de la souris 144
Figure 5.14. Image avec les textes corrigés et sa version en relief avec la légende en Braille 145
Figure 5.15. Positions correctes des étiquettes 148
Figure 5.16. Positions incorrectes des étiquettes 148
Figure 5.17. Impression de l'image en relief. 149
Figure 5.18. Note obtenue par chaque participant et la note moyenne de chaque image 165
Figure 5.19. Temps d'exploration moyen de chaque carte et intervalle de confiance estimé 168
Figure 5.20. Note moyenne de chaque image et intervalle de confiance estimé 168

Figure A1. Les six points d'une cellule Braille 193
Figure B1. Image astronomique (Carte a)..... 195
Figure B2. Carte géographique (Carte b)..... 198
Figure B3. Image histogramme (Carte c) 202
Figure B4. Image camembert (Carte d) 206
Figure B5. Image astronomique (Carte e)..... 209
Figure B6. Carte géographique (Carte f) 213
Figure B7. Image histogramme (Carte g) 217
Figure B8. Image camembert (Carte h) 221
Figure C1. Image astronomique réalisée en relief (phase de la lune)..... 225

Introduction

L'information visuelle est une source d'information très riche à laquelle les non-voyants et malvoyants (ou Personnes Aveugles et Malvoyantes : PAM) n'ont pas toujours accès. En effet, malgré les efforts fournis pour inciter au développement de solutions spécifiques à cette problématique, les avancées observées restent assez limitées. Prenons, par exemple, le cas de l'accès aux documents. Il existe depuis longtemps un système de transcription Braille bien connu qui permet de transformer les documents « texte » en une source d'information tactile pouvant être utilisée par les PAM. Des adaptations ont été effectuées pour la transcription des mathématiques, de la musique [**Jessel 2002**], etc. Toutefois, il est rare de trouver des documents qui ne contiennent que du texte pur. Dans la majorité des documents (qu'ils soient pédagogiques, descriptifs ou autre), on trouve de plus en plus souvent des images, des figures ou des graphiques qui viennent illustrer ou compléter le texte existant. La présence de ces images constitue un réel handicap pour les PAM. C'est précisément à cette problématique que nous nous intéressons dans le cadre de ce travail.

Jusqu'à présent, la transcription des images se fait à la main, ce qui nécessite beaucoup de temps et de moyens. Ajoutons à cela la présence de différentes contraintes qui sont relatives au toucher. En effet, contrairement à la personne voyante, qui peut distinguer facilement un nombre assez riche de couleurs et d'informations visuelles, le non-voyant ne peut distinguer au toucher, qu'un nombre assez limité de textures. Le fait de passer d'une zone à une autre (au niveau texturant) n'est pas aussi évident pour un non-voyant qu'il ne l'est visuellement pour une personne voyante. De ce fait, les contours des différentes zones doivent obligatoirement être renforcés pour une meilleure exploration et compréhension par les PAM. De plus, étant donné la taille fixe de la police des caractères écrits en Braille, on est contraint dans la majorité des cas à établir une légende indépendante de l'image et à introduire une étiquette dans l'image pour faire référence à la légende. Toutes ces contraintes spécifiques à la transcription tactile des documents sont généralement prises en compte pour une transcription manuelle d'images en cartes tactiles.

Le travail présenté dans cette thèse a pour but d'utiliser des méthodes de traitement d'image pour contribuer à l'automatisation du processus de conversion des images en cartes tactiles. Ces dernières seront alors facilement manipulables par les non-voyants et malvoyants. Ce processus repose sur trois étapes importantes. Premièrement une segmentation est effectuée pour extraire les différentes régions, extraire et simplifier leurs contours, puis attribuer les textures adaptées. Puis, pour chaque élément de chaque région extraite, nous proposons une étape de reconnaissance de formes avec un vecteur défini. À la fin, une étape de reconnaissance de texte est utilisée pour reconnaître automatiquement les textes et les transcrire en texte Braille. Finalement, la procédure proposée ici vise également à ouvrir, à terme, l'accès direct à une grande partie des documents numériques qu'on peut trouver sur Internet.

Une partie de ces travaux a bénéficié d'une aide de deux projets: le projet OZALID et le projet CARTASAM. Le premier dans le cadre d'un appel à projets FUI "R&D collaborative", labellisé par le pôle Cap Digital le 28 avril 2011 et le deuxième concerne un appel à projet PICRI (Partenariats institutions-citoyens pour la recherche et l'innovation) 2013 financé par la région ile de France.

Ce manuscrit est organisé en six chapitres :

Le chapitre 1 présente le contexte de l'accessibilité des images et du texte pour les PAM. Un état de l'art sur les technologies de représentation graphique est d'abord présenté suivi d'une description de l'objectif de ces travaux de thèse qui est la conception d'une nouvelle procédure de transcription automatique des images en documents tactiles.

Le chapitre 2 identifie la problématique en s'intéressant au processus de conception d'une image en relief ainsi qu'aux recommandations préconisées pour conception..

Le chapitre 3 présente les différentes méthodes de traitements d'image numérique appliquées pour la segmentation et la détection des contours dans une image. Tandis que le chapitre 4 présente les méthodes dédiées à la reconnaissance de formes et à la reconnaissance du texte.

Dans le chapitre 5, nous décrivons l'interface réalisée avec l'implémentation des méthodes retenues pour chaque tâche. Puis, nous présentons l'étude expérimentale menée avec les non-voyants pour l'évaluation des images créées par notre approche.

Le chapitre 6 présente la discussion et une réflexion sur les possibilités d'applicabilité du travail présenté ainsi qu'une conclusion et des perspectives.

CHAPITRE 1

État de l'art sur les technologies des représentations graphiques pour les malvoyants et les non-voyants

Sommaire :

1.1 Introduction et définition 2
1.2 La perception 5
1.2.1 La vision 5
1.2.2 Le toucher 7
1.2.3 L'audition 13
1.3 La représentation mentale des non-voyants 14
1.4 Les utilisations des représentations graphiques en tactile 19
1.4.1 Image en relief dans le domaine de l'éducation 19
1.4.2 Cartes tactiles pour l'accessibilité des déficients visuels dans les lieux publics 20
1.5 Les autres technologies d'assistance destinées aux malvoyants et aux non-voyants 25
1.5.1 Lecteur d'écran 26
1.5.2 Reconnaissance automatique de la parole 26
1.5.3 Plage Braille 27

1.1 Introduction et définition

« Qu'est-ce donc qu'un aveugle ? S'il en est qui ne bénéficient pas de la moindre perception lumineuse, d'autres perçoivent encore la lumière, les ombres, voire les couleurs, pour peu qu'elles soient fortement éclairées, mais ne parviennent pas à distinguer les formes ; il en est aussi qui, parvenant à distinguer les formes, ne peuvent pas les localiser nettement. Ce n'est pas tout : chez certains, l'acuité visuelle, relativement utile le matin, devient quasi nulle le soir ; il en est qui, pourvus d'une capacité suffisante de vision, la voient compromise par des affections telles que le nystagmus (ou oscillations régulières des yeux) ; il en est encore qui, pourvus d'une vision centrale satisfaisante, ont un champ visuel périphérique réduit. » [**Nino 1960**]

Le mot « cécité » vient du latin *caecitas*, qui représente l'état d'une personne aveugle. Et le mot « aveugle » est une déformation du latin *ab oculis* qui signifie « sans œil ».

Donc un aveugle, c'est la personne qui est privée de ses yeux ou qui est privée de vue. Dans le sens règlementaire français, la cécité commence dès que l'acuité est inférieure à 1/20 [**Berveiller 1991**]. De même, si une personne a en dessous de 3/10 d'acuité visuelle du meilleur œil après la correction, elle est considérée comme malvoyante, ce seuil varie en fonction du pays

À partir des années 1970, au lieu d'utiliser le mot aveugle, le terme « non-voyant » est utilisé de plus en plus souvent considérant que le mot « aveugle » a une connotation péjorative dans la mesure où il tend à englober l'ensemble de la personne alors que seule la vue est touchée. Actuellement, on parle plus volontiers de personnes atteintes de cécité ou de non-voyants.

Le terme de cécité est utilisé aussi pour décrire la condition neurologique qui détruit partiellement ou totalement la fonction des aires visuelles du cerveau. La plupart des aveugles utilisent plus profondément leurs autres sens comme le toucher par exemple.

Grâce à l'alphabet Braille (l'assemblage de points en relief, voir Annexe A), un brailliste (une personne qui sait lire le Braille) peut déchiffrer les lettres, les chiffres avec son toucher. Il existe aussi d'autres sortes d'aide pour les non-voyants et les malvoyants, en particulier grâce aux nouvelles technologies qui permettent d'accéder aux informations écrites à travers la reconnaissance optique de caractères (ROC) et des transcriptions en synthèse de la parole.

On doit distinguer aussi le terme « cécité de naissance » (avant l'âge de 3 ans) et « cécité tardive » (les personnes perdant la vue au cours de la vie) :

1) Cécité de naissance

Les non-voyants de naissance (les aveugles précoces, aveugles congénitaux ou aveugles nés) ont une perception de l'environnement extérieur très différente de celle des autres. Dans leur cerveau, le monde est représenté par les autres sens dont ils disposent : des sons, des odeurs ou bien du toucher. Pour eux certains mots n'ont aucun lien avec le réel, c'est-à-dire que le sens pour eux est inconnu ou incompris (les couleurs par exemple), c'est parce que la vision joue un rôle important dans l'acquisition du langage et peut entraîner des retards dans l'individualisation et la catégorisation des objets et du langage.

2) Cécité tardive

Un non-voyant tardif (ou un aveugle récent) par opposition à le non-voyant de naissance a peut-être eu l'avantage de pouvoir se souvenir ou d'organiser ses connaissances visuelles. Il a déjà bénéficié de la vue, ce qui lui permet plus facilement de découvrir les objets, d'intégrer leur schéma corporel, de se déplacer dans l'espace.

D'après l'organisation mondiale de la santé (OMS), 285 millions de personnes ont une déficience visuelle dans le monde entier, et parmi eux 39 millions sont non-voyants [**WHO 2012**]. On estime à plus de 30 millions les personnes ayant une déficience visuelle

en Europe[1]. Selon une enquête HID (Handicap-Incapacités-Dépendance) [**HID 2003**] réalisée par L'INSEE (Institut National de la Statistique et des Etudes Economiques), parmi une population de 60 millions en 2000, il y a 60 000 non-voyants et près de 1 million de malvoyants (profonds et moyens) en France[2]. La première cause de cécité dans le monde est la cataracte (50% des causes de cécité), les glaucomes en sont la deuxième, suivis du diabète, de la dégénérescence maculaire liée à l'âge (DMLA) et, enfin du trachome.

La déficience visuelle peut réduire largement l'autonomie des personnes, parce que la plupart des informations extérieures sont représentées sous formes visuelles uniquement, et les personnes atteintes d'une déficience visuelle ont des difficultés pour accéder à ces informations. Cette inaccessibilité représente une épreuve importante pour les personnes non-voyantes et malvoyantes en matière de mobilité et d'orientation. Le rapport sur les personnes ayant une déficience visuelle en France [**C2RP 2005**] montre que 56% d'entre eux ont des problèmes de déplacement, ce qui influence grandement leur vie quotidienne et professionnelle. Aujourd'hui, il existe beaucoup de techniques d'assistance pour aider ces personnes à dépasser quelques limitations de leurs vies.

Vanderheiden a présenté dans son travail [**Vanderheiden 2012**] une liste de techniques d'assistance destinées aux personnes non-voyantes. Dans cette liste, l'auteur mentionne différentes modalités comme le Braille, les symboles tactiles, les images en relief et les synthétiseurs de parole. Pour mieux comprendre les difficultés rencontrées par les personnes ayant une déficience visuelle, nous devons d'abord connaître les fonctionnalités, l'importance et les différences de chaque sens chez les non-voyants et les malvoyants.

[1] http://www.euroblind.org/resources/information/#details
[2] http://www.avh.asso.fr/rubriques/association/cecite_malvoyance.php

1.2 La perception

Le système de perception du corps humain contient cinq modalités sensorielles : la vision, le toucher, le goût, l'ouïe et l'odorat. Parmi ces cinq sens, la vision semble être le sens qui apporte le plus d'informations, bien que ces considérations soient très relatives et dépendent de plusieurs critères. L'audition est une perception à distance qui nous permet de localiser les objets, nous localiser et nous orienter dans un espace, mais elle est très limitée parce que tous les objets ne sont pas sonores. Quant au toucher, il joue aussi un rôle important dans l'exploration de l'environnement, les informations manuelles nous permettent aussi de connaître les formes, les grandeurs, les longueurs, les textures, les directions et les orientations à condition que l'espace soit proche de notre corps.

1.2.1 La vision

La vision humaine est l'interprétation cognitive du sens de la vue. Nous utilisons nos yeux pour voir le monde, c'est-à-dire pour recevoir les informations visuelles. La vision nous permet d'avoir une connaissance du monde très rapide et efficace et de contrôler nos mouvements. Elle est le sens le plus important parmi tous les sens humains. D'après Nomura [**Nomura 1990**], la vision apporterait plus de 80% des apports totaux, mais Li [**Li 2005**] la limite à 70%. Elle possède plusieurs facteurs qui lui permettent d'être le meilleur sens pour obtenir les informations spatiales [**Foulke 1992**] [**Milar 1994**] [**Thinus-Blanc 1997**]. Elle a une plus grande résolution spatiale que les autres sens et est aussi mieux adaptée pour coordonner le mouvement dans l'espace [**Cattaneo 2011**]. Chaque œil prend une image en deux dimensions et avec une coopération des deux yeux et du cerveau, il est possible de créer une image mentale en trois dimensions en ayant une estimation de la distance et de la profondeur [**Hinton 1993**].

Le système visuel donne une perception de l'espace, de la direction, de la distance, etc. pour nous aider à nous déplacer dans un environnement. Les informations visuelles prennent les rôles essentiels dans la conception du schéma mental de l'espace chez les voyants. Mais pour les non-voyants, par manque de vision, la mobilité et l'orientation sont largement limitées [**Lahav 2003**]. « La vision permet, en une fraction de seconde et sans

mouvement exploratoire spécifique, de percevoir en même temps la forme, la taille, la texture, la couleur »[**Hatwell 1986**].

Grâce à la vision stéréoscopique, la perception du relief peut être effectuée par les yeux, mais on continue à avoir cette perception de profondeur quand on ferme un œil, d'ailleurs, certains oiseaux ont une très bonne vision de la profondeur, mais ils n'ont pas de vision binoculaire. Donc le problème est de savoir pourquoi la vision a une très bonne perception en troisième dimension. Depuis longtemps, on suppose que la reconstruction visuelle de la troisième dimension est réalisée en fonction d'une collaboration avec les autres modalités, avec le toucher en particulier [**Berkeley 1944**], ou avec des expériences antérieures «enrichissantes »de la vision [**Brunswik 1950**].

Mais l'interprétation cognitive de la vision peut être trompée malgré sa force dans certaines situations, par exemple dans le cas d'illusions qui concernent la perception de la taille, la couleur (figure 1.1) ou la symétrie. L'équivalence entre illusions visuelles et haptiques a été démontrée [**Gentaz 2008**].

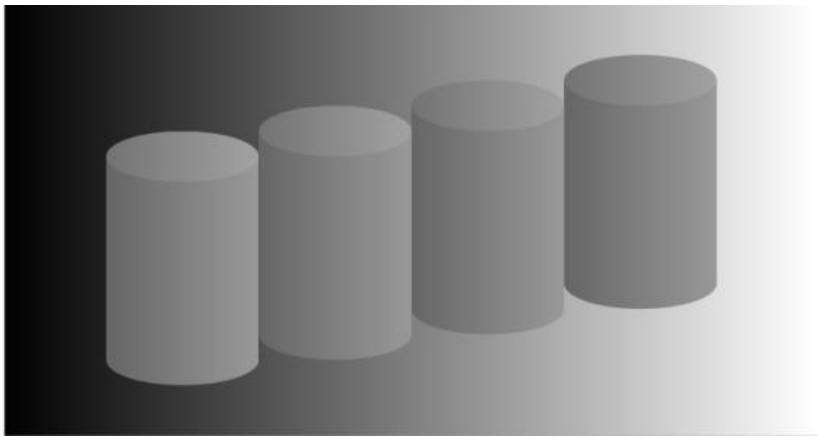


Figure 1.1. Illusion de couleurs: Tous les cylindres sont de la même couleur![1]

[1] https://illusionsoptiques.wordpress.com/plan-du-site/ii-les-differents-types-dillusions/ 15/11/2015

1.2.2 Le toucher

La peau qui enveloppe entièrement le corps est l'organe sensoriel le plus vaste chez l'être humain. Le toucher est le premier sens à se développer chez l'embryon humain [**Montagu 1979**].

« La main est une structure osseuse et articulatoire très ancienne. Le développement qu'elle a connu chez les primates supérieurs, à mesure que s'effectuait le redressement postural qui la libérait de sa fonction locomotrice, a abouti chez l'homme à une forme si perfectionnée du point de vue anatomique et si performant du point de vue fonctionnel qu'on caractérise parfois l'espèce humaine par l'utilisation qu'elle fait de ses mains pour sa suivie et sa maîtrise de l'environnement. » [**Hatwell 1986**].

En effet, les mains ont deux fonctions majeures : une fonction motrice dominante et une fonction perceptive de connaissance des objets [**Hatwell 2003a**]. Avec la première fonction, les perceptions tactiles nous aident à effectuer les actions. La deuxième fonction, elle, est au service des mouvements exploratoires. Quand les adultes voyants explorent un objet sans le voir, une « procédure exploratoire » [**Lederman 1987**] particulière doit être appliquée pour une perception des propriétés de l'objet : suivre des contours pour connaître la forme et la taille, frictionner la surface pour connaître la texture, presser l'objet pour estimer la dureté, soulever l'objet pour connaître le poids, etc. (figure 1.2). Mais « tous ces mouvements sont intentionnels et, s'ils ne sont pas produits, la propriété à laquelle ils correspondent ne sera pas perçue ou le sera mal. De plus, ils sont incompatibles les uns avec les autres du point de vue moteur…et doivent donc être effectués successivement, ce qui accentue encore la lenteur et le caractère séquentiel du toucher… » [**Hatwell 1986**].

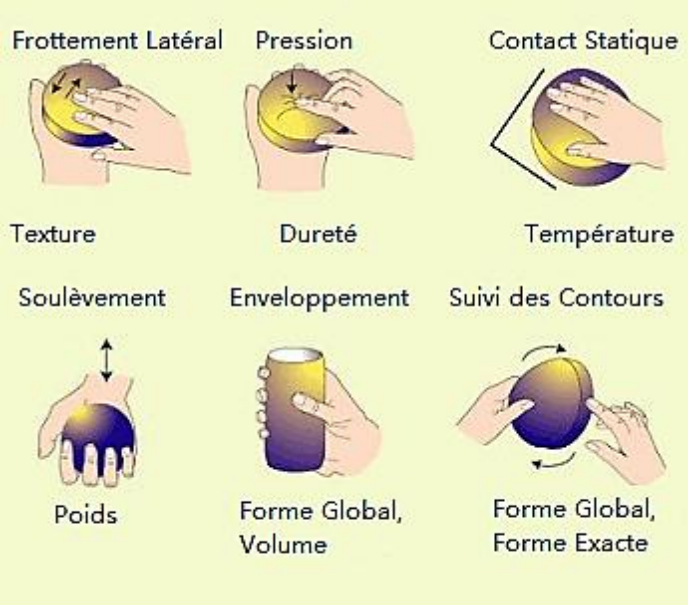


Figure 1.2. Les mouvements de procédure exploratoire manuelle[1] décrits par Lederman et Klatzky [**Lederman 1987**]

La sensibilité tactile est effectuée, d'une part, par les récepteurs de notre peau et d'autre part par les récepteurs situés dans les couches profondes de la musculature et des articulations. Ces derniers nous donnent les informations de la position et du mouvement de chaque partie de notre corps dans un espace proche de notre corps.

Le premier type de récepteur permet de nous donner les perceptions tactiles de l'environnement extérieur. Il existe deux types de touchers : le toucher passif (un contact simple avec un objet) et le toucher actif (un déplacement spatial produit par le mouvement volontaire du sujet) [**Austry 2006**]. Le dernier capte plus d'informations et explore mieux que le toucher passif.

« Dans le cas de certains récepteurs, l'amplitude du potentiel récepteur diminue alors même que le stimulus persiste, un phénomène appelé "adaptation"... Il y a deux

[1] http://www.psy.cmu.edu/~klatzkyfaculty/lab/pictures/index.html 15/12/2015

grandes classes de récepteurs selon la vitesse à laquelle ils s'adaptent: les récepteurs toniques et les récepteurs phasiques. » [**Sherwood 2006**] :

1) Récepteurs toniques (ou récepteurs à adaptation lente)

Les récepteurs toniques donnent une réponse pendant toute la durée de la stimulation (figure 1.3b). Ils renseignent sur la valeur absolue de l'intensité du stimulus et sur sa durée (par exemple, les nocicepteurs).

2) Récepteurs phasiques (ou récepteurs à adaptation rapide)

Ces récepteurs donnent une réponse qui code seulement une partie de la stimulation, ils traduisent les variations du stimulus en fonction du temps (par exemple, les corpuscules tactiles). En réalité, nous avons trois réponses phasiques différentes : la réponse "ON" ; la réponse "OFF" ; la réponse "ON-OFF" (figure 1.3a).

Le mode de fonctionnement "ON-OFF" est fondamental pour la perception du relief, nous devons bouger les doigts pour lire et comprendre les textures et les textes en Braille.

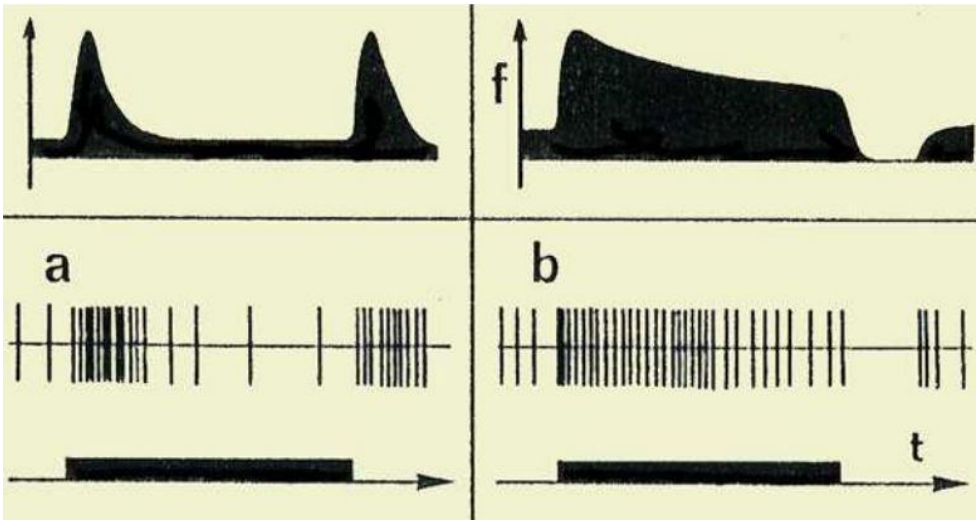


Figure 1.3. Réponse phasique (a) et tonique (b)

(Extraire de [**Buser 1982**])

La peau, un organe plein de sens, est le site principal d'interaction de l'organisme avec son environnement. Elle possède trois types de récepteurs de toucher :

a) Les sensations mécaniques (sensations extéroceptives et sensations proprioceptives) transmettent la sensibilité à la pression, à la vibration ou le tact.

b) Les sensations thermiques pour connaître le froid et le chaud.

c) Les sensations de douleurs pour connaître la fatigue et la destruction tissulaire.

Tous ces types de sensations nous permettent de reconnaître le monde extérieur autour de notre corps et de déplacer notre corps dans un environnement sans information visuelle.

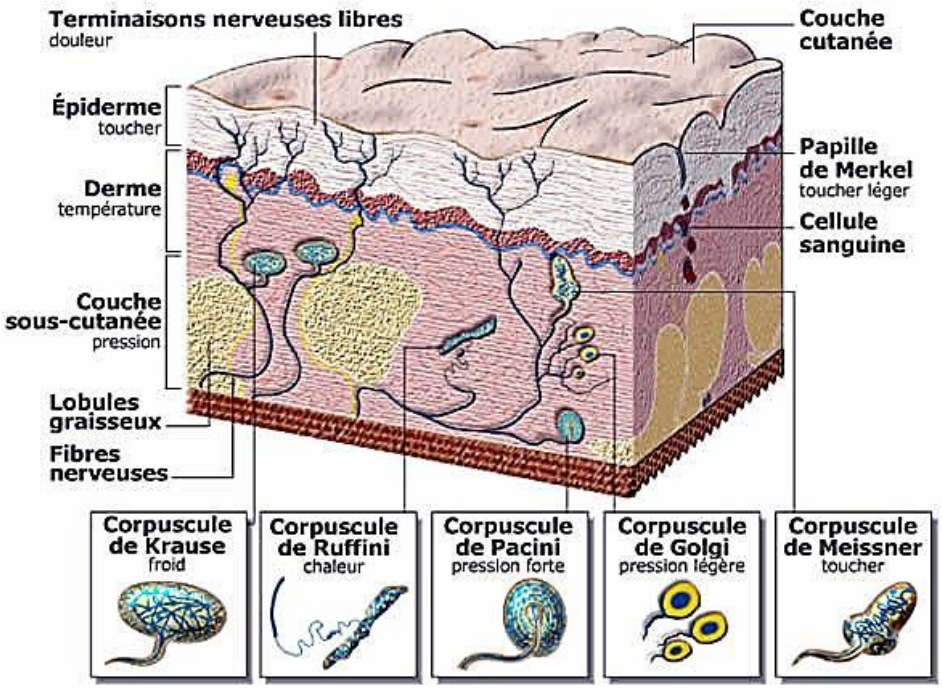


Figure 1.4. Schéma des récepteurs de la peau[1]

[1] http://www.cnrs.fr/cw/dossiers/doschim/imgArt/peau/recepteurs_base.html 15/11/2015

Les différentes parties du corps n'ont pas le même nombre de récepteurs. On trouve le maximum de récepteurs au niveau des lèvres, de la langue, et des doigts (face interne). Au contraire, le nombre de récepteurs en minimal sur l'abdomen et les jambes (voir Homonculus sensitif dans figure 1.5).

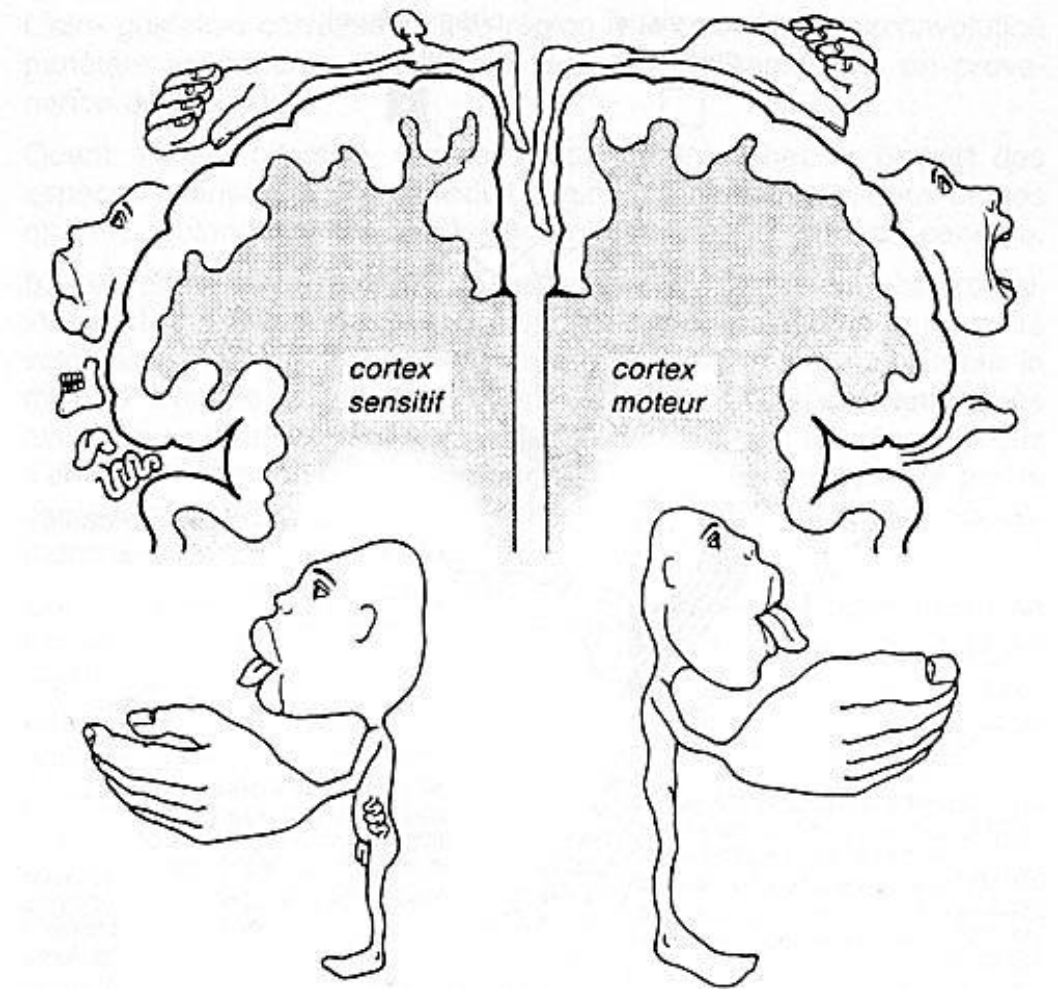


Figure 1.5. Homonculus sensitif (à gauche) et Homonculus volontaire (à droite)[1]

Parmi les doigts, l'index, le majeur et le pouce ont davantage de récepteurs que les autres. Ces doigts ont la sensibilité la plus précise et constituent une sorte de « fovéa

[1] http://www.afblum.be/bioafb/hominisa/hominisa.htm 15/11/2015

tactile » (l'endroit où la densité des récepteurs tactiles est particulièrement élevée). Ils peuvent distinguer les surfaces avec des textures plus fines que les autres.

Le toucher et la vision donnent tous les deux des connaissances spatiales du monde extérieur, nous permettant de connaître certaines caractéristiques des objets. Le champ tactile est plus limité par rapport au champ de vision, car il est limité à la zone de contact. La taille de ce champ varie selon l'utilisation d'un seul doigt, d'une main entière, de deux mains ou de la participation (ou non) des bras et des épaules. Mais certains objets peuvent être considérés comme une extension du corps humain, comme la canne blanche, par exemple [**Gaunet 2005**]. Les performances du système tactile sont plus faibles que celles du système visuel, car les yeux captent les informations visuelles d'un espace extérieur ou d'un objet de manière globale et instantanée. En revanche, la perception tactile est effectuée séquentiellement par un contact direct. En plus, la résolution de la perception tactile est plus limitée par rapport à la vision [**Brock 2015**].

D'autre part, un contact direct donne une capacité de changer l'état physique et spatial des objets, mais la vision a une fonction unique – la perception. La main peut donc aussi accéder à des parties cachées à la vue des objets.

Une exploration tactile est réalisée normalement de manière plus complexe que la vision, chacun des doigts peut avoir une action différente et indépendante. « D'une façon générale, ce qui caractérise ce champ tactile par rapport au champ visuel, c'est à la fois son exiguïté relative et son caractère « élastique » dû à la non-conjugaison des mouvements manuels. Il est en effet rétractable ou expansible à volonté, le sujet pouvant le réduire à la dimension de la pulpe d'un seul index ou l'agrandir en mobilisant toute la face interne des deux mains. » [**Hatwell 1986**]

Le toucher, considéré comme un système perceptif majeur chez les non-voyants malgré ses limitations, leur permet, dans certains domaines, d'avoir des connaissances équivalentes à celles des personnes voyantes. La perception tactile possède une grande

acuité temporelle. Les hommes peuvent distinguer des impulsions successives avec un intervalle de temps aussi petit que 5 ms, alors qu'il est de 25 ms pour la vision [**Choi 2013**].

1.2.3 L'audition

Pour un non-voyant, la perception auditive est plutôt primordiale. L'ouïe, l'un des cinq sens, nous permet de percevoir les informations sonores du monde extérieur. L'oreille (figure 1.6), qui est l'organe principal de l'ouïe, capte les sons, transforme l'information acoustique en signaux nerveux électriques, et finalement transmet ces signaux au cerveau via le nerf auditif.

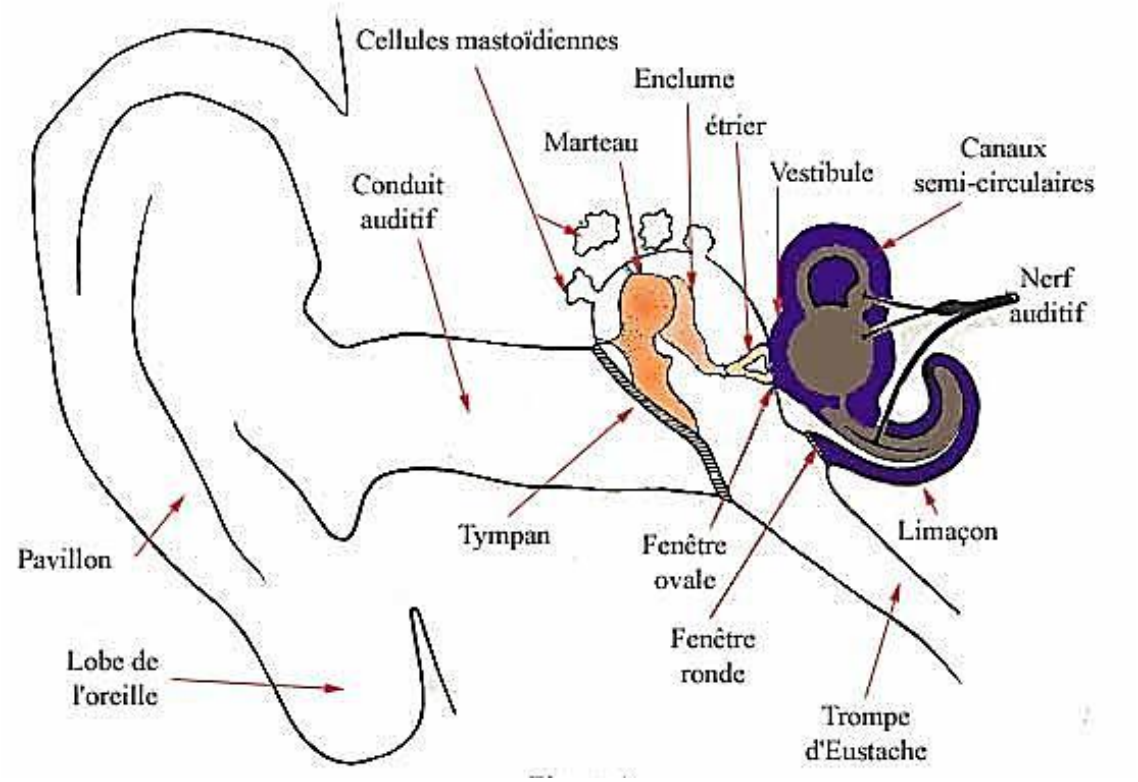


Figure 1.6. Anatomie de l'oreille[1]

[1] http://www.medecine-et-sante.com/anatomie/anatoreille.html 15/11/2015

L'audition permet aussi à la personne de se localiser dans un environnement, mais elle est moins précise que la vision [**Ungar 2000**]. Avec de l'entraînement, une personne non-voyante peut se localiser avec l'aide de l'audition dans un environnement dans lequel il n'y a pas trop de bruits. Grâce au système binauriculaire, la localisation d'une source peut être effectuée par les deux oreilles en estimant sa position relative du corps humain. Morrongiello, Timmey, Humphrey, Anderson et Skory ont fait une expérience très intéressante qui a prouvé que l'ouïe joue un rôle important dans l'élaboration de l'espace [**Morrongiello 1995**].

Le toucher et l'audition sont les deux systèmes réceptifs principaux chez les non-voyants, grâce aux informations tactiles et auditives, les non-voyants peuvent percevoir le monde extérieur.

1.3 La représentation mentale des non-voyants

Avant de parler de la représentation mentale des non-voyants, il est intéressant de présenter "la question de Molyneux". En 1688, William Molyneux a demandé si un non-voyant de naissance qui savait distinguer un cube et une balle avec son toucher, serait capable de distinguer ces objets par la vision une fois qu'il aurait retrouvé la vue? Cette question n'a trouvé sa réponse qu'en 2011. Les travaux de Held et de ses collègues [**Held 2011**] ont montré que les non-voyants de naissance qui ont recouvré la vue après une chirurgie de l'œil n'ont pas réussi immédiatement à reconnaître les formes qu'elles avaient explorées par le toucher auparavant. Mais cette capacité a été réacquise après un très court temps en procédant à un couplage des deux modalités. Aujourd'hui, il est connu que la représentation mentale peut être établie sans avoir d'informations visuelles, mais ce genre de représentation est différent de celle des personnes voyantes.

« Les images mentales non visuelles des non-voyants ont les mêmes propriétés générales fonctionnelles que celles des voyants. Elles sont cependant en partie spécifiques, car le poids des perceptions auditives et tactiles dont elles dérivent y est évidemment plus important que chez les voyants. En conséquence, leur manipulation est plus lente, plus

laborieuse et parfois moins performante que celle des images à dominante visuelle des voyants, surtout lorsqu'il faut traiter une grande quantité d'informations spatiales ou des données typiquement visuelles comme les transformations dues à la perspective » conclue Yvette Hatwell [**Hatwell 2003b**].

Les coordinations entre la vision et le toucher, entre la vision et l'audition jouent des rôles importants dans la reconnaissance du monde extérieur. Pour les non-voyants, en l'absence de vision, le toucher et l'audition sont les deux sens principaux de la manipulation dans un environnement. Un non-voyant de naissance a plus de difficultés pour la perception de l'espace parce qu'il n'a aucune expérience visuelle.

Les adultes humains savent bien saisir un objet qu'ils ont vu et mettre leur main sur cet objet pour avoir une information tactile ou changer son état spatial. Inversement, quand ils manipulent un objet non visible, une reconnaissance sur sa forme, son volume, sa taille peut se développer lorsque cet objet devient visible. « Ces deux types de coordinations intermodales témoignent de l'abstraction de plusieurs informations communes à la vision et au toucher qui apparaissent au cours de développement. » [**Hatwell 2000a**]

À la naissance, quand le nouveau-né ouvre ses yeux, il peut fixer son regard sur un point ou un objet de son environnement, mais il ne peut pas attraper ce qu'il regarde et ne peut pas regarder ce qu'il touche, la coordination entre l'œil et la main est indépendante à ce moment. Ensuite, lorsque la main et l'objet sont tous dans son champ de vision, une coordination limitée s'est produite : c'est-à-dire la main saisit l'objet. Vers 4 mois, le bébé est capable de saisir un objet non visible et de l'amener dans son champ visuel. À partir de 5-6 mois, la vision devient le sens dominant et la coordination entre l'œil et la main est complète : l'œil pilote les mouvements et les gestes manuels, et la main transfère les informations tactiles à la vision.

Pour les non-voyants, manquant d'informations visuelles, un schéma mental de l'espace est construit par les informations captées par les autres sens comme les informations tactiles, auditives, kinesthésiques.

Pour nous localiser dans un espace spécial, nous avons toujours besoin d'un référent. En général, on note qu'il y a deux systèmes de référence : le système de référence égocentré et le système de référence exocentré [**Chellali 2010**].

1. Avec un système de référence égocentré, les actions sont basées sur le corps ou certaines parties de l'opérateur, c'est-à-dire on considère soi-même comme référent.
2. Un système de référence exocentré est un système dans lequel toutes les actions sont basées sur les critères extérieurs.

Chez les non-voyants, la localisation spatiale dans un système égocentré est dominante. Pour eux, le système égocentré est plus accessible que le système exocentré car il manque des perceptions des critères extérieurs. « L'origine des difficultés de représentation spatiale des non-voyants nés tient au fait que le toucher n'est pas adapté à la perception et à la représentation du mouvement. » [**Hatwell 1992**]

Les images mentales d'un non-voyant se construisent à partir des coordinations tactiles liées au toucher ainsi qu'aux mouvements de la main et du corps, les coordinations visuelles et motrices ne jouent plus dans cette procédure. Ces images mentales ne comportent pas les informations visuelles, parce que la vision ne joue pas dans sa construction, mais elles correspondent tout de même à des images spatiales.

Jusqu' à maintenant, nous savons que la représentation spatiale peut être construite sans les expériences visuelles, donc une carte accessible aux non-voyants peut être considérée comme un outil efficace de la cognition spatiale. Les non-voyants peuvent aussi créer une représentation mentale à partir des images accessibles [**Jacobson 1996**]. Une carte représente les informations de l'environnement avec une projection en deux dimensions et une petite échelle. Elle est un dispositif permettant de stocker les informations spatiales par un cartographe [**Lloyd 2000**]. La carte pour les non-voyants est

traditionnellement tactile, les informations sont présentées en relief en utilisant les différentes textures, les symboles et le Braille.

Quand on parle de lecture d'une carte ou d'une image, cela semble être une activité purement visuelle. Très peu de réalisations ont été faites sur l'accessibilité des images pour les non-voyants. En effet, après la création de Braille jusqu'au début des années 1960, il n'y avait pas de moyen efficace et à bas prix pour créer une image en relief. Dans les années 1960, des dispositifs sont apparus sur le marché permettant de tracer un trait en relief ponctué en déplaçant un stylo à bille sur une feuille spéciale posée sur un support en plastique souple. C'est la première fois qu'on a pu créer une image en relief à un prix abordable.

De nos jours, il existe plusieurs méthodes différentes de production d'une image tactile par exemple le gaufrage, le thermoformage et le thermogonflage. Leurs principes sont présentés dans le chapitre suivant. Le dernier utilise une machine spécifique qui peut produire les dessins en relief en utilisant l'effet de chaleur et un papier spécial. (Figure 1.7).

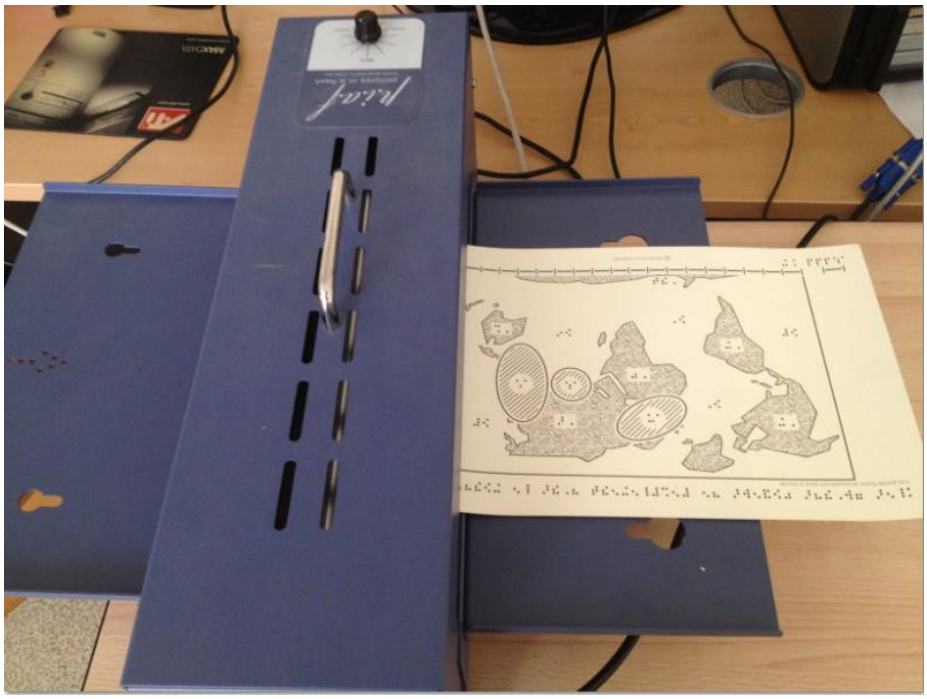


Figure 1.7. Le four à thermorelief et la carte en relief

Le principe de cette technique est le suivant : nous imprimons en noir un document (des textes en Braille, des traits, la surface en texture, les formes, etc.) sur une feuille spéciale qui passe ensuite dans le four à thermoformer, la feuille est chauffée et les parties noires sont gonflées. Cette technique est bien adaptée à notre travail expérimental, cependant, comme nous l'avons expérimenté, sa durée à l'usage est limitée.

Depuis une trentaine d'années, avec ce progrès technique, les cartes, dessins ou diagrammes en relief sont de plus en plus utilisés dans la cadre de la scolarité pour les non-voyants. Mais pour les non-voyants, on ne peut que constater la difficulté à effectuer eux-mêmes des dessins ; puisqu'ils ne peuvent pas contrôler visuellement leur main pour tracer un trait correct. Une autre contrainte est que l'image est une représentation bidimensionnelle d'un objet 3D, pour comprendre ce qu'il représente dans l'image, on a besoin d'une connaissance des lois projectives de 3D vers 2D, mais qui est inconnue par les non-voyants précoces. Pour les voyants, les dessins peuvent être identifiés dès la deuxième année parce que l'objet représenté sur les dessins ressemble visuellement à l'objet réel. En revanche, les dessins tactiles pour les non-voyants ressemblent beaucoup moins aux objets réels, c'est parce que les dessins doivent être modifiés en respectant certaines contraintes (voir Chapitre 2) pour qu'ils puissent être identifiés par les non-voyants via le toucher. En effet, très récemment, la technique d'impression en 3D a été utilisée pour produire des livres tactiles pour les jeunes non-voyants[1]. Elle peut aussi être une méthode possible de la création des images ou des cartes en relief.

[1] http://www.tactilepicturebooks.org/ 15/11/2015

1.4 Les utilisations des représentations graphiques en tactile

1.4.1 Image en relief dans le domaine de l'éducation

Les images en relief sont introduites de plus en plus souvent dans le matériel pédagogique pour non-voyant (figure 1.8). D'après Michel BRIS[1], la nécessité de l'augmentation de l'offre en matière d'image tactile est évidente : un élève non-voyant des cycles de collège et de lycée peut accéder dans le meilleur des cas à moins de 1/1000 des documents imagés dans les seuls manuels. Mais il existe des obstacles qui limitent cette augmentation :

1. La technique : pour l'instant, il n'existe pas de périphériques adaptés pour la création d'images tactiles comme pour la restitution du texte. La création d'une image ne peut se faire qu'après impression sur un support matériel.

2. L'économie : par rapport aux besoins potentiels, le nombre d'images tactiles disponibles est très faible. La tâche est immense. Il est possible d'utiliser les outils informatiques pour créer une image et de définir des éléments de représentation récurrents dans certains registres de documentation graphique. Tout ça peut faciliter et simplifier le travail des transcripteurs.

3. Le statut particulier de l'image tactile : une image tactile doit être conçue ou adaptée aux capacités de discrimination du toucher pour que les contenus soient bien connus par les non-voyants.

[1] Professeur de dessin en relief de l'Institut National Supérieur de formation et de recherche pour l'éducation des jeunes Handicaps et les Enseignements Adaptés (INSHEA)



Figure 1.8. Exemples d'images tactiles pédagogiques
(Petits livres pour enfant, en y collant différentes matières.[1])

1.4.2 Cartes tactiles pour l'accessibilité des déficients visuels dans les lieux publics

À Lyon, l'Office de Tourisme a installé, dans ses locaux Place Bellecour, un plan en relief (figure 1.9) pour accompagner les touristes en situation de handicap visuel dans leur visite du centre-ville. Dans ce plan, les monuments du patrimoine mondial de l'UNESCO et les sites accessibles sont bien indiqués. Il se trouve à l'accueil de l'Office de Tourisme et est accessible à tous les publics.

[1] Images d'enfant aveugle.com, http://www.enfant-aveugle.com/spip.php?article16 15/11/2015



Figure 1.9. Le plan tactile des monuments du patrimoine mondial de l'Unesco de Lyon[1]

La Tour Eiffel est accessible pour les malvoyants grâce aux 4 panneaux tactiles installés au premier étage de la Tour : « la tour de 300 mètres », « La Dame de fer », « Paris au pied de la tour » et « La tour aujourd'hui ». Ces panneaux, créés par Tactile Studio à destination des touristes déficients visuels, sont étudiés pour la lecture tactile. Ils se composent d'images en relief et d'une légende en Braille. En 2014, des nouveaux panneaux tactiles ont remplacé les anciens (figure 1.10).

[1] Images de museumexperts.com, http://www.museumexperts.com/actu_pro/paru/2010/] 15/11/2015



Figure 1.10. Panneau tactile de La Tour Eiffel[1]

La table en relief (figure 1.11), développée en partenariat avec la RATP[2], et avec le soutien du STIF[3], est une carte en relief du quartier autour de la station métro. Installées en 2005, elles sont conçues pour fournir aux non-voyants et malvoyantes des informations intéressantes sur les quartiers autour des sorties de la station.



Figure 1.11. Tables d'orientation dans les stations de métro parisien[4]

[1] Images de Site officiel de la tour Eiffel, http://m.toureiffel.paris/explorer-nouveau-1er-etage 15/11/2015
[2] La Régie autonome des transports parisiens 15/11/2015
[3] Le Syndicat des transports d'Île-de-France 15/11/2015
[4] Photo Michel Lamoireux, http://www.ville-en-mouvement.com/en/content/projet-daide-%C3%A0-la-mobilit%C3%A9-des-aveugles-et-malvoyants-diaporama 15/11/2015

En 2008, les cartes en relief et en grands caractères sur tous les réseaux de transports en commun de l'Île de France sont présentées par la RATP (figure 1.12). Ces cartes ont été conçues pour aider les non-voyants et malvoyantes à planifier facilement leur voyage en visualisant les zones qu'ils veulent visiter.



Figure 1.12. Carte en relief des réseaux de transports en commun de l'Île-de-France[1]

Les images en relief sont largement utilisées comme un outil efficace accessible aux non-voyants, malgré quelques inconvénients. Par exemple, les non-voyants ne sont pas toujours braillistes et le contenu d'une telle carte ne peut pas être adapté de manière dynamique [**Brock 2012**]. Ces cartes tactiles sont donc rapidement obsolètes [**Yatani 2012**]. Il existe aussi d'autres dispositifs pour les non-voyants et les mal voyants. La carte interactive est une façon assez simple et pratique d'avoir accès aux informations pour ces personnes.

Un type de carte interactive combine la carte en relief et de la synthèse vocale [**Brock 2012**], par exemple le Tact2Voice développé par Michel Bris à INSHEA (figure

[1] RATP-Photo Bruno MARGUERITE, http://www.ville-en-mouvement.com/en/content/projet-daide-%C3%A0-la-mobilit%C3%A9-des-aveugles-et-malvoyants-diaporama 15/11/2015

1.13). Les utilisateurs explorent l'image en relief avec leur doigt. En même temps une tablette au-dessous des images donne une explication auditive qui a la même fonctionnalité qu'une légende.



Figure 1.13. Une image thermogonflée avec un système « Tact2Voice »[1]

Une autre technique donne une nouvelle possibilité de présenter une image aux non-voyants, c'est l'écran tactile avec Feedback-électro-vibration (figure 1.14). Olivier Bau et son équipe ont proposé dans leur travail [**Bau 2010**] une nouvelle technique - électro-vibration, qui peut simuler les différentes textures sur l'écran.

[1] Tact2Voice est une application française développée par Michel Bris sur la tablette iPad d'Apple. Son objectif est d'offrir un accompagnement vocal ou sonore, pendant la lecture d'un document en relief, à un utilisateur non-voyant.

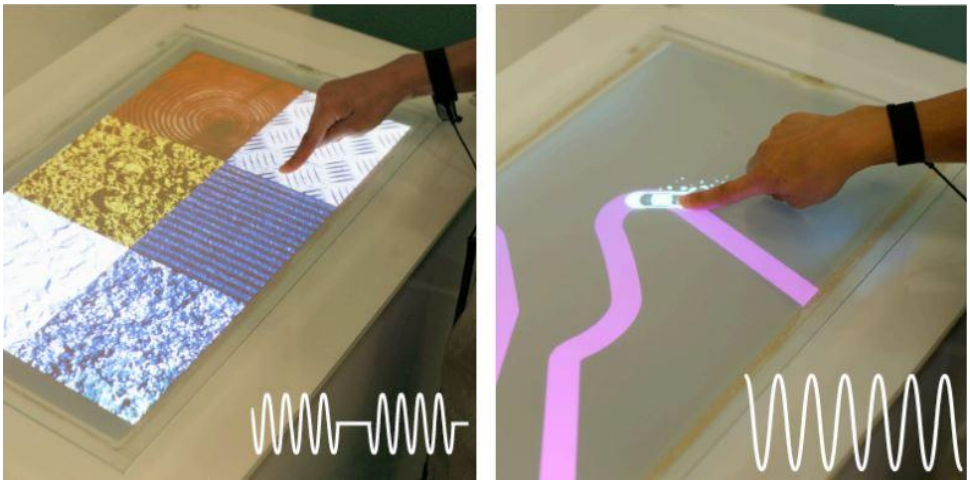


Figure 1.14. Écran tactile avec Feedback-électro-vibration

(Extraite de [**Bau 2010**])

1.5 Les autres technologies d'assistance destinées aux malvoyants et aux non-voyants

Depuis une décennie, grâce au développement rapide et spectaculaire de la micro-information, de nombreuses nouvelles techniques destinées aux personnes handicapées sont apparues sur le marché. Concernant les déficients visuels, ces technologies se rapportent à la synthèse et à la reconnaissance de la parole, aux terminaux à affichage Braille éphémère et aux logiciels de dessin permettant de réaliser des dessins en relief sur papier à microcapsules, sur des imprimantes spécialisées ou à l'aide d'autres techniques numériques.

« Ces technologies sont susceptibles de modifier en profondeur l'enseignement aux handicapés visuels de nombreuses façons :

- En renouvelant l'approche de certains apprentissages traditionnels ;
- En permettant une plus grande autonomie des élèves et des étudiants à l'accès aux informations et aux documents ;
- En facilitant la mise en place de pédagogies différenciées ;
- En induisant la conception de nouveaux supports d'enseignement ;

- En rompant l'isolement des établissements spécialisés et des parents. »

[**Burger 1996**]

1.5.1 Lecteur d'écran

Un lecteur d'écran (ou revue d'écran) est un logiciel destiné aux non-voyants ou fortement malvoyantes. Il transcrit les informations visuelles qui sont affichées sur l'écran d'un ordinateur en informations vocales et/ou tactiles.

Il existe deux principaux modes de restitution de l'information interceptée par le lecteur d'écran :

- la synthèse vocale
- l'afficheur Braille

Le deuxième mode nécessite du matériel supplémentaire (plage Braille, voir 1.5.3), tandis que le premier ne nécessite qu'une sortie audio. Les deux types de conversion de texte peuvent être installés sur un même ordinateur. Mais la synthèse vocale devient le mode le plus utilisé.

1.5.2 Reconnaissance automatique de la parole

La reconnaissance automatique de la parole (ou reconnaissance vocale) est une technique informatique qui permet d'analyser la parole pour la transcrire sous la forme d'un texte exploitable par une machine.

Cette technologie permet de réaliser des interfaces vocales qui ont des interactions se faisant à la voix. Les non-voyants ou malvoyantes peuvent exploiter leur PC avec leur parole et sans souris, ils peuvent aussi dicter un texte sans clavier.

1.5.3 Plage Braille

Une plage Braille est un dispositif électromécanique permettant d'afficher en temps réel des caractères Braille (Figure 1.15).



Figure 1.15. Une plage Braille[1]

Il est aussi possible d'utiliser une plage Braille et une synthèse vocale en même temps. Le non-voyant utilise l'une ou l'autre, ou les deux systèmes simultanément.

Dans ce chapitre, nous avons parlé de la cognition chez les non-voyants, et ensuite montré que la carte tactile peut être considérée comme une façon efficace de représenter l'environnement aux non-voyants. Les images tactiles sont largement utilisées dans beaucoup de domaines pour servir les personnes ayant une déficience visuelle. Dans ce

[1] Photographie prise par Matthieu Faure, http://commons.wikimedia.org/wiki/File:Plage-Braille-Alva.jpg 15/11/2015

travail de thèse, nous allons établir une nouvelle procédure plus facile, rapide et automatique pour créer une image tactile.

CHAPITRE 2

La conception d'une image en relief

Sommaire :

2.1 Introduction 30
2.2 Représentation des éléments en relief 35
2.2.1 Les couleurs et la lumière 35
2.2.2 La forme et les contours 36
2.2.3 Le texte 38
2.2.4 La légende 39
2.2.5 Conclusion 40
2.3 Les contraintes de la conception d'image en relief 40
2.3.1 Le trait 40
2.3.2 La surface 41
2.3.3 L'espace entre les éléments 43
2.3.4 La mise en page 44
2.4 Conclusion et proposition de nouvelle méthode de conception 46

2.1 Introduction

Jusqu'à aujourd'hui, la création d'une image en relief est faite manuellement. Avec un PC, les créateurs utilisent un logiciel de retouche d'image pour redessiner une image à partir de sa version initiale (figure 2.1).

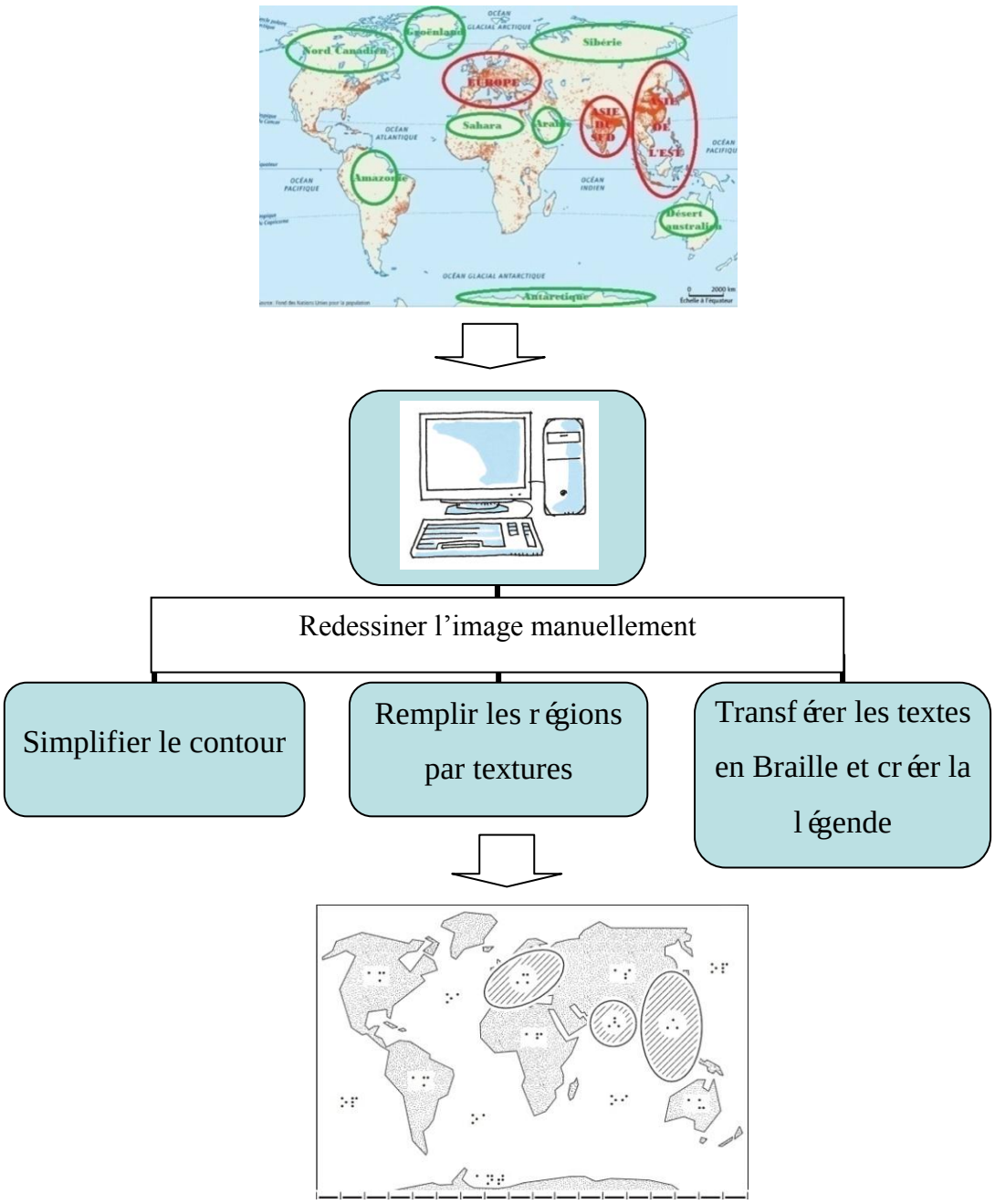


Figure 2.1. Procédure manuelle de réalisation d'une image tactile.
(www.inshea.fr)

Une fois l'image bien redessinée, en cherchant à respecter des critères propres à faciliter une exploration tactile, comment mettre cette image en relief ? Il existe plusieurs procédés, dont les plus connues, le gaufrage, le thermoformage, et le thermogonflage. De nos jours, les deux dernières sont le plus souvent utilisées pour créer une image en relief :

1) Le Gaufrage :

Le gaufrage est une technique d'impression permettant d'obtenir des motifs en relief sur du papier. Cette technique utilise une plaque en bois ou en métal gravée en creux. Cette plaque est fixée sur une table de la presse typographique. Lorsque le tympan (un volet mobile sur lequel on fixe la feuille) est pressé contre la plaque gravée, la feuille qui est mise sous presse entre ces deux éléments se déforme et devient une image en relief en gardant cette déformation (figure 2.2).

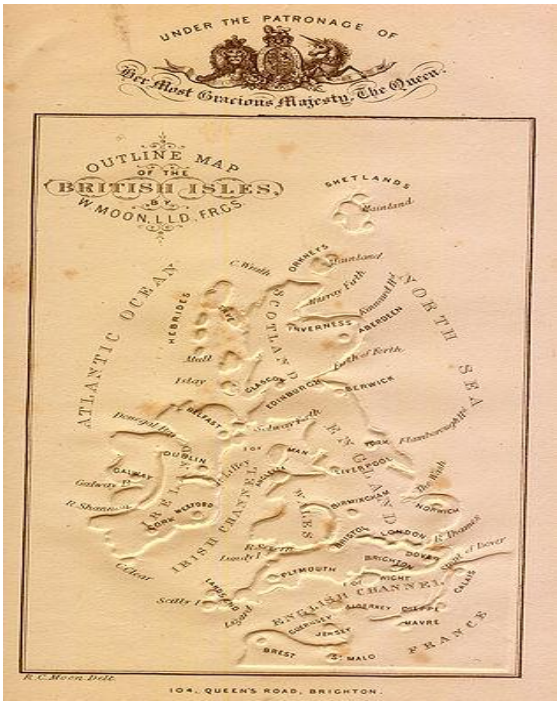


Figure 2.2. Carte gaufrée des îles Britanniques[1]

[1] https://commons.wikimedia.org/wiki/File:Moon-type-map.jpg 15/11/2015

2) Le Thermoformage :

Le thermoformage est un procédé qui transforme des plaques (verre, plastique) : la plaque ductile se déforme quand elle est ramollie par chauffage puis mise en forme grâce à un moule. Lorsque le matériau refroidit, il durcit et garde cette déformation (figure 2.3). Avec cette méthode, nous obtenions une image en relief permanente. Il est possible de créer une image ayant plusieurs niveaux de profondeur qui peuvent augmenter la lisibilité de cette image [**Edman 1992**]. Mais pour cette méthode, l'inconvénient majeur est le coût de la production et des équipements nécessaires qui sont trop chers [**Tatham 1991**].

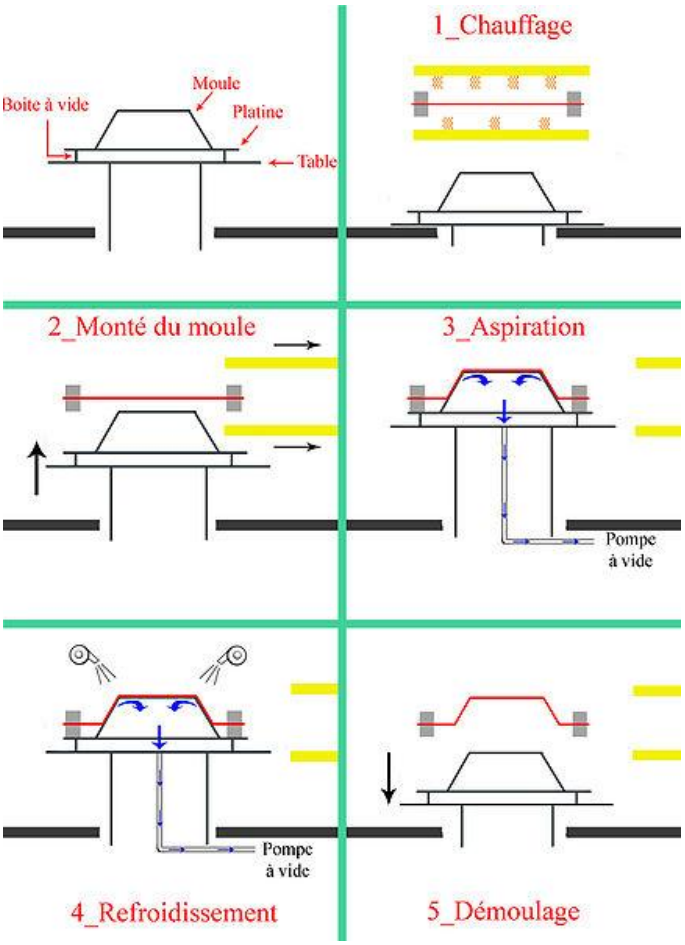


Figure 2.3. Déroulement d'un cycle de thermoformage

(Figure réalisée par la société APTE SA)

3) Le Thermogonflage

Le thermogonflage est une technique qui consiste à faire gonfler une zone de feuille de papier sous l'effet de la chaleur. Ce procédé a besoin d'un photocopieur ou d'une imprimante pour copier ou imprimer un dessin sur un papier spéciale multicouche, recouvert d'une couche de mousse de faible épaisseur composée de microcapsules. Après impression, la feuille est passée au four spécifique et les parties encrées en noir sont gonflées (figure 2.4). Le thermogonflage n'offre que deux niveaux de relief – le fond et le figuré, ce dernier est construit par deux éléments principaux : des traits et des textures de motif. On ne peut pas jouer sur la hauteur des motifs.



Figure 2.4. Un dessin thermogonflé « Saint Sébastien soigné par Irène »[1]

La production des images en relief avec le thermogonflage est plus facile et moins chère (environ un euro par feuillet) que le thermoformage [**Lobben 2012**]. Nous utilisons une imprimante ordinaire. Seul le four coûte un peu cher (environ 1300 euros). Les images peuvent être préparées avec un ordinateur avant d'être imprimées et il est très facile de reproduire les mêmes images. Il est possible de mettre des informations visuelles sur ces images en relief (en gris non gonflable) qui pourront être compréhensibles aussi par les voyants. Il s'agit d'une technique bien adaptée pour des images en faible quantité. En plus, avec cette technique, toutes les personnes peuvent imprimer les images en relief sans difficultés.

[1] Images de Polymorphe Design.com, http://www.polymorphe-design.fr/les-sens-de-l-art.php 15/11/2015

Mais cette technique a un inconvénient important : l'image thermogonflée n'est pas durable. Pendant les tests expérimentaux, nous avons observé qu'une image thermogonflée est de moins en moins lisible après certaines explorations par les non-voyants (environs 3 ou 4 personnes), les textures et les textes en Braille étant usés par les doigts. Dans ce cas, nous devions réimprimer les images. La durabilité de l'image thermogonflée est faible par rapport aux images des autres types (image de thermoformage par exemple) parce que le thermogonflage utilise le papier comme matériau de l'image, alors que les autres techniques utilisent des matériaux plus durables (plastique, métal, etc.). En pratique, nous pouvons choisir les différentes méthodes de fabrication selon les informations à présenter dans les images : pour les informations récentes qui changent rapidement (un planning de la semaine par exemple), le thermogonflage est le meilleur choix ; pour les informations constantes (le plan d'un immeuble par exemple), c'est mieux d'utiliser les matériaux plus durables.

Les cartes et images tactiles sont considérées comme une façon pratique de présenter les informations spatiales aux personnes ayant une déficience visuelle. Il existe cependant des critiques ; une des critiques majeures est la reconstitution manuelle d'image en relief avec un PC qui prend vraiment beaucoup de temps [**Rice 2005**]. Et en plus, pour les créateurs, il faut avoir des connaissances sur les logiciels de retouche d'image. Les autres critiques portent sur les contenus présentés parce que le toucher a des limites de perception : sur les cartes tactiles, le nombre d'informations présentées est moindre en comparaison avec une carte visuelle.

En effet, avec le développement technologique, il est possible d'utiliser les techniques de traitement d'image pour rendre plus automatique et plus rapide la procédure de la reconstitution d'images en relief ; ces techniques sont présentées dans le chapitre 3. Dans ce chapitre, on énonce les règles et les contraintes que l'on doit respecter pour créer une image tactile.

2.2 Représentation des éléments en relief

Une image en relief présente les informations via les traits, les symboles et les textures. Le Braille est utilisé pour annoter les informations tactiles [**Edman 1992**]. Dans le chapitre précédent, nous avons parlé des limitations du sens tactile par rapport à la vision. Pour qu'une image tactile puisse être compréhensible, elle doit être simplifiée et en même temps contenir les informations utiles et importantes [**Hatwell 2000b**].

Dans ce travail, l'image dont on parle, n'est pas une image de type photographique. Ce genre d'image est difficilement accessible aux non-voyants. Les images sur lesquelles nous travaillons sont des images pédagogiques comme les cartes, les images statistiques, les images astronomiques, etc.

Une image contient différents éléments visuels qui présentent différentes informations, par exemple les couleurs, les différentes formes, les tracés, les textes, etc. Quand on transcrit ces éléments visuels en éléments tactiles, il faut respecter certaines règles et contraintes pour avoir une image tactile discernable par le toucher. Dans cette partie, nous parlerons du choix des éléments visuels pour les représenter en relief, et dans la partie suivante nous parlerons des recommandations pour créer une image en relief.

2.2.1 Les couleurs et la lumière

La couleur et la lumière sont purement des informations visuelles. Pour les non-voyants de naissance, ces informations sont incompréhensibles. En effet, dans certaines images, les différentes couleurs représentent les différentes informations importantes; quand on traduit une image couleur ou noire et blanche en image en relief, il faut donc instaurer un code pour les couleurs et la lumière, par exemple, en utilisant un système de trames différentes pour représenter chacune couleur ou distinguer les parties éclairées et non éclairées s'il s'agit d'une image en noir et blanc (figure 2.5). Pour ne pas saturer la représentation mentale d'un non-voyant, le nombre de trames ne doit pas être supérieur à cinq, aussi une explication sur chaque trame doit être ajoutée dans la légende.

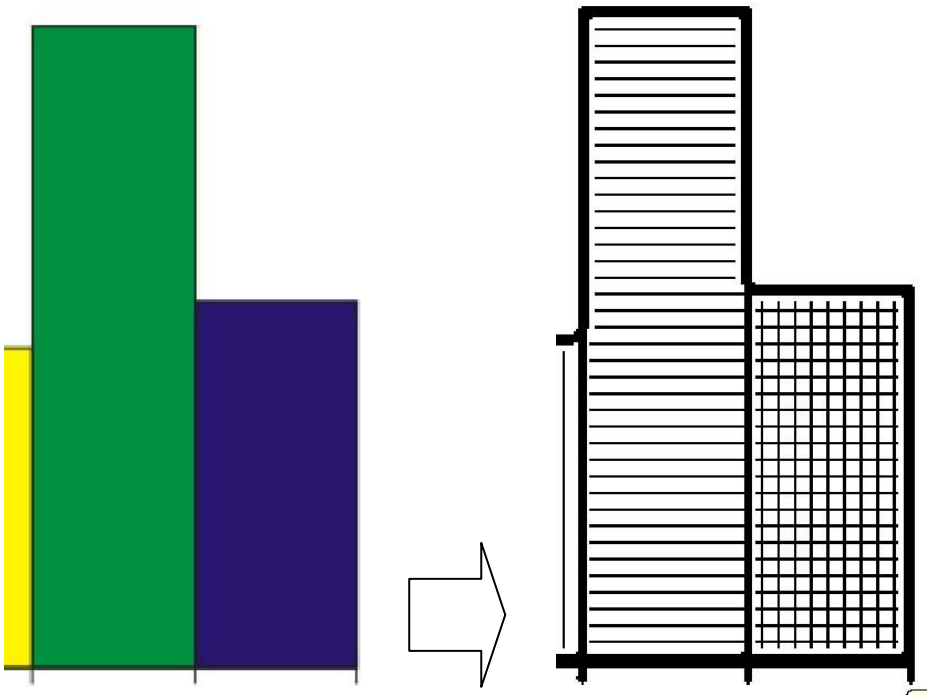


Figure 2.5. Les couleurs sont représentées par les trames

En effet, le surcroît des données rend la lecture d'une image en relief plus délicate, mais dans certaines images (image géographie par exemple), les couleurs sont des éléments importants qu'on ne peut ignorer. Par conséquent, il ne faut pas surcharger l'image tactile et éviter que des informations peu compréhensibles ne viennent « polluer » la perception de l'ensemble.

2.2.2 La forme et les contours

Comme nous avons dit qu'une image est une représentation bidimensionnelle des objets tridimensionnels, il faut avoir une connaissance des lois projectives de 3D vers 2D pour comprendre ce qui est présenté par cette image (figure 2.6). Quand les voyants regardent une image présentant une scène tridimensionnelle, certaines informations visuelles sur l'image comme la lumière et les ombres, la taille relative des objets, la superposition, la perspective linéaire nous permettent d'avoir une connaissance de la profondeur. Notre cerveau va faire la reconstitution des plans et des figures à l'aide de

l'expérience visuelle acquise, et la scène présentée sur l'image est perçue par le cerveau comme une reproduction de la réalité.

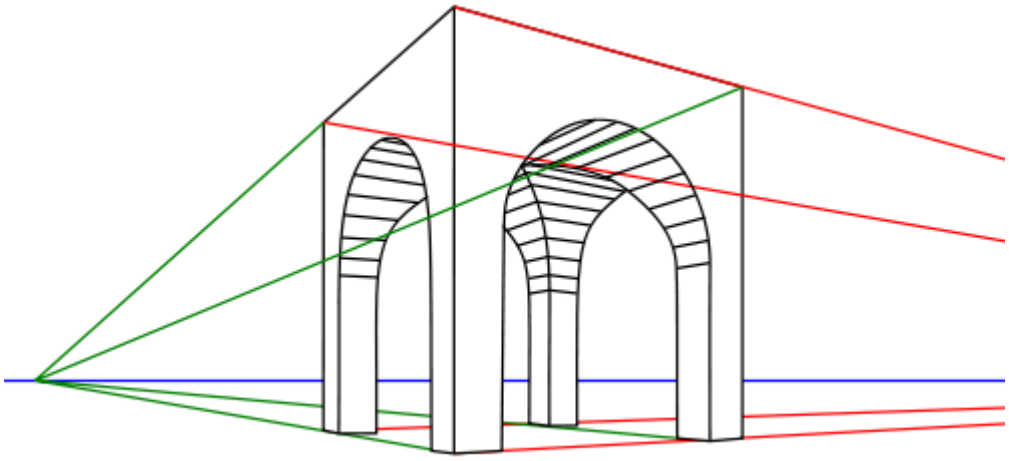


Figure 2.6. Perspective à 2 points de fuite
(Auteur : Wolfram Gothe 2003)

Pour un non-voyant, la loi projective est inconnue, une représentation de profondeur en relief n'a aucun sens pour lui. Par conséquent, quand on transcrit un objet en relief, une simplification de sa forme doit être prise en compte pour qu'il puisse être identifié facilement par les non-voyants. S'il y a un objet qui a un contour très compliqué, on doit aussi simplifier ce contour pour que les non-voyants puissent le lire avec le toucher (figure 2.7).

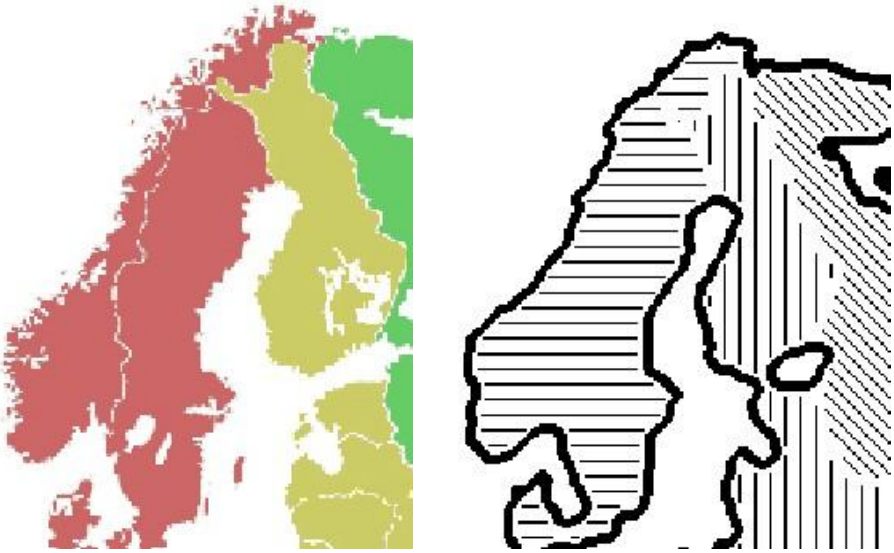


Figure 2.7. Une simplification de contour.

2.2.3 Le texte

Le texte sur une image doit être traduit en Braille. la traduction peut être manuelle, ou bien nous pouvons utiliser la technique ROC et un processus automatique (voir dans le Chapitre 3) pour faciliter ce travail.

Les alphabets Braille sont l'assemblage de points en relief, inventés par Louis Braille en 1824 (voir annexe A). Aujourd'hui, c'est un système standard d'écriture et de lecture pour les non-voyants. Mais le Braille est difficile à apprendre et la vitesse de lecture du Braille dépend fortement de l'âge d'apprentissage. Si le Braille est appris par un non-voyant à son plus jeune âge, la vitesse de lecture est plus rapide [**Tobin 2003**].

Comme le Braille est un système standard d'écriture et de lecture chez les non-voyants, il est aussi une méthode standard pour annoter les images tactiles, mais il n'est pas facile d'introduire les textes en Braille sur une image tactile. Un inconvénient majeur est que les textes en Braille prennent plus de place que les textes écrits et qu'ils vont cacher les autres éléments de l'image [Hinton1993]. Si le texte est long, quand on le transcrit en Braille, nous utilisons une étiquette (par exemple un numéro en Braille) et nous la mettons sur l'image pour remplacer tous les textes, et ensuite nous donnons une explication en Braille dans la légende (figure 2.8).

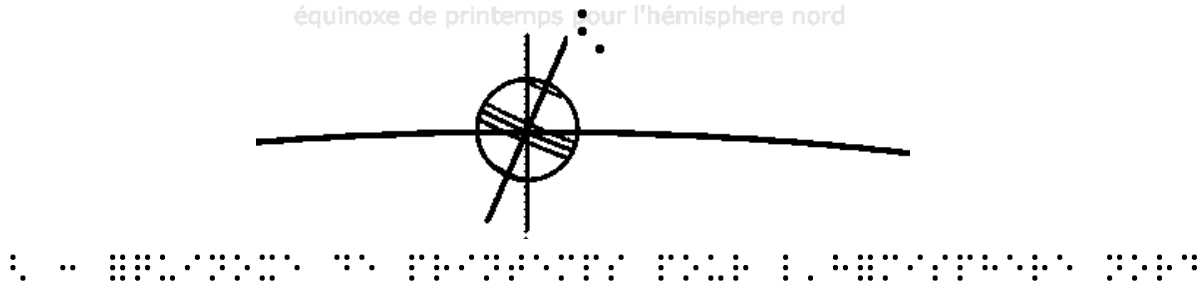


Figure 2.8. Étiquette en Braille d'une phrase et son explication en Braille dans la légende.

2.2.4 La légende

La légende contient normalement les explications supplémentaires sur les éléments présentés par l'image. Quand un non-voyant exploite une image en relief, la légende est nécessairement présentée à côté pour qu'il puisse comprendre ce qu'il est en train de lire par le toucher.

Dans la légende, nous allons mettre les explications sur les formes, les étiquettes, et aussi les textures que nous avons appliquées sur l'image (figure 2.9).

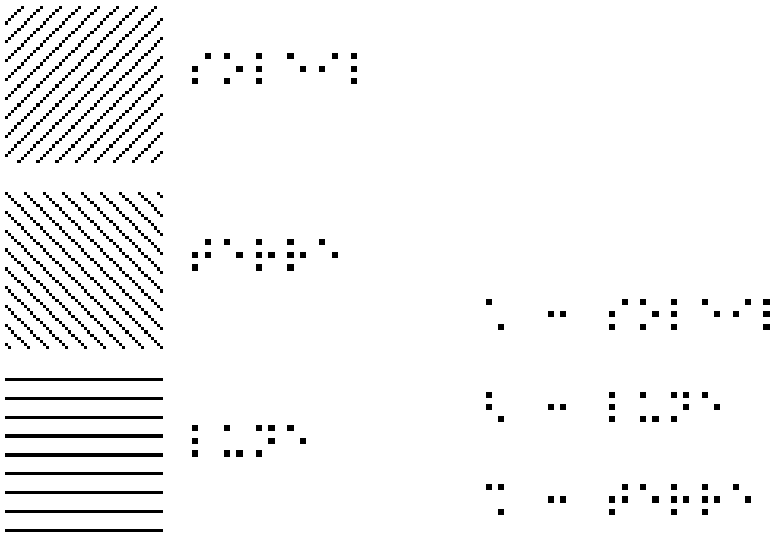


Figure 2.9. Un exemple de légende

2.2.5 Conclusion

Pour créer une image en relief, il faut partir avec l'idée de faire simple : ignorer la perspective par exemple, simplifier certains éléments (les contours ou les formes) et décomposer les images en plusieurs éléments pour faciliter la compréhension. Les différents éléments d'une image originale (les couleurs, les contours, les formes et les textes) doivent être transformés en version tactile (les textures, les traits, les symboles et le Braille), pour que les informations visuelles puissent être compréhensibles par le toucher.

2.3 Les contraintes de la conception d'image en relief

Les cartes et images tactiles utilisent des symboles et des textures différentes ; pour l'instant il n'y a pas de règles strictes concernant la création des cartes tactiles. Sans ces règles, il est plus difficile pour les personnes ayant une déficience visuelle de lire les cartes [**Lobben 2012**]. L'adaptation d'une image aux capacités discriminatoires du toucher demande un travail de traduction important. Heureusement il existe certaines recommandations sur la création des cartes tactiles [**Bris 2004**], et de plus une normalisation des règles est en train d'être établie par les chercheurs. Par exemple, Lobben et Lawrence ont proposé une normalisation des symboles de la carte de ville en relief [**Lobben 2012**]. Dans cette partie on va présenter ces recommandations selon les éléments d'une image tactile.

2.3.1 Le trait

Pour montrer la forme ou la frontière d'un objet ou d'une surface, il faut figurer ses contours. Un trait peut représenter les objets linéaires ou les contours des objets en deux dimensions [**Bris 1999**]. Un contour est perçu comme une succession de déplacements et de changements de direction. « Dans ce cas le trait n'est que le support matériel des propriétés géométriques du contour et doit en permettre l'accès le plus juste. Sa neutralité figurative est donc sa première qualité. » [**Bris 2004**]

Les traits doivent donc être créés en respectant des caractéristiques précises pour avoir une bonne discrimination au toucher (figure 2.10) :

- La hauteur : la valeur de la hauteur du trait ne doit pas être inférieure à 0.4mm.
- L'épaisseur : la valeur minimale de l'épaisseur est 0.4mm.
- La forme : la forme idéale doit être proche de celle d'un trapèze aux sommets arrondis (r=0.1mm)

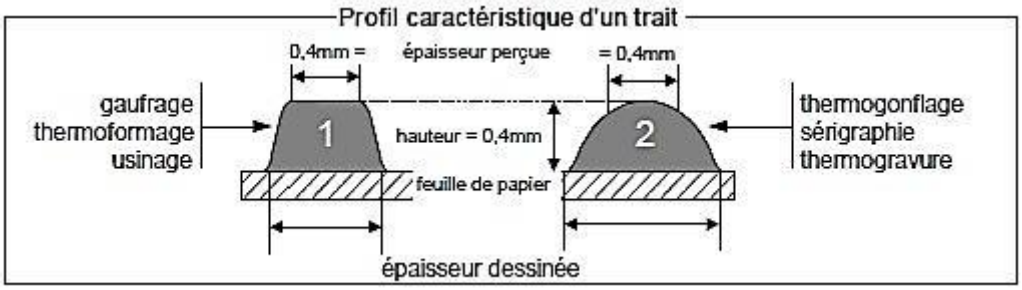


Figure 2.10. Profil caractéristique d'un trait.
(Extraite de [**Bris 2004**])

Pour faciliter la discrimination entre les traits, la variation de l'épaisseur du tracé doit être définie suivant une raison géométrique de 2 : 0,4mm, 0,8mm, 1,6mm, etc. Le nombre maximal de modèles de traits différents doit être entre 3 et 5 [**Bris 2004**].

2.3.2 La surface

Les surfaces sont représentées par les textures en différents motifs. Elles sont particulièrement utilisées pour discriminer l'intérieur et l'extérieur d'un objet, la continuité de cet objet, la séparation de deux objets. Les textures doivent aussi respecter les contraintes du toucher. Elles peuvent être créées soit avec les motifs conventionnels de type point et hachures, soit avec des motifs qui ont valeur de signe, à condition, évidemment, qu'ils soient identifiés en tant que tels.

Les textures sont discriminées par le toucher passif, pour faciliter l'identification de chaque texture, nous devons utiliser des textures qui soient facilement distinguées [**Lederman 1979**]. Il faut donc choisir les valeurs des paramètres attribués aux composants : le motif, l'orientation, le grain (figure 2.11).

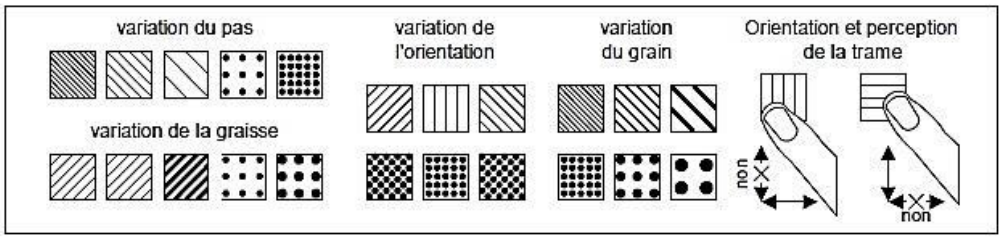


Figure 2.11. Les paramètres de la texture
(Extraite de [**Bris 2004**])

Le motif a deux composants : la forme donnée à l'élément constitutif de base et la structure qui est le mode de répartition de ces éléments dans le plan. Pour les non-voyants, la reconnaissance de ces deux paramètres est très difficile parce que la discrimination du motif fait appel à une grande acuité. Donc, il faut choisir des motifs très différents et très simples, par exemple les points et les traits.

L'orientation du motif peut donner une mauvaise perception. Le doigt glisse sur la texture pour avoir la sensation tactile constituée par les différences de pression. Les mouvements de balayage du doigt les plus économiques sont les mouvements horizontaux, puis les mouvements verticaux. Si l'orientation d'un motif est parallèle à ces mouvements, les différences de sensation sont quasi nulles. Il est donc préférable d'utiliser un motif avec une orientation différente de l'horizontale et de la verticale pour que la texture soit bien lisible. En conséquence, une trame à motif discontinu dans toutes les directions est préférable.

Le pas est la distance entre deux éléments d'un motif (voir figure 2.11). Si le pas est trop grand, la sensation de trame disparaît. Cette variable est la plus pertinente pour une trame à motif constant (le point par exemple). La variation du pas doit être définie avec une raison géométrique minimale de 1,5. Pour une trame de points avec 0,5mm de

diamètre, la variation du pas doit être de 1mm, 1,5mm, 2,3mm, etc. Si les éléments du motif sont très proches, cela risque d'engendrer des amalgames d'encre après gonflage ; dans ce cas, il faut augmenter le pas.

La largeur : si la largeur est trop forte, la trame peut être considérée comme un dessin d'objet. En conséquence si la texture est à l'intérieur d'un tracé, elle doit être plus fine que le tracé pour permettre de bien distinguer la texture et le tracé.

2.3.3 L'espace entre les éléments

« Les graphistes ont l'habitude de dire qu'on lit autant la forme par les vides que par les pleins. Les bons lecteurs de Braille repèrent autant les points Braille que la forme des "blancs" qui séparent les lettres » ceci est expliqué par Michel Bris [**Bris 2004**].

Les blancs de séparation sont toujours appliqués sur les images tactiles pour faciliter la lecture. C'est important de séparer les différents éléments graphiques par l'espace vide pour qu'ils puissent être distingués plus rapidement et plus facilement (figure 2.12) [**Tatham 1991**].

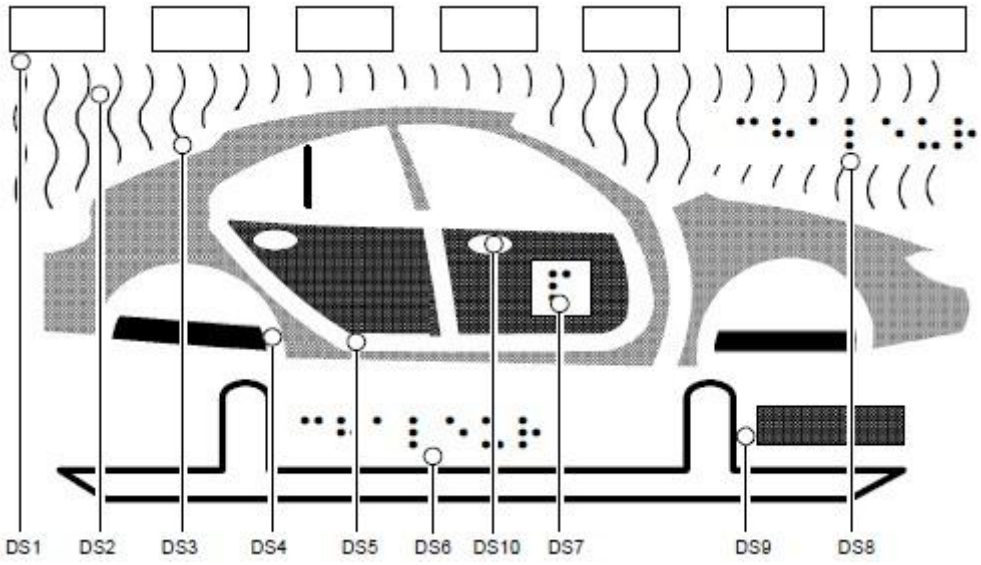


Figure 2.12. Exemple de dessin avec les différents cas de blanc de séparation.
(Extraite de [**Bris 2004**])

Dans le tableau suivant, Michel Bris a donné les recommandations sur chaque cas mentionné dans le dessin ci-dessus :

Tableau 2.1. La valeur minimale de chaque cas de distance de séparation.

(Extrait de [**Bris 2004**])

Situation	Description	Distance minimale
DS1	Intersection de deux traits fins	1 à 2mm
DS2	Deux traits fins	3 à 4mm
DS3	Intersection d'un trait fin avec une surface	4 à 5mm
DS4/DS5	Surface en vis-à-vis	5 à 6mm
DS6/DS7/DS8	Texte vis-à-vis d'un élément tait ou surface	4 à 5mm
DS9	Trait fin parallèle à une surface	4 à 5mm
DS10	Vide minimale	6 à 8mm

2.3.4 La mise en page

Pour avoir une très bonne lisibilité d'une planche, certains critères doivent être respectés. Afin d'obtenir un confort lié à la lisibilité de la lecture tactile, il faut respecter la norme AFNOR : *NF Q 60-502 Technologie graphique-critère de visibilité de la communication graphique*, ainsi que les règles générales qui sont applicables aux produits graphiques destinés à des non-voyants pour avoir l'organisation et la fluidité du document. Il y a quelques éléments particuliers concernant la mise en page d'un document relief :

1) Le format

   Le format, c'est la taille et l'orientation de la page. Pour choisir un bon format, on doit considérer certaines contraintes suivantes :

   - Contraintes économiques et techniques
   - Contraintes environnementales
   - Densité de l'information

- Contraintes ergonomiques de la lecture tactile

Le troisième critère demande de conserver le maximum d'informations ce qui exige d'agrandir les dessins, mais le dernier critère demande d'adapter le coût de la lecture (temps et complexité), ce qui conduit à limiter la taille. Donc il faut d'abord trouver un compromis entre ces deux critères.

La taille et l'orientation du format influent beaucoup sur le temps de lecture d'un document tactile. Ici, plus le format est grand, plus il y a de nombre de « poses » de la main appliquée sur le document (figure 2.13).

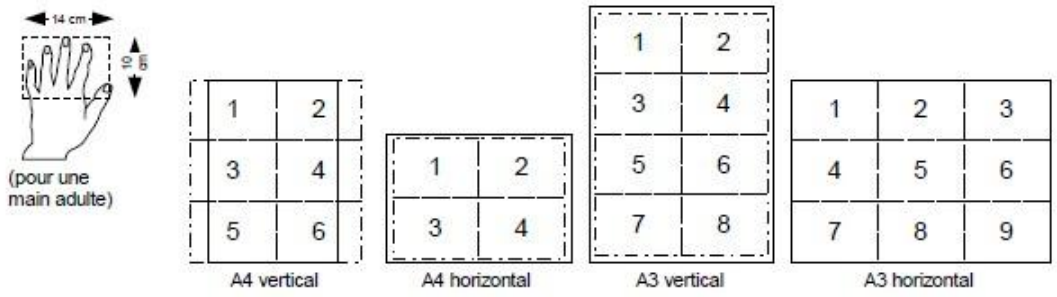


Figure 2.13. Évaluation globale du temps et de la complexité de lecture à l'aide de la valeur de l'empan
(Extraite de [**Bris 2004**])

Michel Bris propose que « le format A3 soit la limite, au-delà de laquelle l'ensemble de l'information ne peut plus être rassemblé en un message unique ». On peut considérer à partir du schéma que le format A4 horizontal est le plus efficace si la planche comporte une grande image et peu de texte. Selon les recherches réalisées par Brock [**Brock 2013**], elle a montré aussi que les formats A3 et A4 sont plus adaptés au toucher pour explorer une image en relief par les utilisateurs non-voyants.

2) les marges

Les marges ont deux fonctions principales :

- Séparer les éléments
- Permettre le repérage des alignements ou des correspondances.

Il ne faut pas mettre d'éléments en bord de page, car cette partie est difficile à lire.

3) le repère

Lorsqu'on lit un document tactile, les mains font un mouvement du haut du document à gauche vers le bas à droite. Un marqueur (grand triangle noir par exemple ou une coupure d'un coin de la feuille) en haut à gauche doit être mis sur tous les documents, permettant ainsi l'orientation de la page.

2.4 Conclusion et proposition de nouvelle méthode de conception

Pour un document tactile, la représentation globale n'est pas une donnée permanente comme pour la vision, elle est construite après une exploration beaucoup plus lente. Afin de rendre la lecture plus facile et plus rapide, les contraintes mentionnées dans ce chapitre doivent être respectées au maximum.

Jusqu'à maintenant, la transformation d'une image dans sa version tactile est réalisée manuellement avec des logiciels de retouche d'image, la durée de ce travail est très variable, entre quelques heures et quelques journées selon la complexité de l'image originale. Les créateurs construisent leur image tactile en appliquant leurs propres règles, les résultats de la transformation d'une même image originale sont donc différents selon les créateurs ; pour un seul créateur, les planches obtenues vont être différentes même après plusieurs essais.

Afin de faciliter et de normaliser cette proc édure, nous proposons de construire une interface op érationnelle sous Windows® pour r éaliser une transformation automatique (ou semi-automatique) en appliquant certaines m éthodes de traitement d'image (segmentation, reconnaissance de formes, ROC, etc.) Dans le chapitre suivant, nous expliquons tous les m éthodes que nous avons test ées et finalement utilis ées.

CHAPITRE 3

La création automatique d'images en relief avec l'aide des technologies de traitement d'image --traitements bas niveaux

Sommaire :

3.1 Introduction 50
3.2 La segmentation de l'image 54
3.2.1 Introduction de la segmentation de l'image 55
3.2.2 Méthodes testées 62
3.2.3 Conclusion 70
3.3 Traitement de contours 70
3.3.1 Introduction de la détection de contours 70
3.3.2 Méthodes testées 73
3.3.3 Conclusion 81
3.4 Conclusion 82

3.1 Introduction

Le traitement de l'image est une discipline de l'informatique qui analyse, modifie ou collecte les informations d'une image numérique. Il s'agit d'une application du traitement du signal destinée soit aux images soit aux données dérivées comme la vidéo.

Le but de notre travail est de réaliser une procédure automatique ou semi-automatique afin de créer plus rapidement et plus facilement une image tactile utile aux personnes non-voyantes en utilisant différentes techniques de traitement d'image. Nous nous limiterons à rendre des images pédagogiques accessibles (image géographique, image statistique, image astronomique, etc.).

Une approche alternative de la création automatique d'une carte en utilisant les méthodes de traitement d'image numérique est proposée dans le travail de Wang [**Wang 2009**] et Kostopoulos [**Kostopoulos 2007**]. Dans le travail de Wang, l'auteur et son équipe ont utilisé:

1) la détection du texte ;

2) la ROC ;

3) l'analyse de composants connexes.

Kostopoulos propose dans son travail d'utiliser :

1) la détection des symboles ;

2) la détection du texte ;

3) la ROC ;

4) l'analyse de la correspondance entre les noms et les rues.

Nous pouvons voir que les deux travaux ont utilisé des techniques presque similaires. les cartes qu'ils ont traitées sont des plans de ville (figure 3.1 et 3.2).

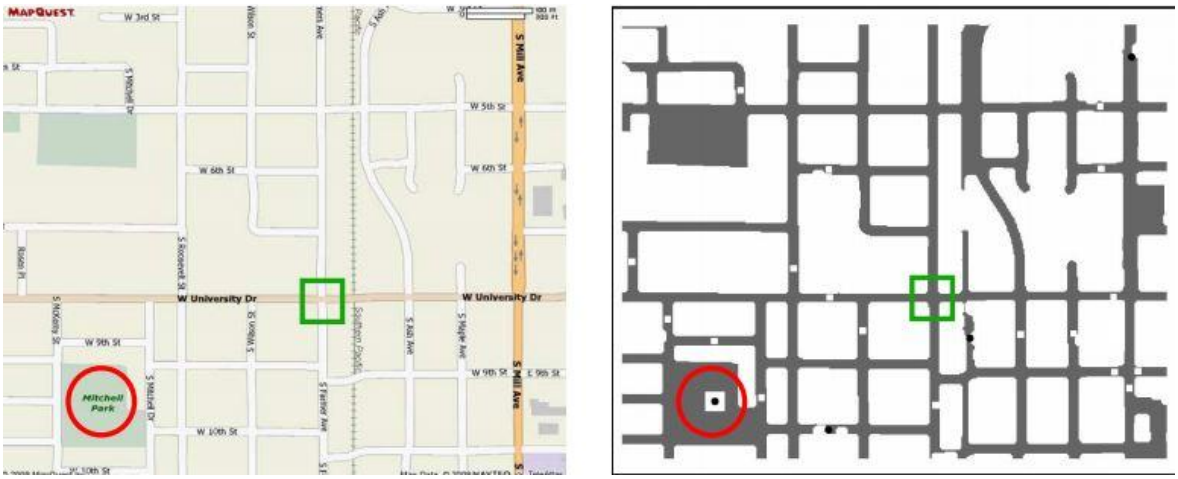


Figure 3.1. Carte en relief créée par Wang

(Extraite de [**Wang 2009**])

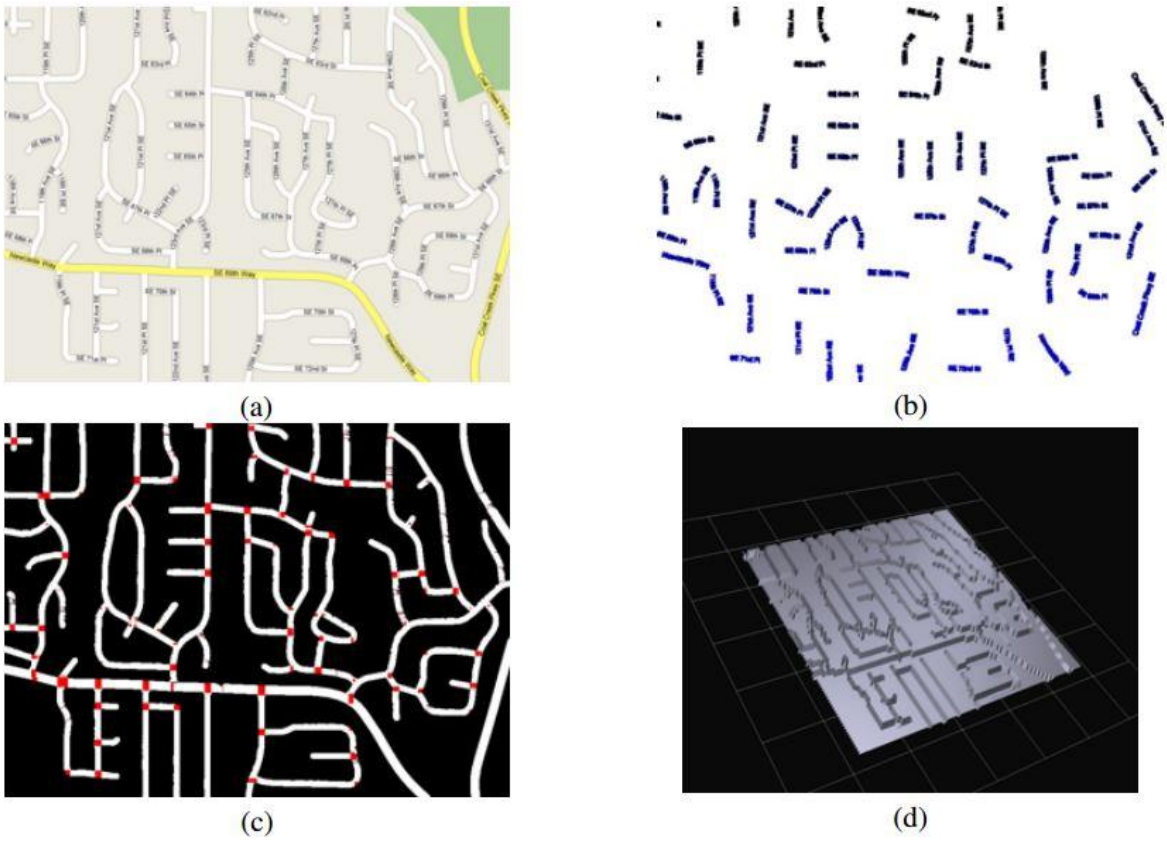


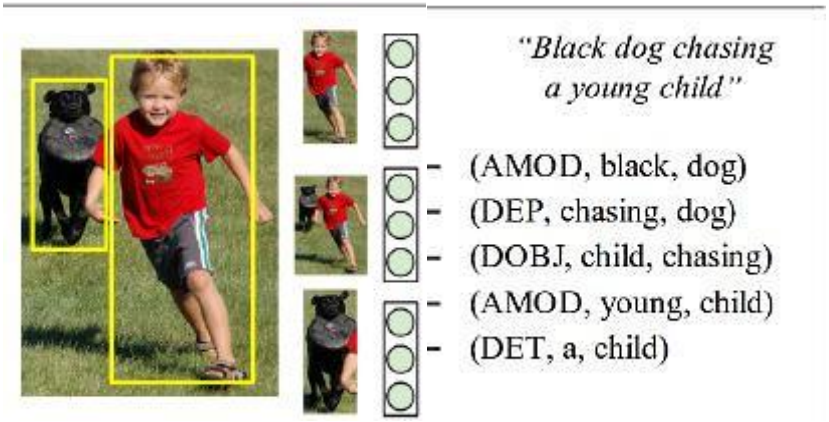
Figure 3.2. Carte en relief créée par Kostopoulos

(Extraite de [**Kostopoulos 2007**])

Les cartes créées sont finalement présentées aux non-voyants via un système de toucher-audio. Wang a montré un retour positif après un test expérimental avec les non-voyants.

Mais dans ces deux travaux, les images traitées sont limitées aux cartes de ville. Les cartes sont simples (peu de couleurs différentes et très peu de formes). En plus, dans la carte obtenue par Wang (figure 3.1), certains endroits (en vert) ne sont pas détectés (ils sont confondus avec les rues). Dans notre travail, les images à traiter sont des images pédagogiques qui sont plus complexes que celles de Wang et Kostopoulos. Les techniques utilisées dans les travaux de Wang [**Wang 2009**] et Kostopoulos [**Kostopoulos 2007**] ne sont plus suffisantes. Nous devons donc étudier et utiliser d'autres méthodes pour réaliser notre travail de recherche.

Dans les travaux de Andrej Karpathy et Fei Fei Li [**Karpathy 2014a**] [**Karpathy 2014b**], une méthode plus intelligente de description d'image, avec la langue naturelle, a été proposée.



(a) (b)

Figure 3.3. Image segmentée et phrase segmentée
(Extraite de [**Karpathy 2014b**])

Selon l'approche proposée, une image est tout d'abord segmentée en certains objets différents (par exemple : la personne, le chien, le ballon, le terrain, etc.) (figure 3.3a). Une

phrase entière de la langue naturelle peut aussi être divisée en certains fragments (figure 3.3b). Le système utilise un modèle de réseau neurone MRNN (Multimodal Recurrent Neural Network en anglais) pour générer automatiquement la description d'une image. Une phrase d'apprentissage avec deux bases de données (une base de données d'images segmentées et une base de données des fragments de la langue naturelle) est nécessairement demandée, avec laquelle le système doit apprendre à associer correctement les différents objets d'image (figure 3.3a) et les fragments de la langue (figure 3.3b) et ensuite à organiser une description correcte globale ou partielle de l'image traitée (figure 3.4 et figure 3.5).



Figure 3.4. Description globale de l'image

(Extraite de [**Karpathy 2014a**])

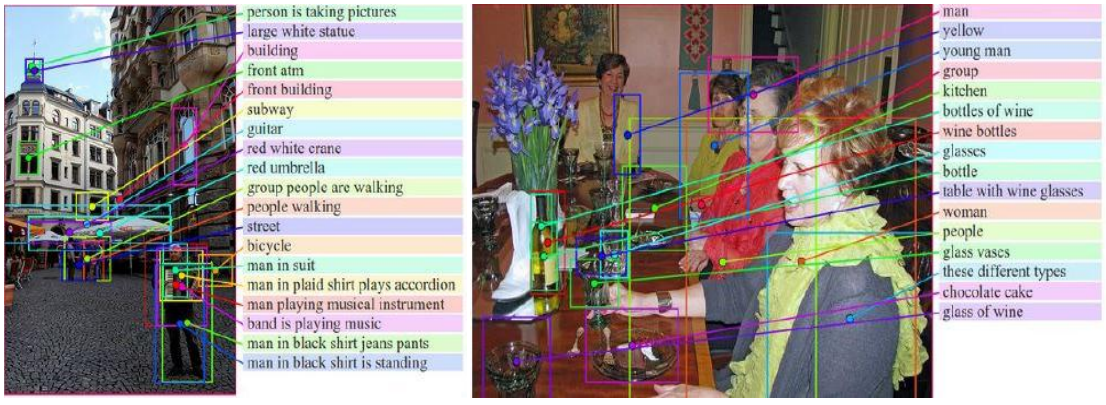


Figure 3.5. Description partielle de l'image

(Extraite de [**Karpathy 2014a**])

Ce système peut traiter les images très complexes (les photos des scènes naturelles par exemple). Il donne une possibilité automatique de rendre les images accessible aux non-voyants. Par exemple, les descriptions des images peuvent être transformées en informations auditives avec une synthèse vocale ou en informations tactiles avec une plage Braille. Mais il reste un problème majeur à résoudre. L'image est toujours invisible pour les non-voyants. Une transformation de l'image originale (visuelle) en image en relief (tactile) doit être mise en place, parce que une image est complètement accessible avec une combinaison de sa version tactile et sa description (ou sa légende) auditive ou tactile.

Dans notre travail de recherche, nous nous sommes focalisés sur la transformation d'une image à vocation pédagogique (extraites des cours du collège) en sa version en relief et la création automatique de sa légende tactile. Notre approche est construite par les différentes méthodes de traitement d'image numérique (la segmentation, la reconnaissance de forme, la reconnaissance de texte, etc.). Nous présentons dans la première partie de ce chapitre les méthodes que nous avons utilisées. Ces méthodes comportent la segmentation de l'image et le traitement des contours. Elles sont les traitements bas niveaux ou les traitements de base. Le but de la segmentation est de trouver les différentes régions sur l'image. Ensuite les différentes textures sont appliquées aux régions segmentées pour que les non-voyants puissent distinguer ces régions texturées.

3.2 La segmentation de l'image

L'étape de segmentation d'image est une étape importante qui permet de distinguer ou de différencier les régions présentes dans une image/carte. Ceci se fait généralement en regroupant des pixels selon un certain critère prédéfini. On distingue généralement deux types de segmentations : celle basée sur les régions et celle basée sur les contours. Les deux catégories de méthodes présentent des avantages et le choix se fait plutôt en fonction de l'application. La segmentation par contours est efficace lorsque les zones séparées par les contours ont des intensités moyennes différentes. Les méthodes basées sur les régions sont beaucoup moins précises sur les contours. Elles sont plutôt fondées sur le regroupement des régions présentant les mêmes attributs, par exemple la luminosité ou la texture. Ces méthodes se basent généralement sur les algorithmes de classification.

3.2.1 Introduction de la segmentation de l'image

La segmentation a pour but de regrouper les pixels de l'image en fonction d'informations caractéristiques (par exemple, les niveaux de gris ou de couleur). Ces groupes de pixels sont appelées « les classes ».

Pour une image I et ses régions $R_1, \cdots, R_n$, la segmentation peut être définie par la partition suivante :

$$I = \bigcup_{i=1}^{n} R_i \text{ et } R_i \cap R_j = \varnothing (i \neq j) \qquad (1)$$

Le domaine de la segmentation d'image est un domaine très vaste qui regroupe de très nombreuses approches. Pour segmenter une image, plusieurs méthodes sont possibles. Il n'existe pas de segmentation idéale. Une bonne méthode de segmentation est celle qui permet d'avoir une bonne interprétation sans trop réduire les informations que l'image contient.

Nous présentons ici les approches les plus fréquentes, elles sont :

- La segmentation fondée sur les régions ;
- La segmentation fondée sur les contours ;
- La segmentation fondée sur le seuillage ;
- La segmentation par approche de « classification ».

3.2.1.1 La segmentation fondée sur les régions

Les méthodes de cette approche s'appliquent directement aux régions. Étant donné un critère d'homogénéité H, la segmentation d'une image I peut être définie par le modèle suivant :

$$\bigcup_{i=1}^{n} X_i = I \qquad (2)$$

($X_1, \cdots, X_n$ sont n régions homogènes de I).

- Pour tout i, X_i est connexe
- Pour tout i, $H[X_i]$ est vrai
- Pour tout couple (X_i, X_j) de régions voisines, $H[X_i, X_j]$ est faux

Pour avoir une bonne segmentation, le choix du critère est le plus important. De plus, la décomposition obtenue n'est pas unique. Pour un critère donné, il existe plusieurs solutions.

1) L'algorithme de «décomposition et fusion»

L'algorithme de «décomposition et fusion» (Split and Merge en anglais), proposé par Horowitz et Pavlidis en 1974 [**Horowitz 1974**], exploite les caractéristiques propres de chaque région. Le processus de cet algorithme est décomposé en deux étapes.

Dans la première, l'étape de décomposition (Split en anglais), on considère en premier temps que l'image est comme un premier quadrant. Si celui-ci est homogène, on arrête. Sinon, on divise ce quadrant en quatre nouveaux quadrants et on applique le même processus à chaque nouveau quadrant jusqu'à ce que tous les quadrants soient homogènes (figure 3.6).

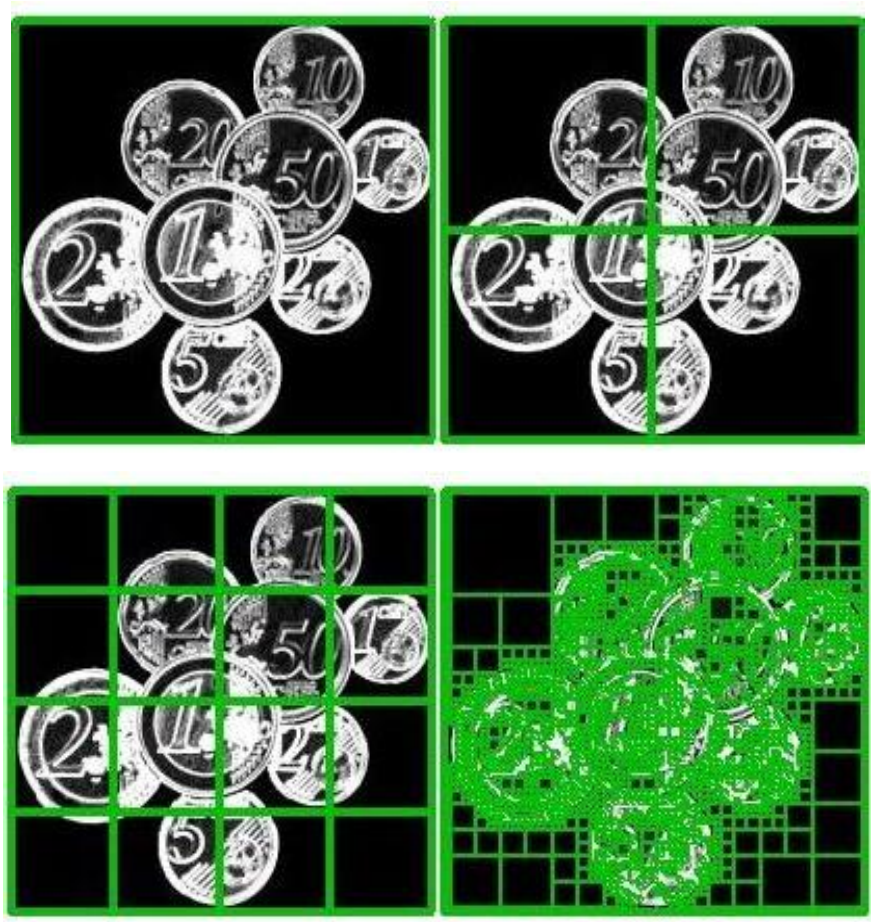


Figure 3.6. L'étape de décomposition[1]

La deuxième étape est la fusion (Merge en anglais). Cette étape traite l'ensemble des rectangles homogènes de l'image segmentée de la première étape et redonne l'ensemble des régions homogènes. Elle s'applique à tous les couples de régions voisines. Si deux régions voisines peuvent former une nouvelle région homogène, elles vont fusionner. Cette étape est répétée jusqu'à ce qu'il n'y ait plus de nouvelle région homogène (figure 3.7).

[1] Images de developpez.com http://xphilipp.developpez.com/articles/segmentation/regions/?page=page_3 15/11/2015

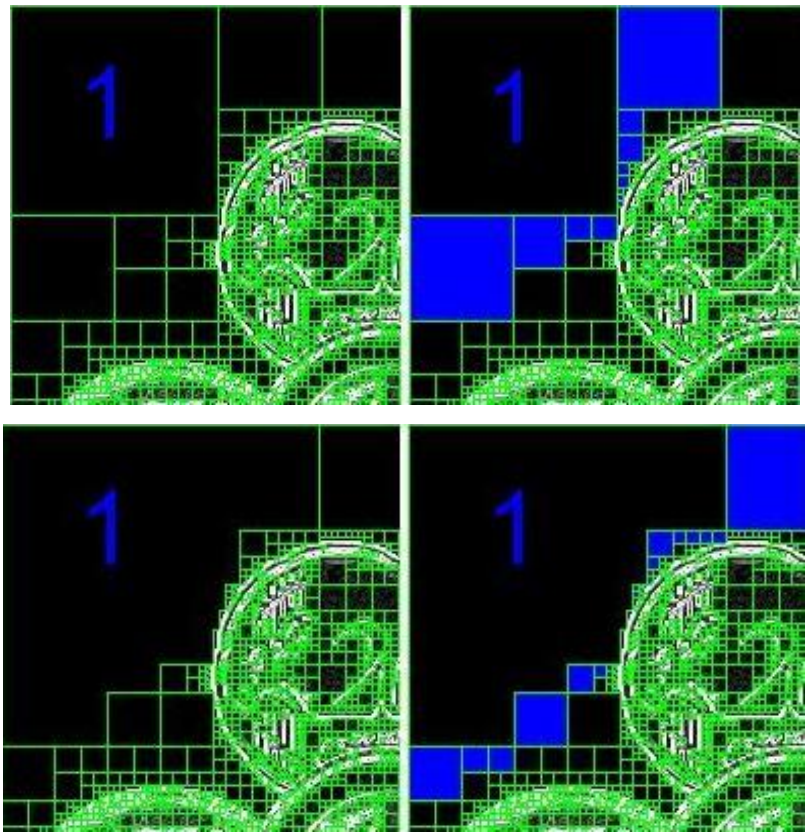


Figure 3.7. L'étape de fusion [1]

2) L'algorithme par croissance de régions (growing-region)

Ce type de technique consiste à faire progressivement grossir les régions à partir d'un point de départ - germe (seeds en anglais) [**Zucker 1976**] [**Tremeau 1997**] [**Adams 1994**].

Le principe de ce type d'algorithme comporte deux étapes :

a) Fixer un point de départ dans l'image ;

b) Grossir les régions par agglomération des pixels voisins.

[1] Images de developpez.com http://xphilipp.developpez.com/articles/segmentation/regions/?page=page_3 15/11/2015

Dans la première étape, le point de départ peut être donné par un utilisateur. Le choix des points de départ est une étape importante. Par conséquent, il vaut mieux choisir les points de départ dans les zones les plus homogènes possibles. En effet, si le point de départ est dans une zone non homogène, la mesure de similarité va produire de fortes variations et la croissance va s'arrêter très tôt. Pour trouver les points de départ, nous pouvons reprendre la première étape de l'algorithme «décomposition et fusion», qui nous fournit une liste de blocs homogènes. Nous pouvons choisir le centre du bloc le plus gros pour obtenir le bon point de départ (figure 3.8).



Figure 3.8. Les points de départ.[1]

La deuxième étape a pour but de grossir les régions. Pour ce faire, on prédéfinit un critère d'homogénéité de la région. Les pixels voisins du point de départ sont ajoutés à cette région s'ils satisfont ce critère (figure 3.9).

[1] Images de developpez.com http://xphilipp.developpez.com/articles/segmentation/regions/?page=page_4 15/11/2015

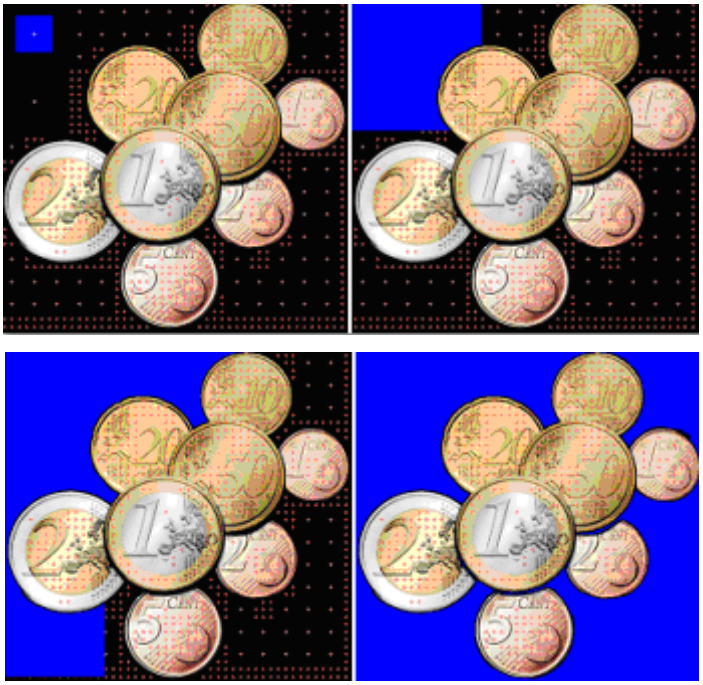


Figure 3.9. Croissance progressive d'une région.[1]

3.2.1.2 La segmentation fondée sur les contours

Cette approche s'applique à la frontière détectable entre deux régions connexes. Il existe plusieurs opérateurs classiques de détection de contours, comme le filtre de Canny [**Canny 1986**] ou le filtre de Sobel [**Sobel 1968**] [**Danielsson 1990**]. On peut aussi utiliser des modèles déformables à l'aide de courbes paramétriques, comme la courbe de Bézier, par exemple. Une fois la frontière entre deux régions bien caractérisée, un autre problème se pose : la fermeture des contours. Pour y parvenir, il existe plusieurs méthodes, par exemple l'algorithme de Deriche et Cocquerez [**Deriche 1988**].

L'intérêt de cette approche de segmentation est de diminuer le nombre d'opérations en cas de traitement de séries d'images qui ont peu de différence entre elles (cas des images vidéo). S'il y a très peu de différences entre les images, une fois qu'on trouve les

[1] Images de developpez.com http://xphilipp.developpez.com/articles/segmentation/regions/?page=page_4 15/11/2015

contours des régions dans la première image, il est plus efficace d'utiliser le modèle déformable que de tout recalculer.

3.2.1.3 La segmentation fondée sur le seuillage

L'objectif est de segmenter une image en utilisant son histogramme. Si dans cet algorithme, chaque classe de l'image est caractérisée par sa distribution de niveaux de gris, à chaque pic de l'histogramme sera associée une classe.

Il existe beaucoup de méthodes dans cette approche. Ces méthodes sont utilisées très souvent pour traiter le cas particulier de la segmentation d'une image en deux classes (vers une image binaire) (figure 3.10).

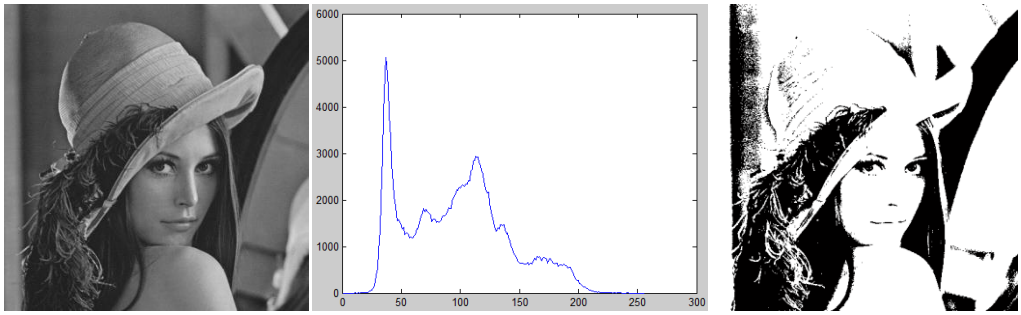


Figure 3.10. Image en binaire conservée par un seuillage de l'histogramme.

3.2.1.4 La segmentation par approche de « classification »

Les méthodes de classification sont utilisées très souvent pour réaliser la segmentation d'une image. Leur but est de regrouper l'ensemble des données en groupes en minimisant la variance intra-classe et en maximisant la variance inter-classe. On considère que la première application de la classification à la segmentation d'image correspond aux travaux de Wacker en 1969 [**Wacker 1969**]. Aujourd'hui, il existe plusieurs algorithmes de classification.

Les différents algorithmes de classification suivent tous les mêmes étapes:

- calcul des indicateurs pour chaque pixel ;
- décorrélation des indicateurs et réduction du nombre d'indicateurs ;
- classification selon les indicateurs ;
- segmentation de l'image.

On peut aussi classer ces méthodes en deux types :

1) la classification non supervisée, qui permet de séparer automatiquement l'image en clusters naturels sans aucune connaissance préalable des caractéristiques des classes ;
2) la classification supervisée, qui fonctionne en connaissant chacune des classes définies.

Les algorithmes supervisés ont besoin des connaissances caractéristiques de chaque classe, mais ce type de méthodes n'est pas considéré comme une bonne solution pour notre travail parce que nous voulons créer une procédure plus automatique. Il est donc plus intéressant d'étudier les méthodes non supervisées.

Dans ce travail, nous avons testé deux méthodes non supervisées : algorithme K-Means et algorithme Fuzzy C-Means(FCM).

3.2.2 Méthodes testées

3.2.2.1 Algorithme K-moyennes (K-Means en anglais)

L'algorithme des K-Moyennes, proposé par MacQueen en 1967 [**MacQueen 1967**], est un algorithme classique qui permet de répartir un ensemble de données relevant des statistiques et de l'apprentissage non supervisé. Le but de cet algorithme est de regrouper les observations en *K* clusters dans lesquelles chaque observation appartient au cluster selon la valeur moyenne la plus proche.

L'idée principale de cet algorithme est de définir K centres de gravité, un pour chaque cluster, qui doivent être choisis d'une manière intelligente parce que les différentes positions des centres peuvent donner différents résultats. Un choix possible est de les placer le plus loin possible les uns des autres.

L'étape suivante consiste à prendre chaque point appartenant à un ensemble de données et l'associer au centre de gravité défini le plus proche. Lorsque tous les points sont classés, l'étape pré-groupement est terminée.

Ensuite, nous avons besoin de recalculer les K nouveaux centres de gravité pour les K groupes donnés par l'étape précédente. Nous obtenons ainsi des nouveaux centres de gravité. Nous revenons à l'étape précédente pour regrouper les points avec les K nouveaux centres.

Les étapes ci-dessus vont être bouclées jusqu'à ce qu'il n'y ait plus de changement sur les centres de gravité, c'est-à-dire que les positions de ces centres soient stables. Enfin, supposant que nous ayons un vecteur $\left(x_1, x_2, x_3, \cdots, x_n \right)$ classé en K clusters, cet algorithme vise à minimiser la fonction objective suivante :

$$J = \sum_{j=1}^{K} \sum_{i=1}^{n} \left\| x_i - c_j \right\|^2 \qquad (3)$$

où x_i est le pixel à classer, c_j est le centre de gravité de chaque cluster, $\left\| x_i - c_j \right\|$ est la distance entre ces deux points.

L'algorithme se compose des étapes suivantes (figure 3.11) :

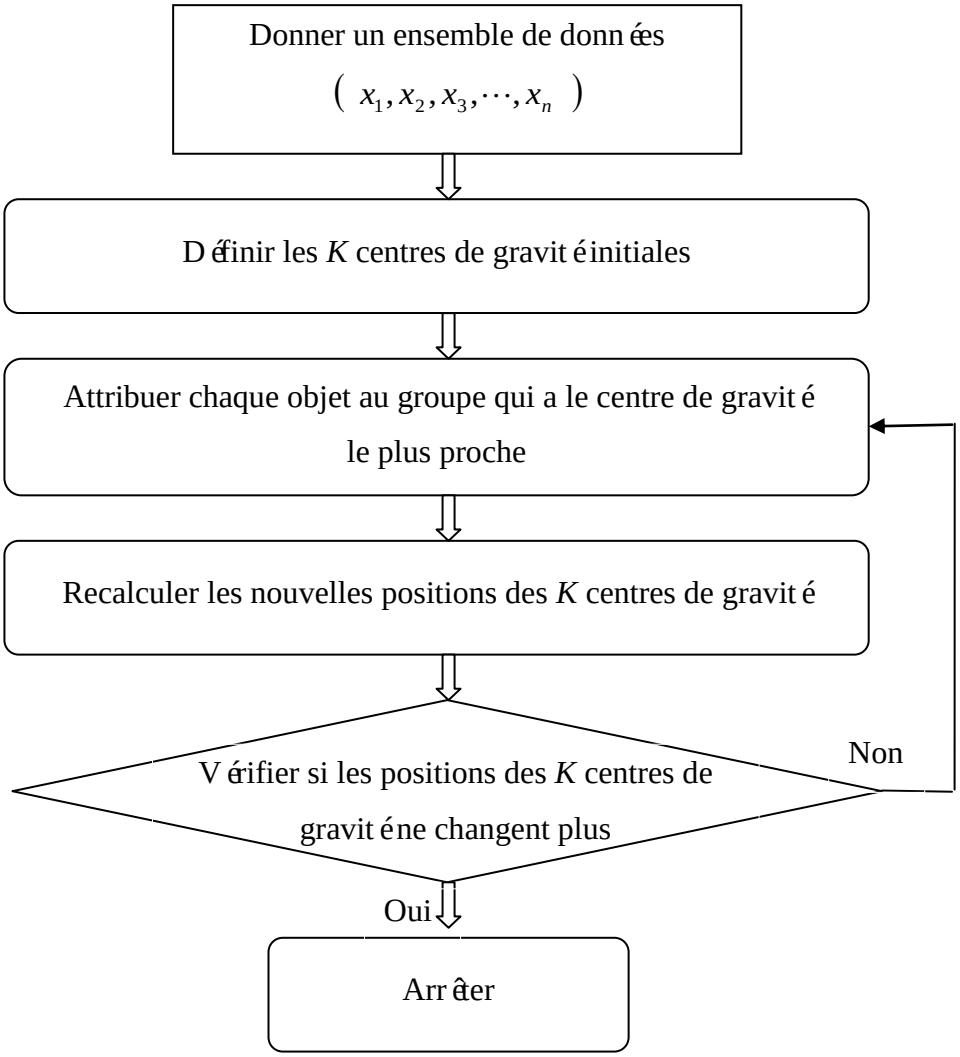


Figure 3.11. Les principales étapes de K-Moyennes.

Pour tester cette algorithme, nous avons utilisé deux types d'image (une photo paysage et une carte géographique, figure 3.12 et 3.13), les résultats de segmentation de chaque image sont donnés dans les figures suivantes (figure3.14 et 3.15) :



Figure 3.12. Image testée (Photo paysage)

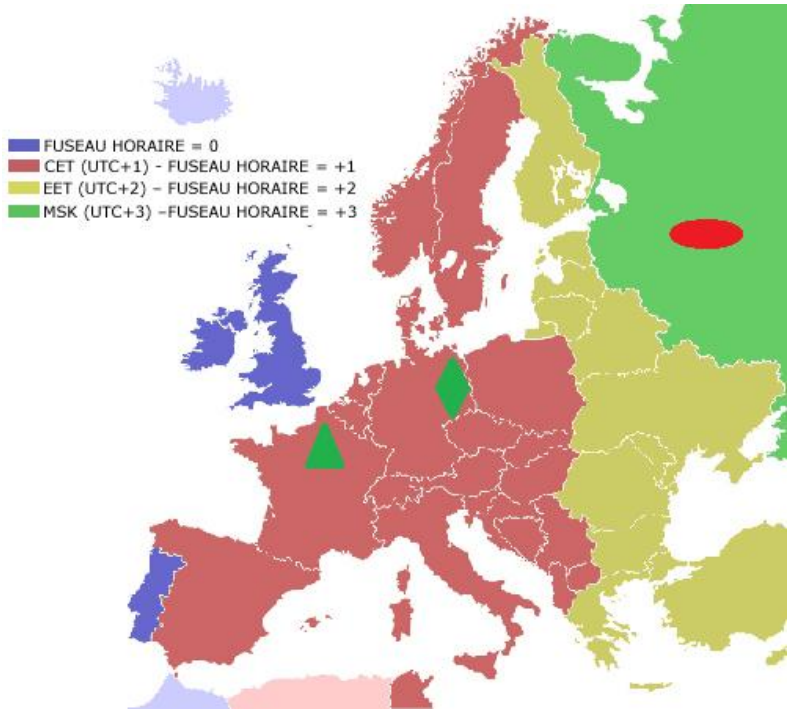


Figure 3.13. Image testée (carte fuseau horaire)

(10 classes en totales : 8 régions en couleur plus les textes en noir et le fond en blanc)

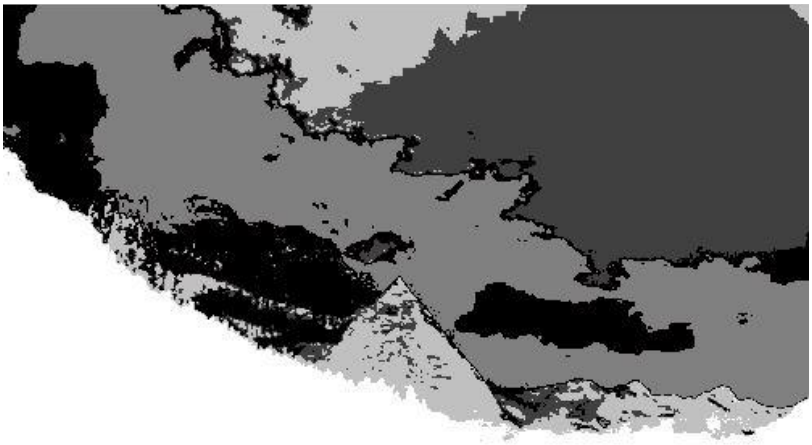
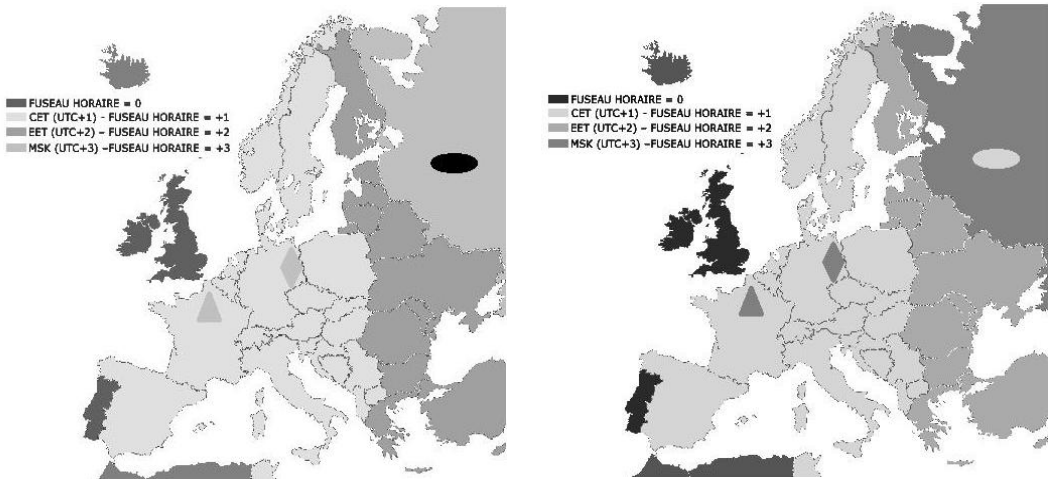


Figure 3.14. Résultats de segmentation par l'algorithme K-moyennes (Photo paysage) (Nombre de classes = 5)



(a) premier essai (b) deuxième essai

Figure 3.15. Résultats de segmentation par l'algorithme K-moyennes (carte fuseau horaire) (Nombre de classes = 10)

Les résultats obtenus montrent que l'algorithme K-moyennes ne donne pas le même résultat au premier et au deuxième essai, mais que le nombre de classes reste similaire. En effet, la qualité de la segmentation de cet algorithme dépend beaucoup de l'initialisation des points de départ : si les points de départ sont bien définis, on peut obtenir un très bon résultat de segmentation, sinon, le résultat est très mauvais. Dans notre procédure, les

points de départ sont définis aléatoirement par le système, c'est pour cette raison qu'il est possible d'avoir des résultats différents à chaque essai.

3.2.2.2 Algorithme des C-Moyennes floues (FCM)

L'algorithme des C-Moyennes floues (Fuzzy C-Means en anglais) [**Priya 2011**] [**Shambharkar 2011**] [**Narayan 2012**] est un algorithme de classification non-supervisée floue. Il permet d'obtenir, pour chaque pixel de l'image, un degré (entre 0 et 1) d'appartenance à une classe donnée.

L'algorithme a besoin de connaître le nombre de classes au préalable et génère les classes par un processus itératif en minimisant une fonction objective (fonction N°6). Pour une image en couleur, nous avons l'ensemble des pixels $\{x_1, x_2, \cdots, x_n\}$, x_i est un vecteur de trois composantes (RVB par exemple). Pour un nombre de classe connu C, on peut définir une matrice d'appartenance U :

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1k} & \cdots & u_{1n} \\ \vdots & \ddots & & & \vdots \\ u_{i1} & & u_{ik} & & u_{in} \\ \vdots & & & \ddots & \vdots \\ u_{C1} & \cdots & u_{Ck} & \cdots & u_{Cn} \end{bmatrix} \text{ avec } 1 \le k \le n,\ 1 \le i \le C \qquad (4)$$

où u_{ik} est le degré d'appartenance du pixel i à la classe k :

$$u_{ik} = \sum_{j=1}^{C} \left(D_{jk}\right)^{\frac{2}{m-1}} \Big/ \left(D_{ik}\right)^{\frac{2}{m-1}}, \text{ avec } D_{jk} = \left\| x_k - c_j \right\| \qquad (5)$$

Cet algorithme vise à chercher la matrice d'appartenance U en minimisant une fonction objective :

$$J = \sum_{k=1}^{n} \sum_{i=1}^{C} \left(u_{ik}\right)^m \cdot \left\| x_i - c_k \right\|^2 \qquad (6)$$

où :

m>1 est un paramètre contrôlant le degré de flou (généralement m=2), c_k est le centre de gravité de la classe k.

L'algorithme se compose des étapes suivantes (figure 3.16) :

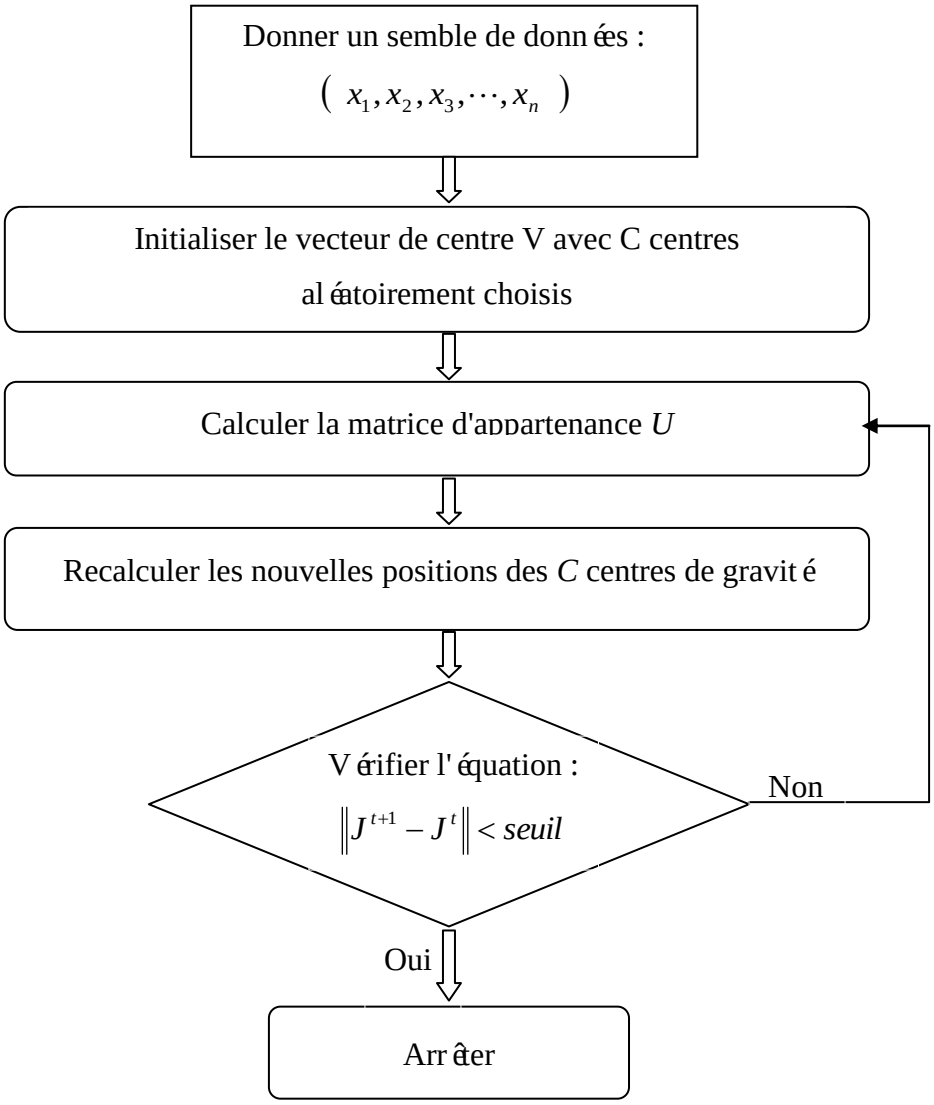


Figure 3.16. Les principales étapes de *C-Moyens flou.*

Les mêmes images sont utilisées pour tester l'algorithme FCM, les résultats sont présentés dans les figures suivantes (figure 3.17 et 3.18) :

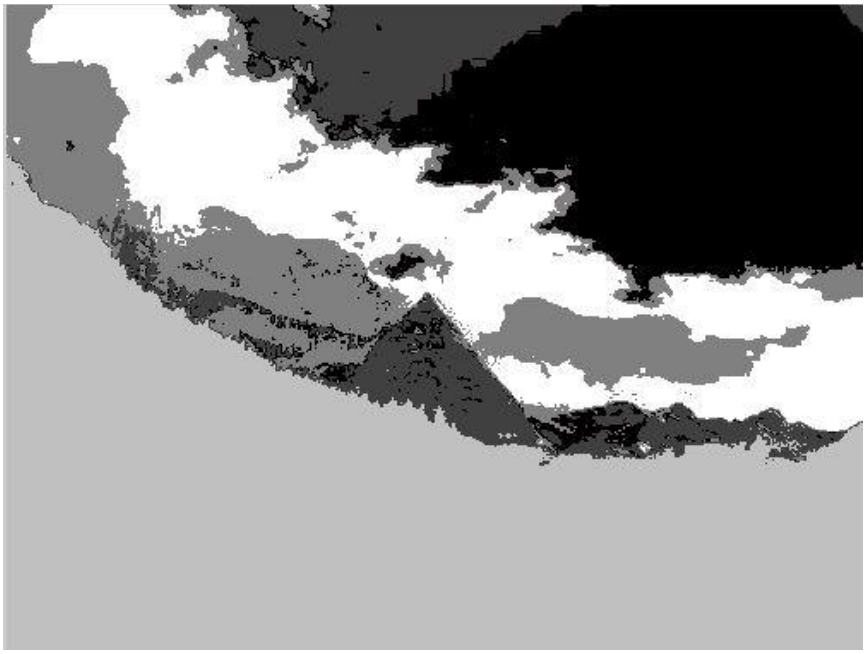
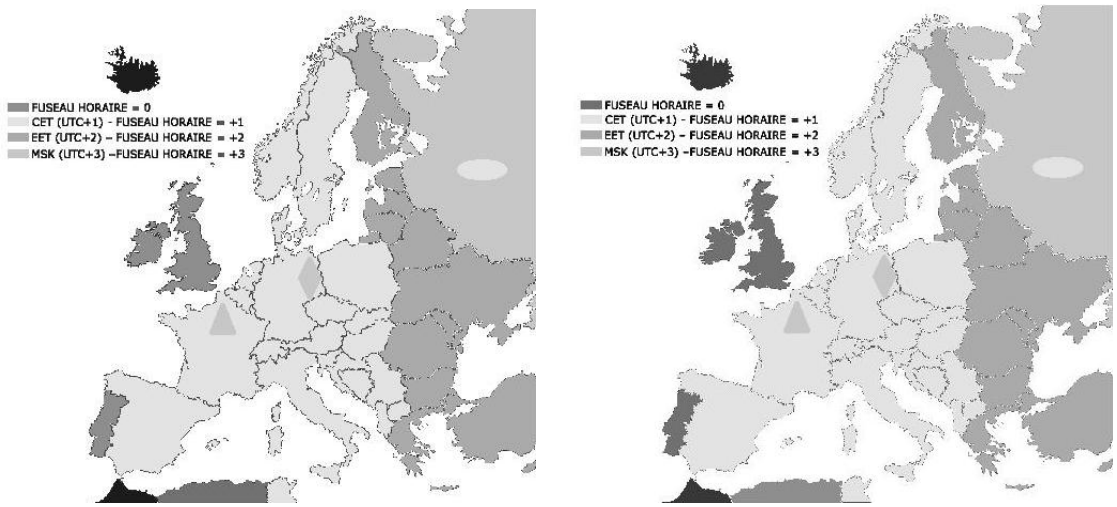


Figure 3.17. Résultats de segmentation par l'algorithme FCM (Photo paysage)
(Nombre de classes = 5)



(a) premier essai (b) deuxième essai

Figure 3.18. Résultats de segmentation par l'algorithme FCM (carte fuseau horaire)
(Nombre de classes = 10)

Les résultats montrent que la qualité de segmentation par l'algorithme FCM est toujours bonne, même avec une définition automatique et aléatoire du vecteur de centre.

3.2.3 Conclusion

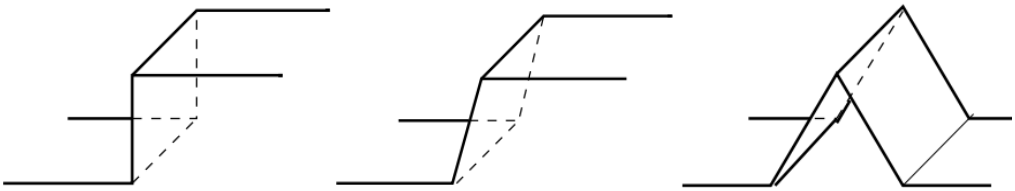
Les tests menés sur ces deux algorithmes (K-moyennes et FCM) montrent un inconvénient majeur pour l'algorithme K-moyennes : la qualité de la segmentation dépend de l'initialisation des points de départ. Ainsi, puisque nous cherchons une procédure automatique dans notre travail, l'algorithme K-moyennes n'est pas un bon choix car il peut produire des erreurs si nous déterminons automatiquement les points de départ. Au contraire, l'algorithme FCM fournit toujours le même résultat que l'initialisation soit manuelle ou automatique. Nous avons donc décidé d'utiliser l'algorithme FCM pour notre travail.

3.3 Traitement de contours

Dans notre recherche, le traitement de contours représente une partie très importante. En effet, les contours détectés et renforcés permettent de bien séparer les différents éléments de l'image pour augmenter sa lisibilité lorsqu'elle est en relief. Ces contours ont la même fonction que les blancs de séparation. Pour chaque élément de l'image, ils donnent une première perception visuelle de sa propriété géométrique. Dans le cas d'une image tactile, les contours doivent être mis en relief en respectant certaines contraintes pour que les non-voyants puissent les lire par le toucher. Pour ce but, les contours d'une image originale doivent être détectés, simplifiés et finalement renforcés pour être facilement accessibles au toucher.

3.3.1 Introduction de la détection de contours

La détection de contours est une étape importante dans le traitement d'image par ordinateur. Le but de la détection est de trouver et de regrouper les points d'une image qui correspondent à un changement brutal de l'intensité lumineuse. La figure 3.19 donne quelques modèles de contours.



(a) marche (b) rampe (c) toit

Figure 3.19. Quelques modèles de contours[1]

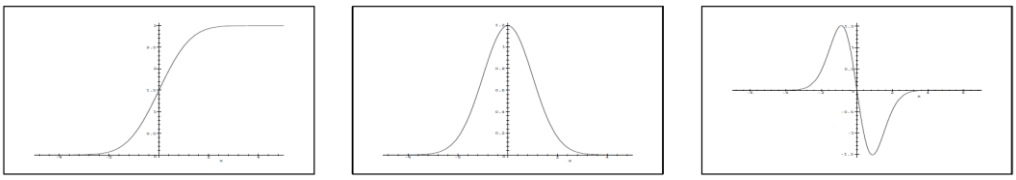
Il existe plusieurs méthodes de détection. En général, la plupart de ces méthodes peuvent être classées en deux catégories :

1. Les détecteurs de gradient.

   Cette famille de détecteurs recherche en général les maximums locaux de l'intensité du gradient. C'est-à-dire la détection de la valeur maximum absolue de la première dérivée (figure 3.20b).

2. Les détecteurs Laplacien.

   Ce genre de détecteur cherche en général les annulations du Laplacien ou d'une expression différentielle non linéaire. C'est-à-dire la détection du passage par zéro de la seconde dérivée (figure 3.20c).



(a) (b) (c)

Figure 3.20. Le détecteur Laplacien

(a : Contour de type marche, b : Dérivée première, c : Dérivée seconde,

Extraite de [**Fortier 1999**]).

[1] H. Maître, La détection de contour dans les images, http://perso.telecom-paristech.fr/~bloch/TDI/poly_contours.pdf

Ces méthodes peuvent être classées ensuite suivant la manière d'estimer les dérivées de la fonction d'intensité:

1. Différence finie [**Ames 1977**].

   Étant donné qu'une image est discrète par nature, les premières approches approximent les dérivées par différence :

$$\nabla_u I(u,v) = I(u,v) - I(u-n,v) \qquad (7)$$

   ou :

$$\nabla_u I(u,v) = I(u+n,v) - I(u-n,v) \qquad (8)$$

   avec $n = 1$ par défaut.

   Pour obtenir ces dérivées, on calcule la convolution entre l'image et un masque de différence (Opérateur de Roberts, Opérateur de Prewitt, Opérateur de Sobel, etc.).

2. Dérivation par filtrage optimal [**Canny 1986**]

   Les dérivations présentées dans la première approche consistent à convoluer l'image par des masques. Ces approches sont donc dépendantes de la taille des images et elles sont aussi très sensibles au bruit.

   Un autre type d'approche est proposé qui repose sur la définition de critère d'optimalité de la détection de contours. Les critères d'optimalité sont :

   - Bonne détection : le contour doit être détecté en minimisant les fausses réponses,
   - Bonne localisation : le contour doit être détecté en minimisant la distance entre les points détectés et le vrai contour,
   - Réponse unique : il faut minimiser le nombre de réponses pour un seul contour.

Ces critères débouchent sur des filtres de lissage optimaux (Filtre Canny, Filtre de Shen -Castan, Filtre Gaussien etc.) :

Soit h le filtre de lissage, $I(x,y)*h(x)*h(y)$ est l'image lissée, $I(x,y)*h'(x)*h(y)$ et $I(x,y)*h(x)*h'(y)$ sont les images des dérivées suivantes x et y. Enfin, on peut calculer l'image de Laplacien :.

$$I(x,y)*\left(h''(x)h(y)+h(x)h''(y)\right) \qquad (9)$$

3.3.2 Méthodes testées

3.3.2.1 Opérateur de Roberts

L'opérateur de Roberts [**Roberts 1963**] permet de calculer le gradient bidimensionnel d'une image de manière simple et rapide. Il est constitué par deux masques de convolution :

a) Filtre horizontal :

$$h_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad (10)$$

b) Filtre vertical :

$$h_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad (11)$$

Si $p(x,y)$ est un pixel d'une image, les amplitudes des gradients selon x et y peuvent s'écrire :

$$G_x = p(x+1,y) - p(x,y) \qquad (12)$$

$$G_y = p(x,y+1) - p(x,y) \qquad (13)$$

et l'orientation du gradient est donnée par :

$$D(x,y) = act\tan(G_y / G_x) \qquad (14)$$

La norme du gradient du pixel $p(x,y)$ est :

$$G(x,y) = \sqrt{G_x^{\ 2} + G_y^{\ 2}} \qquad (15)$$

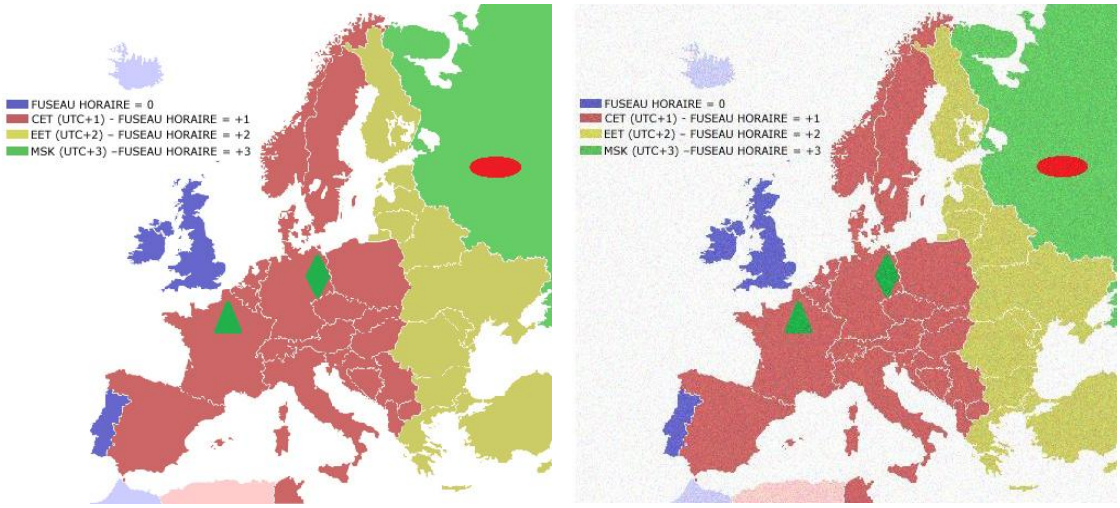
Cet opérateur est rapide parce que la taille du masque est petite. Mais il est très sensible au bruit parce qu'il augmente, par dérivation, le bruit présent dans l'image. Si les bords ne sont pas très forts, le filtre aura tendance à ne pas les détecter.

Les résultats des tests effectués montrent les caractéristiques de cet opérateur. Nous testons cet algorithme sur deux types d'images : l'une est une photo (figure 3.21), l'autre est une carte géographique (figure 3.22). Nous avons appliqué les bruits (type de poisson) sur ces images pour tester la sensibilité au bruit. Sur chaque image, nous effectuons deux tests : l'un avec les seuils auto-choisis ; l'autre avec les seuils choisis manuellement. Les résultats sont donnés dans les figures suivantes (figure 3.23 et 3.24) :



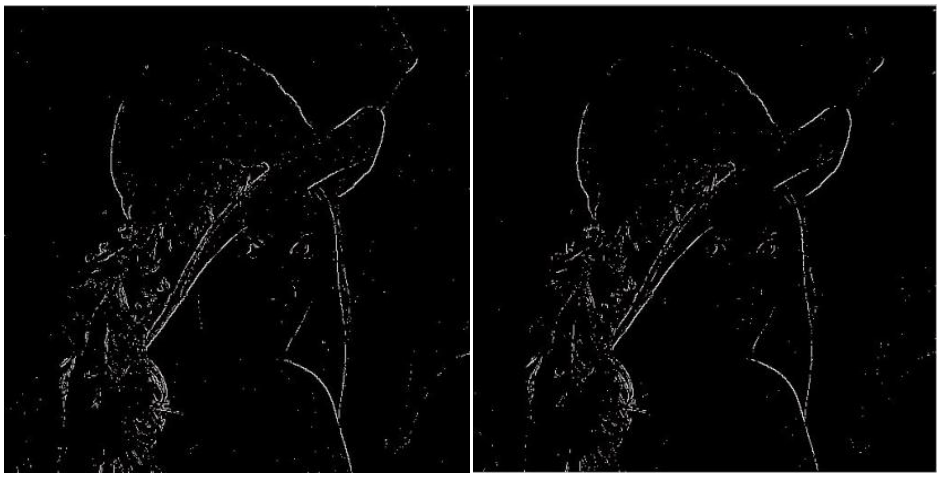
(a) Image originale (b) Image niveau de gris bruitée

Figure 3.21. Image testée «Lena»



(a) Image originale (b) Image niveau de gris bruitée

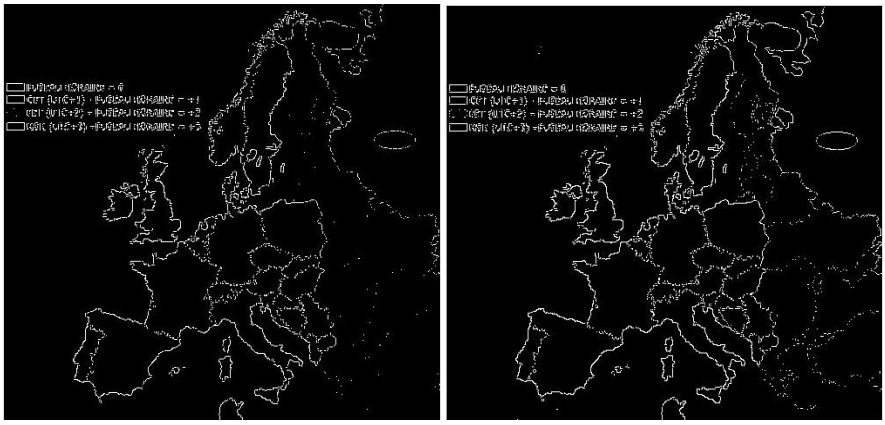
Figure 3.22. Image testée (carte géographique-fuseau horaire)



(a) ta = 0.13 (b) tm = 0.15

Figure 3.23. Contours détectés par l'opérateur de Roberts «Lena»

(ta : seuil auto-choisi, tm : seuil choisi manuellement)



(a) ta = 0.19 (b) tm = 0.15

Figure 3.24. Contours détectés par l'opérateur de Roberts (carte géographique-fuseau horaire)

(ta : seuil auto-choisi, tm : seuil choisi manuellement)

Les résultats montrent que l'opérateur de Roberts est très sensible au bruit. Pour les deux différentes procédures testées (la procédure automatique et la procédure manuelle), la plupart des contours ne sont pas détectés pour l'image de «Lena», et certains contours sont perdus pour la carte des fuseaux horaires. Cet opérateur n'est pas un bon choix pour notre travail.

3.3.2.2 Opérateur de Prewitt et de Sobel

Les opérateurs de Prewitt [**Prewitt 1970**] et de Sobel [**Sobel 1968**] sont aussi des opérateurs de dérivation, les deux masques de convolution sont :

a) Filtre horizontal :

$$h_x = \begin{pmatrix} -1 & 0 & 1 \\ -c & 0 & c \\ -1 & 0 & 1 \end{pmatrix} \qquad (16)$$

b) Filtre vertical :

$$h_y = \begin{pmatrix} -1 & -c & -1 \\ 0 & 0 & 0 \\ 1 & c & 1 \end{pmatrix} \qquad (17)$$

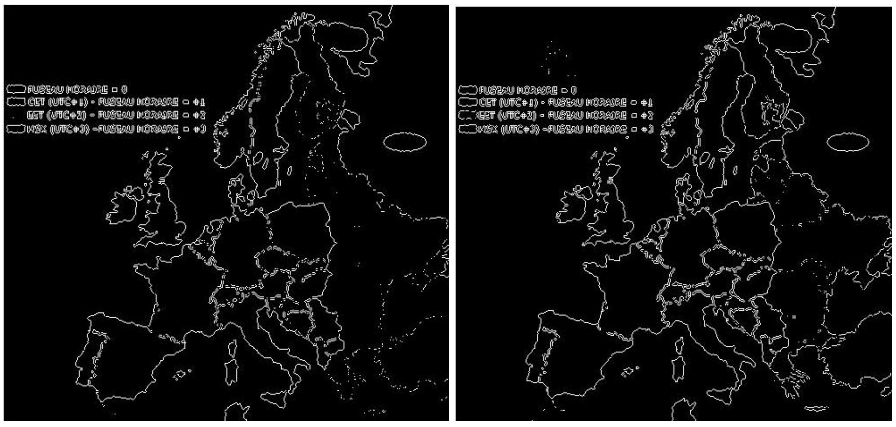
Si $c = 1$, ce sont des opérateurs de Prewitt, et si $c = 2$, ce sont des opérateurs de Sobel. Par rapport aux opérateurs de Roberts, ces filtres sont aussi rapides, mais moins sensibles au bruit, car ils effectuent un lissage dans la direction orthogonale.

Pour bien connaître ces caractéristiques, nous avons appliqué le même test que celui effectué avec l'opérateur de Robert. Les résultats sont donnés dans les figures suivantes (figure 3.25 et 3.26) :



(a) ta = 0.097 (b) tm = 0.1

Figure 3.25. Contours détectés par l'opérateur de Sobel «Lena»
(ta : seuil auto-choisi, tm : seuil choisi manuellement)



(a) ta = 0.13 (b) tm = 0.1

Figure 3.26. Contours détectés par l'opérateur de Sobel (carte géographique-fuseau horaire)
(ta : seuil auto-choisi, tm : seuil choisi manuellement)

Les résultats montrent que l'opérateur de Sobel est moins sensible au bruit. La qualité de la détection de contours est meilleure que celle de l'opérateur de Roberts. On ne perçoit pas beaucoup de différentes entre la procédure automatique et la procédure manuelle.

3.3.2.3 Détecteur de Canny

Le détecteur de Canny est un opérateur de la détection de contours développé par John F. Canny en 1986 [**Canny 1986**]. Le but de développer cet opérateur est de trouver un algorithme pour satisfaire les trois critères (voir 3.3.1). Cet opérateur comporte 4 parties :

1) Réduction du bruit

Avant de commencer à détecter les contours, un filtrage gaussien de deux dimensions est appliqué sur l'image originale pour réduire le bruit. Ceci permet d'éliminer les points isolés qui pourraient donner une fausse réponse lors du calcul du gradient.

Le filtrage gaussien est donné par le modèle suivant :

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (18)$$

Normalement, la taille du filtre est plus petite que l'image filtrée. Si le masque est plus grand, le détecteur est moins sensible au bruit et la localisation des contours sera de ce fait moins précise.

2) Gradient d'intensité et direction des contours

Après la réduction du bruit, deux masques de convolution sont utilisés pour calculer les gradients d'intensité de chaque point suivant les directions horizontales et verticales. Les deux masques sont :

$$h_x = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \quad (19)$$

$$h_y = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad (20)$$

L'amplitude du gradient et la direction peuvent être calculées de la même manière que pour les filtres de Roberts.

3) Suppression des non-maxima

Dans la carte des gradients obtenus, chaque point a une intensité. En fait, pour un point, une forte intensité donne une forte probabilité de présence d'un contour. Pour trouver les points de contours, on cherche les maxima locaux sur cette carte.

4) Seuillage des contours

Pour filtrer les différents contours, cette méthode applique un seuillage à hystérésis qui a besoin de deux seuils, un seuil haut et un seuil bas, qui seront comparés à l'intensité du gradient de chaque point. Si l'intensité du gradient est :

- plus petite que le seuil bas, ce point est rejeté;
- plus grande que le seuil haut, ce point est retenu ;
- entre le seuil bas et le seuil haut, ce point est retenu s'il est connecté à un point déjà retenu.

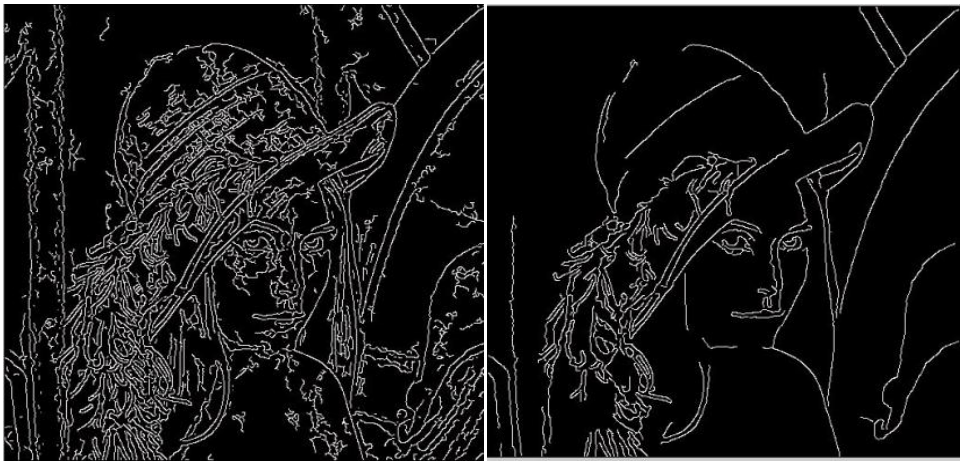
L'algorithme de Canny contient certains paramètres qui peuvent influencer le temps de calcul et l'acuité. Les deux paramètres les plus importants sont la taille du filtrage gaussien et les deux seuils :

a) Taille du filtre : le filtre de lissage utilisé dans la première étape affecte directement les résultats de cet algorithme. Les filtres de petite taille causent

moins de flous et permettent de détecter les petites lignes. Les filtres de taille plus grande produisent un effet de flou plus important, ce qui permet une détection des contours moins nets.

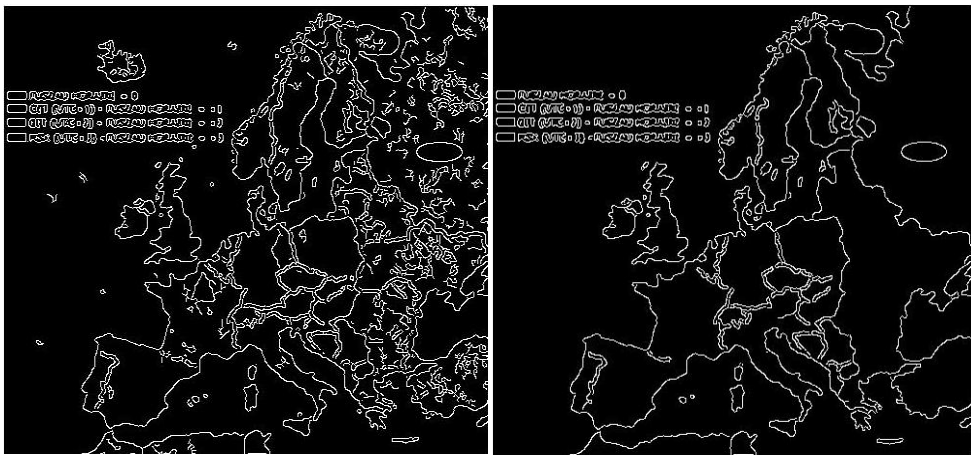
b) Seuils : l'utilisation de deux seuils (au lieu d'un seul) augmente la flexibilité, mais deux problèmes se présentent : si le seuil est trop élevé, les informations importantes risquent d'être perdues, si le seuil est trop bas, on risque d'obtenir une réponse fausse. Pour l'instant, il n'existe pas de méthode générique pour déterminer des seuils qui peuvent donner des résultats satisfaisants pour tous les types d'images.

Nous utilisons toujours les mêmes images pour réaliser le test. Les résultats obtenus sont présentés dans les figures suivantes (figure 3.27 et 3.28) :



(a) ta = (0.037, 0.094) (b) tm = (0.1, 0.3)

Figure 3.27. Contours détectés par l'opérateur de Canny «Lena»
(ta : seuil auto-choisi, tm : seuil choisi manuellement)



(a) ta = (0.031, 0.078) (b) tm = (0.1, 0.3)

Figure 3.28. Contours détectés par l'opérateur de Canny (carte géographique-fuseau horaire)

(ta : seuil auto-choisi, tm : seuil choisi manuellement)

Nous voyons que les contours sont bien détectés par la procédure manuelle, mais en revanche, pour la procédure automatique, les résultats sont très mauvais. Il est, en effet, très difficile de trouver les bonnes valeurs pour les 3 paramètres de façon automatique. Nous avons dû effectuer plusieurs essais pour définir les bons paramètres lors de la procédure manuelle.

3.3.3 Conclusion

En comparant les résultats obtenus, on observe que, au niveau de la sensibilité au bruit, les algorithmes de Sobel et de Canny ont donné des meilleures résultats que l'algorithme de Roberts. Mais la qualité de la réponse de Canny dépend 3 paramètres (seuil bas, seuil haut, et taille de la fenêtre), qui doivent être bien définis pour obtenir un bon contour. En revanche, l'algorithme de Sobel dépend d'un seul paramètre.

L'objectif de notre travail est d'utiliser des procédures automatiques. Or, en comparant les résultats obtenus automatiquement par les algorithmes de Sobel et de Canny, nous remarquons que l'algorithme de Sobel donne toujours des meilleurs résultats et est plus robuste au bruit. Et comme il est plus facile d'établir une procédure automatique à

partir d'un seul paramètre, c'est l'algorithme de Sobel que nous retenons pour notre travail. Il nous garantit d'obtenir de bons résultats même avec un seuillage auto-choisi.

3.4 Conclusion

Dans ce chapitre, nous avons présenté les deux premières étapes de notre procédure : la segmentation d'image et le traitement des contours. Ces deux étapes sont des étapes considérées comme un traitement de bas niveau qui ne prend en compte aucun contenu sémantique. Les autres étapes, la reconnaissance de formes et la reconnaissance de texte, vont s'effectuer en se basant sur les résultats des deux étapes précédentes. Par exemple, les formes à reconnaître sont extraites par l'étape segmentation ; la détection de texte est basée sur la détection de contours. Dans le chapitre suivant, nous allons présenter les étapes qui tentent d'interpréter un contenu sémantique de quelques formes simples pour rendre l'image accessible.

CHAPITRE 4

La création automatique d'images en relief avec l'aide des technologies de traitement d'image --traitements haut niveaux

Sommaire :

4.1 Introduction 84
4.2 Étape de la reconnaissance de formes 84
4.2.1 Introduction de la reconnaissance de formes 84
4.2.2 Méthodes testées 88
4.2.3 Conclusion 104
4.3 Étape de détection de flèches 105
4.3.1 Introduction et problématique 105
4.3.2 Méthodes testées 107
4.4 Étape de la reconnaissance de texte 117
4.4.1 Introduction de la ROC 117
4.4.2 Méthodes testées 119
4.4.3 Conclusion 132
4.5 Conclusion 132

4.1 Introduction

Dans ce chapitre, nous présentons les étapes des traitements haut niveau de l'image dans notre procédure, y compris l'étape de la reconnaissance de formes, l'étape de la reconnaissance de textes et l'étape de la détection des flèches, cette étape doit associer un contenu sémantique à une étiquette ou à un texte de manière à faciliter sa compréhension.

4.2 Étape de la reconnaissance de formes

Il est assez fréquent que les images contiennent des formes spécifiques qui correspondent à certaines informations. Ces dernières sont données par les légendes correspondant à ces images. Dans ce travail, nous proposons une détection automatique et une reconnaissance intelligente de ces formes dans l'image.

4.2.1 Introduction de la reconnaissance de formes

La reconnaissance de formes est un domaine très large en informatique. Les formes (ou parfois les motifs), en général, ne sont pas seulement des formes géométriques, elles peuvent être aussi de natures très variées. Par exemple, il peut s'agir d'éléments visuels comme les visages ou sonores comme la parole, ou d'image médicale comme l'image de radio.

Donc le terme - reconnaissance de formes (ou reconnaissance de motifs), en générale, est un ensemble de techniques visant à identifier des formes à partir de données brutes en définissant les caractéristiques de ces formes.

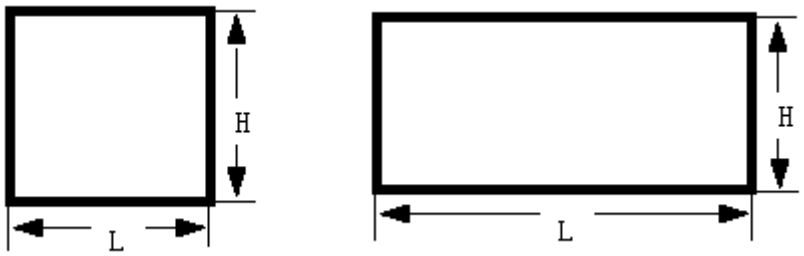
Il existe en général deux approches de la reconnaissance des formes : l'approche statistique et l'approche structurelle (ou syntaxique).

4.2.1.1 L'approche statistique

L'approche statistique est l'approche standard en reconnaissance de formes simples. Une forme ou un objet à reconnaître peut être représenté sous la forme d'un vecteur caractéristique. Le but de cette approche est de construire ce vecteur, et ensuite, de reconnaître la forme en classifiant ce vecteur.

Par exemple, nous avons deux classes simples :

Classe A : les rectangles carrés, et Classe B : les rectangles non-carrés (figure 4.1).



Classe A Classe B

Figure 4.1. Deux classes de rectangle : rectangle carré et non-carré avec deux caractéristiques.
(L : largeur, H : hauteur)

Un vecteur de caractéristiques peut être défini par deux attributs : $V(L,H)$. Ces deux caractéristiques définissent un espace de caractéristiques (figure 4.2). Dans cette espace, chaque groupe est associé à un nuage de points.

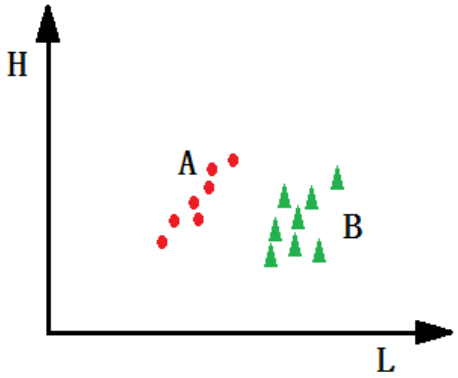


Figure 4.2. L'espace de caractéristique de rectangle
(axe horizontal : largeur, axe vertical : hauteur, A : Classe rectangle carrée, B : Classe rectangle non-carré, la largeur est plus grande que la hauteur)

Lorsque les objets à traiter sont représentés par certaines caractéristiques, nous devons choisir une procédure adéquate pour classifier ces objets d'après leur vecteur de caractéristiques. Le choix des caractéristiques est très important. Il faut qu'elles soient le plus discriminantes possible. Pour une forme géométrique simple, elles peuvent être : largeur, hauteur, surface, excentricité..

La procédure générale de cette approche est présentée dans la figure 4.3 :

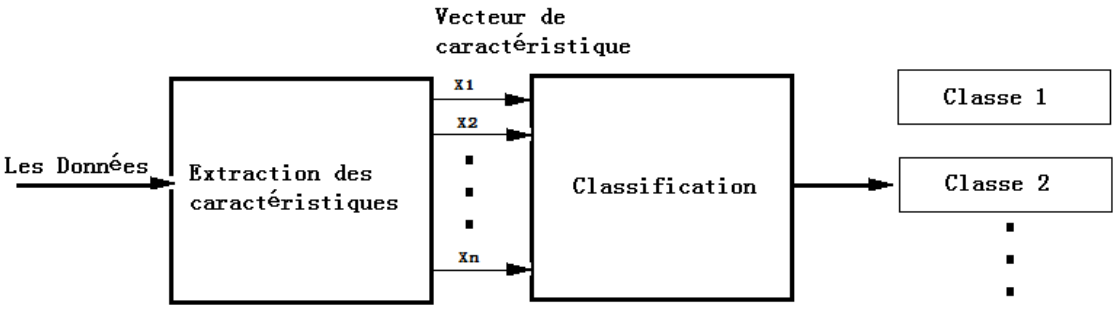


Figure 4.3. La procédure générale de la reconnaissance de forme statistique.

4.2.1.2 L'approche structurelle

La reconnaissance de formes structurelles est utilisée pour traiter des objets plus complexes. Pour décrire un objet, cette approche utilise trois paramètres : les éléments primitifs, les relations, et les organisations (figure 4.4).

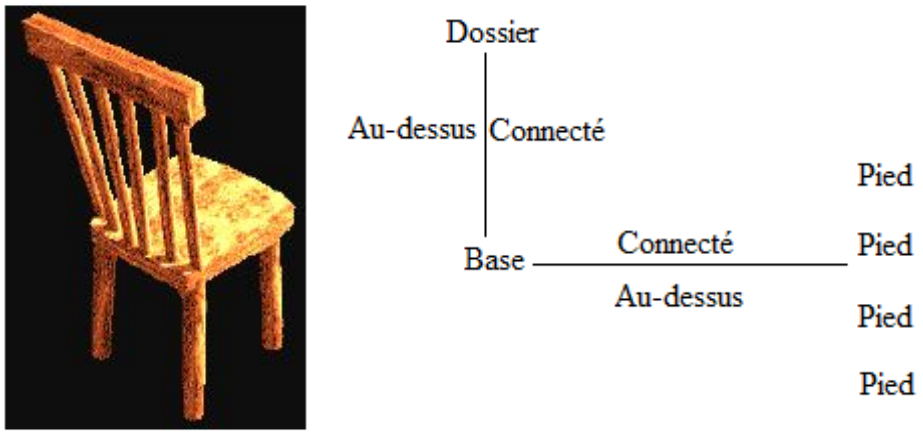


Figure 4.4. Une chaise est décrite par l'approche structurelle

Sur la figure 4.5, on présente la procédure de cette approche :

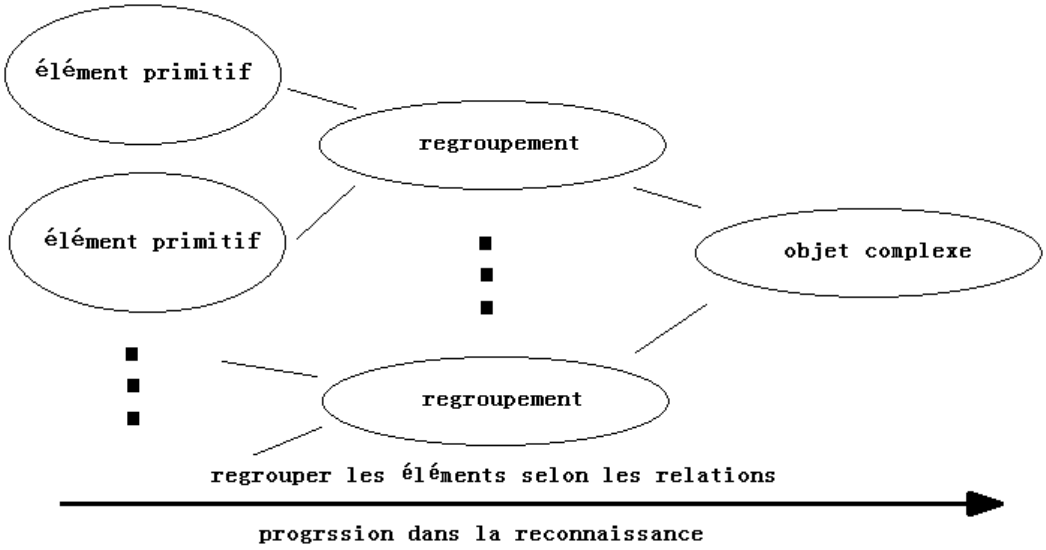


Figure 4.5. La procédure de la reconnaissance de formes structurelles.

Dans notre travail de recherche, nous travaillons sur des images pédagogiques : il s'agit de formes géométriques simples. Nous allons étudier les méthodes de la première approche.

4.2.2 Méthodes testées

4.2.2.1 Transformation de Hough

La transformation de Hough est une méthode de reconnaissance de formes dans l'image. Elle a été inventée par P.V.C Hough en 1962 [**Hough 1962**]. Ensuite, pendant plusieurs années, cette méthode a été à la base d'autres travaux comme ceux de : Rosenfield [**Rosenfield 1969**], Duda et Hart [**Duda 1972**]. Depuis les années 80, la transformation de Hough est appliquée dans le domaine de la vision par ordinateur et le traitement d'image, elle est devenue un outil très efficace pour la détection des lignes droites et des cercles dans l'image. Un état de l'art sur l'utilisation de cette transformation a été présentédans [**Hart 2009**].

Principe :

On définit une droite dans l'image (figure 4.6). Elle peut être présentée par ses deux coordonnées sur les axes vertical et horizontal (x, y). a et b sont deux paramètres prédéfinis de cette ligne droite :

$$y = ax + b \qquad (21)$$

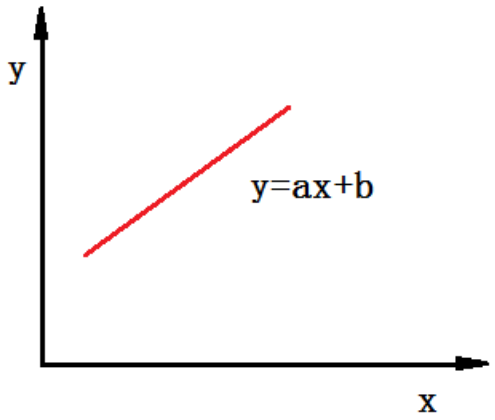


Figure 4.6. Une ligne droite dans le plan d'image
(a : la pente, b : l'ordonnée à l'origine des abscisses)

Sachant que cette ligne est construite par un ensemble de points $P_i(x_i, y_i)$ en respectant l'équation $y_i = ax_i + b$. Selon la transformation de Hough, pour chaque point $P_i(x_i, y_i)$ de l'image, il est possible de définir une droite $b = -ax_i + y_i$ dans l'espace des paramètres appelé espace de Hough (figure 4.7). Par exemple, deux points $P_i(x_i, y_i)$ et $P_j(x_j, y_j)$ sur la droite forment deux droites dans l'espace de Hough :

$$b = -ax_i + y_i \qquad (22)$$

et

$$b = -ax_j + y_j \qquad (23)$$

ces deux droites se croisent au point $N(A, B)$.

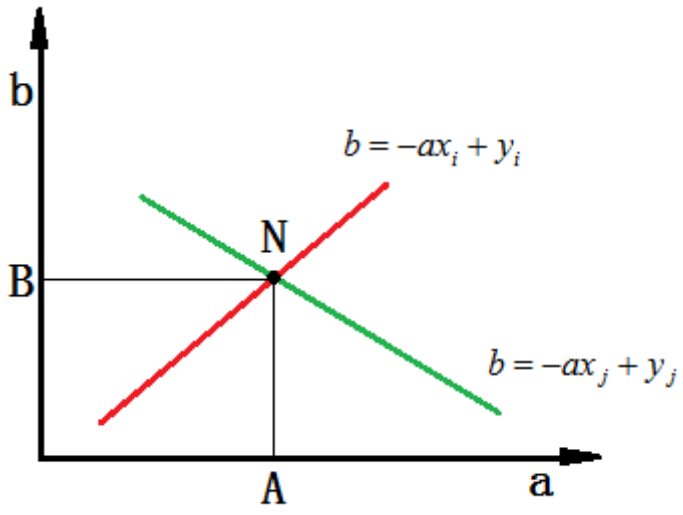


Figure 4.7. Transformation de Hough

En fait, toutes les droites formées dans l'espace de Hough par les points de droite dans l'espace d'image se croisent sur le point $N(A, B)$, les coordonnées (A, B) sont les valeurs des paramètres (a, b) de la droite $y = ax + b$.

En général, l'équation de droite est définie par un paramétrage polaire qui est plus simple à détecter. Donc nous avons la définition d'une droite :

$$\rho = x\cos\theta + y\sin\theta \qquad (24)$$

avec ρ la distance perpendiculaire à la droite de l'origine et θ l'angle entre cette distance et l'axe horizontal.

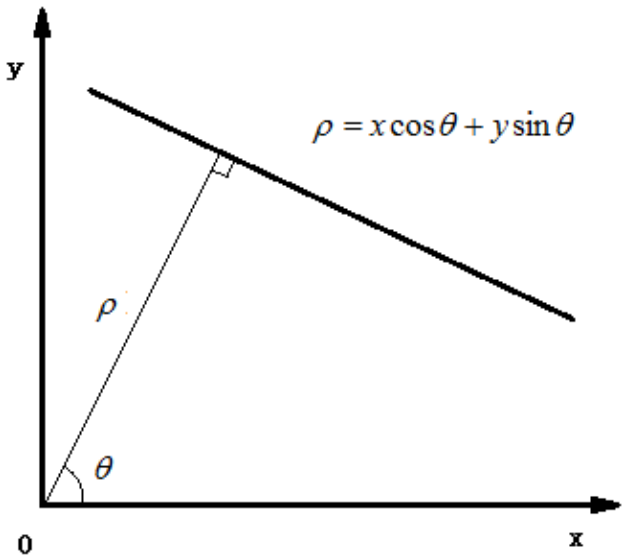


Figure 4.8. Paramétrage polaire d'une droite

Donc si on a un point $P_i(x_i, y_i)$ sur la droite, on peut définir une sinusoïde dans l'espace de Hough : $\rho = x_i\cos\theta + y_i\sin\theta$. Une droite sera représentée par un point $N(\rho_0, \theta_0)$ intersection de l'ensemble des sinusoïdes (figure 4.9).

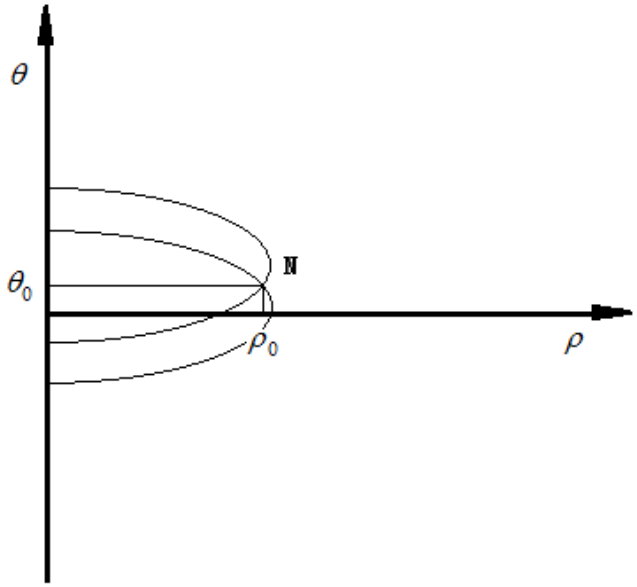


Figure 4.9. Transformation de Hough (paramétrage polaire)

La transformation de Hough présentée au-dessus ne peut que détecter les lignes droites, sa version généralisée [**Ballard 1981**] peut aussi détecter les autres types de courbes plus complexes. Xie et Ji [**Xie 2002**] ont inventé une façon efficace pour détecter l'ellipse en utilisant cette transformation. La transformation de Hough a deux formes : *m à 1* et *1 à m* [**Maitre 1985**] [**Lopez-Krahe 1988**]:

- La version "*m à 1*" :

Soit $B(P)$ l'ensemble des bipoints formés par les points sur la droite,

$$B(P)=\{(P_1,P_2),(P_1,P_3),\ldots,(P_1,P_k),\ldots,(P_k,P_{k+1}),\ldots,(P_{p-1},P_p)\} \qquad (25)$$

dans la version "*m à 1*", pour chaque bipoint (P_m,P_k) de $B(P)$, les coordonnées du point $N(\rho_0,\theta_0)$ peuvent être calculées en utilisant les équations suivantes :

$$\theta=arctg\left(\frac{y_m-y_k}{x_m-x_k}\right), \text{ avec } \left(-\frac{\pi}{2}<\theta\le\frac{\pi}{2}\right) \qquad (26)$$

$$\rho=\frac{\left|x_m y_k-x_k y_m\right|}{\sqrt{(x_k-x_m)^2+(y_k-y_m)^2}} \qquad (27)$$

S'il y a h points sur la droite, donc il aura $h(h-1)/2$ fois comptées par l'accumulateur de ces paramètres.

- La version "*1 à m*" :

S'il y a un point $P_m(x_m, y_m)$ dans l'espace image, il y aura tous les compteurs correspondant à la sinusoïde :

$$\rho = x_m \cos\theta + y_m \sin\theta \qquad (28)$$

Pour h points de la droite, l'accumulateur de ces paramètres compte h fois.

Nous avons utilisé 4 formes géométriques simples (figure 4.10) pour tester cet algorithme.

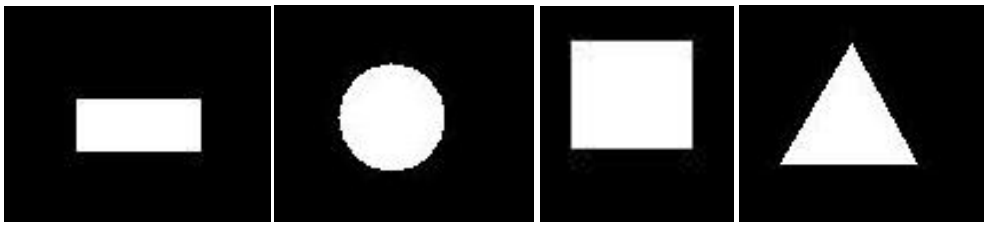
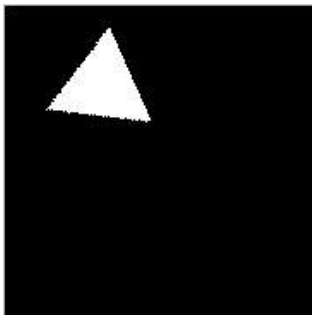
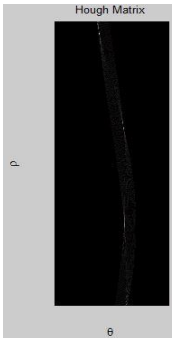
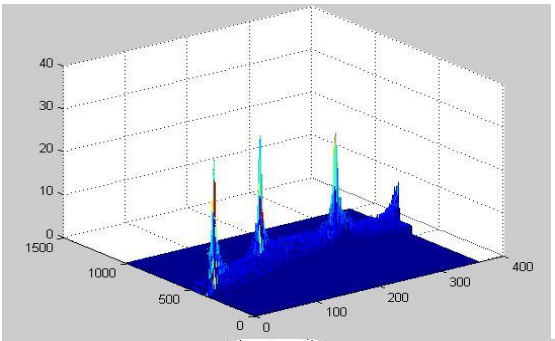
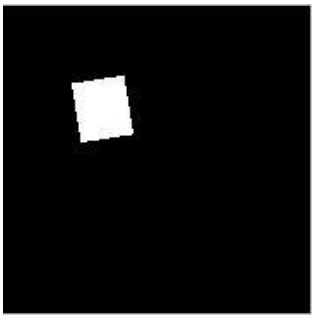
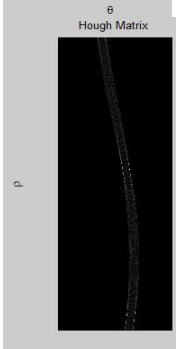
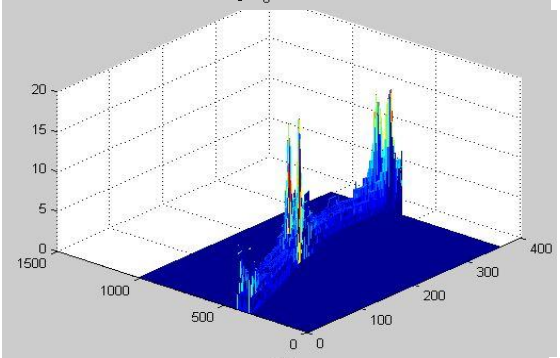
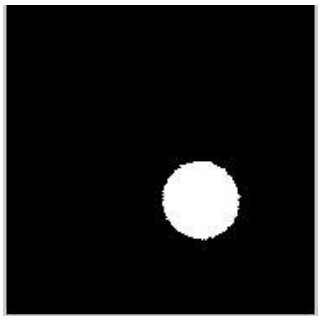
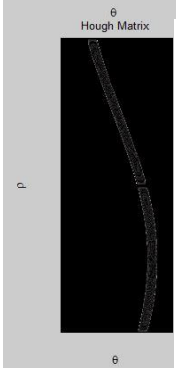
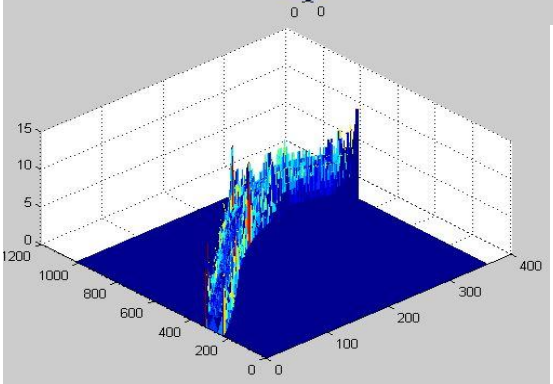
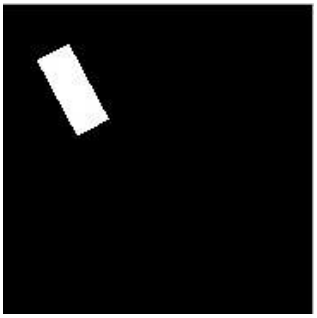
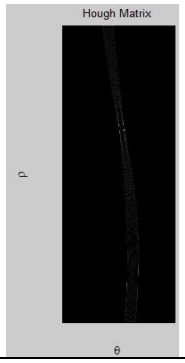
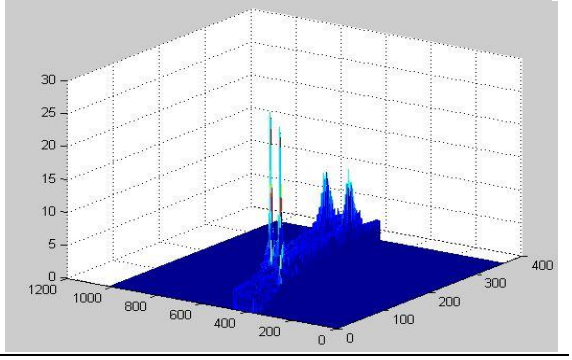


Figure 4.10. Les formes géométriques testées
(Rectangle, cercle, carré, triangle)

Le tableau suivant (tableau 4.1) montre les résultats de la transformation en Hough pour chaque forme testée. Nous avons appliqué sur chaque forme une rotation et translation aléatoires pour vérifier la robustesse de cette méthode.

Tableau 4.1. Les différentes formes et leur transformation en Hough

Forme	Matrice Hough	Matrice Hough en 3D
		
		
		
		

Les résultats du tableau 4.1 ont montré qu'avec la méthode de la transformation de Hough, les lignes droites sont représentées par des pics dans l'espace de Hough, notamment pour les lignes parallèles ; c'est donc plus facile de les détecter [**Lopez-Krahe 1988**]. Mais pour les formes plus complexes, qui se composent de plusieurs lignes, la

procédure de détection sera probablement plus complexe. Pour le cas du cercle, on doit appliquer la transformation générale de Hough, donc la complexité du système augmente. Nous décidons donc d'étudier et d'utiliser les autres méthodes.

4.2.2.2 Moments invariants de HU

Les moments invariants sont utilisés très souvent dans le domaine du traitement de l'image, de la reconnaissance des formes et de la classification. Les moments d'image ne peuvent fournir que les caractéristiques d'un objet qui représentent sa forme. Ils sont invariants par rapport à la translation, l'échelle et la rotation de la forme. En 1962, Hu [**Hu 1962**] a établi, le premier, le fondement mathématique pour les moments invariants en deux dimensions en démontrant leurs applications à la reconnaissance de formes. Dans le papier de Dudani [**Dudani 1997**], les auteurs ont montré qu'il est rapide et fiable d'utiliser ces moments pour détecter les formes d'avions.

- Les moments statistiques :

Sachant qu'une image est considérée comme une matrice $I(N,M)$, $f(x,y)$ est la valeur de gris du point (x,y). Le moment statistique d'ordre $(p+q)$ est défini comme :

$$m_{pq} = \sum_{x=1}^{N}\sum_{y=1}^{M} x^p y^q f(x,y) \qquad (29)$$

- Les moments centrés :

Soit les coordonnées de centre de gravité d'image données par les équations suivantes :

$$\bar{x} = \frac{m_{10}}{m_{00}} \qquad (30)$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \qquad (31)$$

Les moments centrés sont définis comme :

$$\mu_{pq} = \sum_{x=1}^{N}\sum_{y=1}^{M}(x-\bar{x})^p (y-\bar{y})^q f(x,y) \qquad (32)$$

Les moments centrés normalisés sont définis comme :

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{(2+p+q)/2}} \qquad (33)$$

avec $p+q \geq 2$.

- Les moments invariants

En particulier, Hu [**Hu 1962**] a défini sept valeurs calculées en moments centrés normalisés qui peuvent représenter une forme. Ces sept valeurs sont invariantes par rapport à l'échelle, la translation et l'orientation. Les sept moments sont donnés par les équations suivantes :

$$\begin{aligned}
M_1 &= \eta_{20} + \eta_{02} \\
M_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
M_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
M_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \\
M_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \\
M_6 &= (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
M_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{30})^2\right] - \\
&\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]
\end{aligned} \qquad (34)$$

J. Flusser [**Flusser 2000**] et T. Suk [**Flusser 2006**] ont montré que les moments invariants de Hu ne sont pas totalement indépendants, M_3 n'est pas très utile, car elle dépend des autres. En plus, Hu a oublié un moment de l'ordre de trois qui est indépendant des autres :

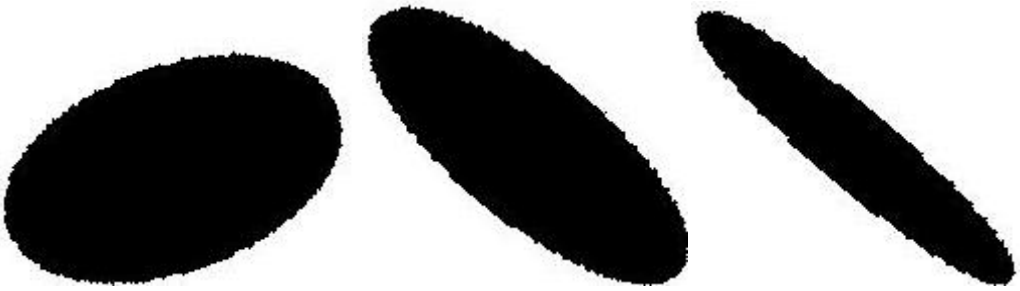
$$M_8 = \eta_{11}\left[(\eta_{30}+\eta_{12})^2 - (\eta_{03}+\eta_{21})^2\right] - (\eta_{20}-\eta_{02})(\eta_{30}+\eta_{12})(\eta_{03}+\eta_{21}) \quad (35)$$

Nous avons testé la méthode Moments invariants, dans le tableau suivant (tableau 4.2), on montre que les résultats sur 3 ellipses différentes, nous calculons les 7 moments invariants pour chaque ellipse avec la rotation, la translation et un changement d'échelle (figure 4.11).



1 2 3

(a) les ellipses initiales



1 2 3

(b) les ellipses avec changement de rotation et d'échelle

Figure 4.11. Les 3 ellipses testées

Tableau 4.2. Les moments invariants des ellipses différentes

	Ellipse 1	Ellipse 1 avec une rotation de 20 ° et un facteur d'échelle 1.5	Ellipse 2	Ellipse 2 avec une rotation de -20 ° et un facteur d'échelle 1.5	Ellipse3	Ellipse 3 avec une rotation de -40 ° et un facteur d'échelle 1.5
M_1	0.1835	0.1835	0.2416	0.2436	0.4329	0.4360
M_2	0.0083	0.0083	0.0330	0.0334	0.1621	0.1647
M_3	$3.0967e^{-9}$	$1.0678e^{-8}$	$7.8478e^{-9}$	$5.0346e^{-8}$	$1.0080e^{-6}$	$5.6050e^{-7}$
M_4	$1.2445e^{-10}$	$8.8322e^{-10}$	$7.2582e^{-10}$	$1.0907e^{-8}$	$7.3707e^{-6}$	$3.7788e^{-7}$
M_5	$-6.5954e^{-21}$	$1.1572e^{-18}$	$-4.8310e^{-19}$	$1.0047e^{-16}$	$6.2908e^{-13}$	$1.7697e^{-13}$
M_6	$-1.1241e^{-11}$	$3.5580e^{-11}$	$-6.2444e^{-11}$	$1.2699e^{-9}$	$2.9647e^{-7}$	$1.4863e^{-7}$
M_7	$-1.6581e^{-20}$	$1.4736e^{-18}$	$-1.4874e^{-18}$	$1.6836e^{-16}$	$7.2254e^{-14}$	$-2.4895e^{-13}$

Les résultats nous ont montré que pour la même ellipse, les moments restent invariants même si celle-ci a eu une rotation ou un changement d'échelle. Mais si l'ellipse est déformée, les moments ne restent plus constants. La méthode des moments invariants est utile pour la détection d'une forme connue (retrouver un template donné dans une image à chercher par exemple). Mais pour identifier le type de forme dans une image inconnue et sans template donné, il n'est pas adapté.

4.2.2.3 Approche de la logique floue

La logique floue (fuzzy logic en anglais), est une forme de logique plurivalente, par rapport à la logique classique (où les variables ne peuvent prendre que des valeurs vraies 1 ou fausses 0). Les variables de la logique floue peuvent avoir un degré d'appartenance variés entre 0 et 1.

La théorie de la logique floue a été établie autour des années 1965 par le professeur Lotfi A. Zadeh [**Zadeh 1965**]. Elle est basée sur sa théorie mathématique des ensembles

flous. Jusqu'à aujourd'hui, la logique floue a été appliquée à beaucoup de domaines. En effet, par rapport à la logique classique, le mode de raisonnement en logique floue est plus intuitif. Il permet de comprendre et de définir plus facilement les règles des phénomènes naturels, imprécis et difficiles.

Nous allons prendre un exemple classique pour expliquer cette méthode, par exemple la conduite automobile à l'approche des feux tricolores :

Tableau 4.3. Règle de conduite automobile à l'approche d'un carrefour contrôlé par des feux tricolores.

Si le feu est rouge	Si ma vitesse est élevée	Et si le feu est proche	Alors je freine fort
Si le feu est rouge	Si ma vitesse est faible	Et si le feu est éloigné	Alors je maintiens ma vitesse
Si le feu est orange	Si ma vitesse est moyenne	Et si le feu est éloigné	Alors je freine doucement
Si le feu est vert	Si ma vitesse est faible	Et si le feu est proche	Alors j'accélère

Le tableau 4.3 donne une situation très fréquente dans notre vie quotidienne, nous pouvons prendre un autre modèle moins flou :

Si le feu est rouge, si ma vitesse dépasse 85,6 Km/H et si le feu est à moins de 62,3 mètres, j'appuie sur la pédale de frein avec une force de 33,2 Newtons !!!

En comparant les deux modèles précédents, nous concluons que le cerveau humain fonctionne plutôt en logique floue (le modèle présenté dans le tableau 4.3).

La logique floue a trois éléments de base, ce sont :

1) les variables linguistiques,

Ce sont les descriptions de certaines situations, d'un phénomène et en général d'une situation avec des adjectifs flous.

2) les fonctions d'appartenance

La logique floue affecte à chaque variable des degrés d'appartenance variés entre 0 et 1, qui sont appelés "degré d'appartenance" μ . Les ensembles flous sont décrits par leur fonction d'appartenance. Les fonctions d'appartenance ont diverses formes selon leur définition (triangulaire, trapézoïdale, sigmoïdes...) (figure 4.12)

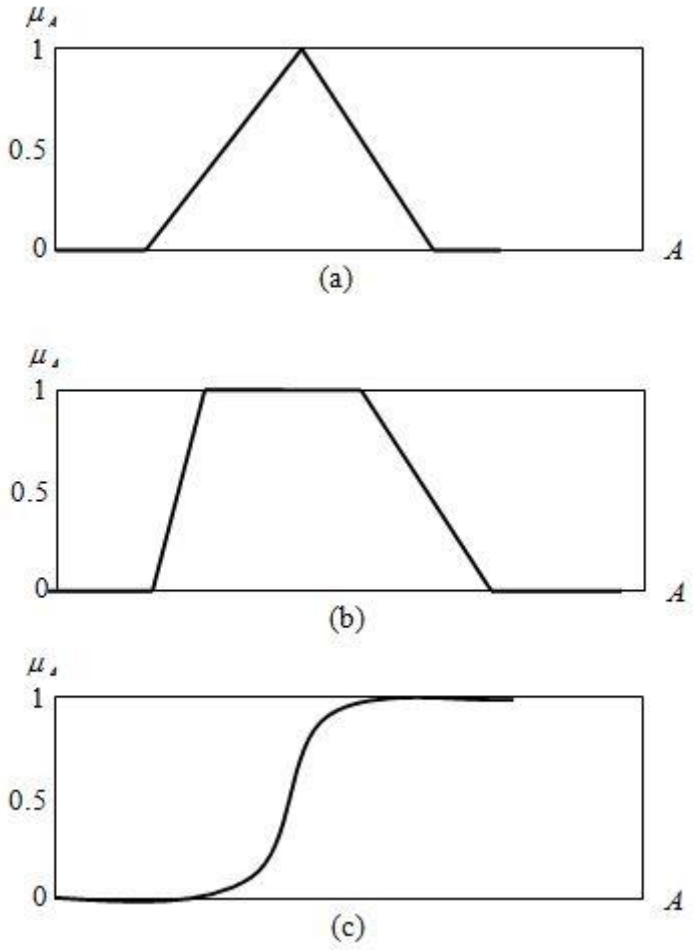


Figure 4.12. Fonctions d'appartenance.

(a : triangulaire, b : trapézoïdale, c : sigmoïdes)

3) les déductions aux inférences

Les différentes variables sont liées par des règles qui permettent de tirer des conclusions. Les règles peuvent être sous la forme de :

- si condition 1 alors action1 ou...
- si condition 2 alors action 2 ou...
- si condition *n* alors action *n* .

La condition peut avoir plusieurs variables qui sont liées par des opérateurs OU ou ET.

La logique floue est utilisée aussi dans le domaine de la reconnaissance des formes. Fonseca et Jorge [**Fonseca 2000**] ont établi une nouvelle méthode de détection de formes géométriques dessinées à la main. Ce travail est une extension du travail de Kimura, Apte et Van [**Kimura 1993**] qui avait établi une méthode simple de reconnaissance de formes en utilisant certaines caractéristiques géométriques.

Selon la méthode de Fonseca, la première étape consiste à extraire les propriétés caractéristiques des formes, ensuite à utiliser un filtrage pour supprimer les formes non utiles, et la troisième étape utilise la logique floue [**Bezdek 1992**] pour identifier les formes.

Cette méthode peut reconnaître certaines formes géométriques simples comme les triangles, les rectangles, les cercles, les ellipses, etc. Le modèle géométrique et les caractéristiques utilisées par Fonseca sont présentés dans la figure 4.13 et le tableau 4.4 :

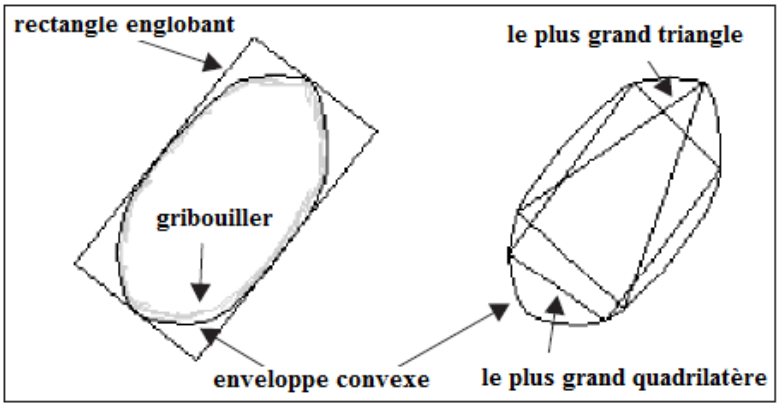


Figure 4.13. Modèle de polygone pour estimer les caractéristiques

(Extraite de [**Fonseca 2000**])

Tableau 4.4. Caractéristiques utilisées par les formes.

(Pch^2 : périmètre au carré, Ach : surface de l'enveloppe convexe, Her : hauteur du rectangle englobant, Wer : largeur du rectangle englobant, Tl : longueur du trait, Aer : surface du rectangle englobant, Alq : surface du plus grand triangle, Plt, périmètre du plus grand triangle, Hm : mouvement horizontal, Wbb : largeur du rectangle englobant, Ns : nombre de traits, Vm : mouvement vertical, Hbb : hauteur du rectangle englobant)

(Extrait de [**Fonseca 2000**])

	Hollowness	Pch^2/Ach	Her/Wer	Tl/Pch	Ach/Aer	Alq/Ach	Plt/Pch	Hm/Wbb	Alt/Ach	Alq/Aer	Plq/Pch	Alt/Alq	Pch*Ns/Tl	Vm/Hbb	Alt//Aer
Ligne			X												
Flèche	X		X		X		X				X				
Triangle	X						X		X						
Rectangle	X				X					X					
Cercle	X	X													
Ellipse	X	X				X									
Losange	X	X				X				X		X			
Supprimer	X		X	X	X										
Ligne ondulée			X	X				X							
Copie		X		X				X			X			X	
Mouvement		X		X			X		X				X		
Défaire	X	X		X											X

Pour que la méthode de Fonseca puisse être plus adaptée à notre travail, nous avons proposé un nouveau choix des caractéristiques pour que le système de reconnaissance soit plus simple, plus rapide et plus efficace. Dans le tableau 4.5, nous présentons les caractéristiques utilisées dans notre travail.

Tableau 4.5. Valeurs des paramètres du vecteur caractéristique correspondant à différentes formes (Extrait de [**Chen 2014a**]).

(X : n'importe quelle valeur)

Forme	**Paramètres composant le vecteur caractéristique**					
	Nombre de sommets	***excentricité***	A/A_{ch}	A/A_b	P_{ch}^2/A_{ch}	P/P_{ch}
Triangle	3	X	1	0.5	X	1
Carré	4	0	1	1	X	1
Rectangle	4	X	1	1	X	1
Losange	4	X	1	0.5	X	1
Cercle	0	0	1	$\pi/4$	4π	1
Ellipse	0	X	1	$\pi/4$	X	1
Etoile	10 (5 sur l'enveloppe convexe)	X	0.6	0.4	X	X

Les caractéristiques sont les paramètres mathématiques suivants :

- Le nombre de sommets : qui est donné par le détecteur de Harris [**Harris 1988**]. La figure 4.14 représente les sommets de quelques formes géométriques simples. Ces sommets sont représentés dans cette figure par des croix rouges.
- Le rapport entre la surface concave et la surface convexe A/Ach : dans la figure 4.15, nous représentons la partie convexe et la partie concave de quelques formes géométriques.
- L'excentricité mathématique : qui est un paramètre réel positif caractéristique d'une courbe type ellipse.
- Le rapport entre la surface concave et la surface de l'enveloppe englobante A/Ab : la figure 4.16 représente la différence entre ces deux entités.
- Le rapport entre le périmètre de l'enveloppe convexe élevé au carré et sa surface Pch2/Ach.

- Le rapport entre le périmètre de la forme et le périmètre de son enveloppe convexe P/Pch.

Les figures suivantes présentent quelques notions utilisées pour la détermination des paramètres [**Chen 2014b**]:

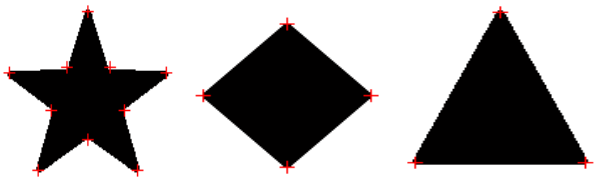


Figure 4.14. Représentation des sommets de différentes formes.

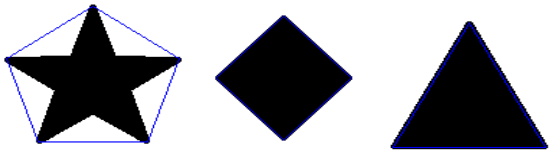


Figure 4.15. Différence entre les surfaces concaves et convexes de différentes formes.

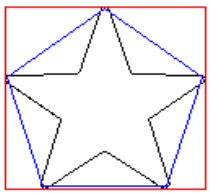
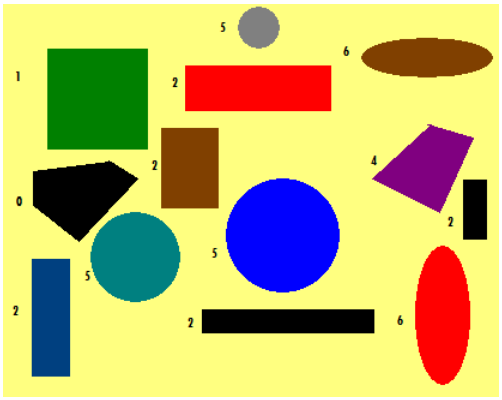


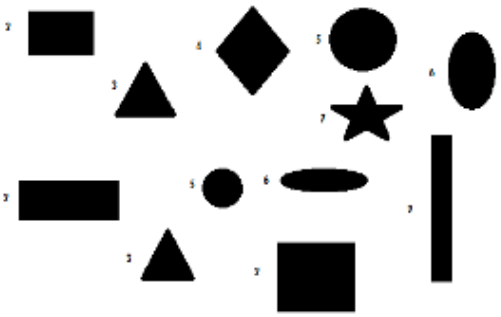
Figure 4.16. Représentation de l'enveloppe convexe et de l'enveloppe englobante.

Une fois le vecteur caractéristique calculé pour chaque forme, la reconnaissance automatique de la forme se fait en appliquant l'algorithme de classification floue.

Nous appliquons la méthode de la logique floue pour réaliser la détection des formes selon des caractéristiques choisies, et les résultats sont donnés dans les figures suivantes :



(a) Les formes testées en couleur



(b) Les formes testées en noir et blanc

0	Inconnu
1	Carré
2	Rectangle
3	Triangle
4	Losange
5	Cercle
6	Ellipse
7	Etoile

(c) L'index de chaque forme

Figure 4.17. Les résultats de détection des formes avec la méthode de la logique floue.

Les résultats montrent que les formes que nous avons testées sont bien identifiées par cette méthode.

4.2.3 Conclusion

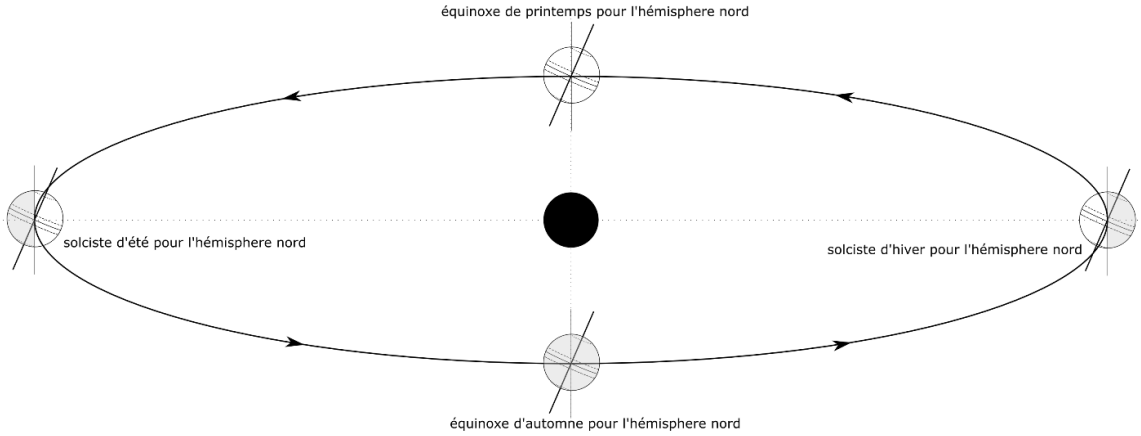
Sachant que dans les images que nous avons à traiter, les formes sont assez simples (comme les formes montrées dans la figure 4.17). Chaque type de formes a certaines caractéristiques communes, et d'autres différentes des autres types de formes. Par exemples, les triangles ont toujours trois sommets, tandis que les rectangles en ont quatre. La logique floue est donc considérée comme une méthode assez performante. Si on veut aller plus loin pour traiter les formes plus complexes, une combinaison des 3 méthodes

mentionnées peut assurer les détections. Mais pour l'instant, la logique floue est retenue pour réaliser la détection et l'identification des formes considérées dans notre travail.

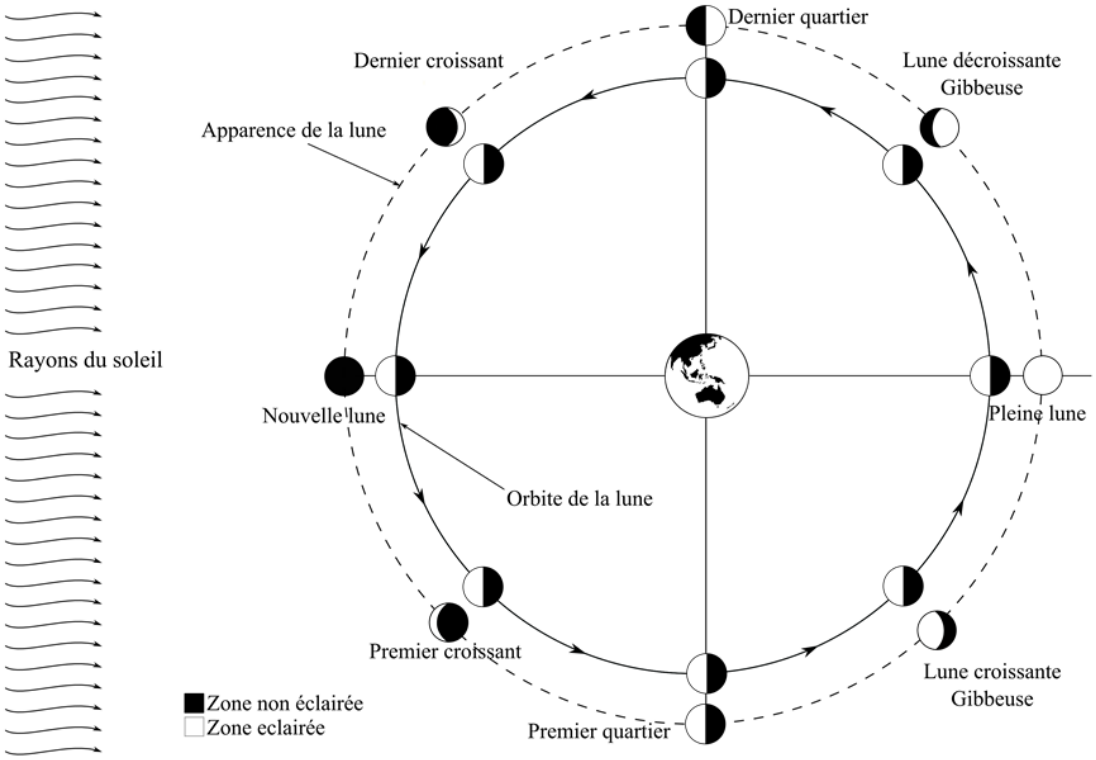
4.3 Étape de détection de flèches

4.3.1 Introduction et problématique

Après plusieurs tests sur les différentes images, nous avons noté que dans certaines d'entre elles, il y a des éléments de petite taille, mais qui comportent des informations très importantes, comme les petites flèches qui se trouvent dans les images montrées ci-dessous. Dans ces deux images, les petites flèches indiquent la direction du mouvement des planètes qui est effectivement une information importante pour une bonne compréhension de l'image. Nous avons donc travaillé sur la détection et l'agrandissement des flèches pour qu'elles puissent être lisibles par des personnes non-voyantes sur les systèmes tactiles.



(a) Image ‘4 saisons’



(b) Image ‘phases de lune’

Figure 4.18. Images comprenant des petits symboles de flèches

Comme les flèches à détecter sont petites (hauteur et largeur inférieures à 10 pixels) et qu’en plus ces flèches sont de types variables, les méthodes de détection de formes qui ont été mentionnées auparavant ne sont plus adaptées à ce type de forme. Nous proposons alors d’utiliser d’autres méthodes.

Deux méthodes sont testées dans cette partie du travail : l'appariement des images via les points d'intérêt et via la corrélation des images. Ces deux méthodes utilisent une même image de template qui modélise la flèche à détecter.

4.3.2 Méthodes testées

4.3.2.1 L'appariement des images via les points d'intérêt

Du point de vue psycho-physique, un point d'intérêt est un point d'image qui se distingue dans son voisinage spatial local par sa luminosité. En général, les points d'intérêt correspondent à des doubles discontinuités de la fonction d'intensité. Ce sont par exemple : les coins, les jonctions en T, les points de fortes variations de texture ou le point de changement de courbure (figure 4.19).



Figure 4.19. Exemples de types de points d'intérêt.
(Extraite de [**Chen 2014c**])

Il existe de nombreuses méthodes de détection de points d'intérêt. Elles peuvent être classées en trois grandes familles :

- Approches fondées sur les contours ;
- Approches fondées sur les intensités ;
- Approches à base de modèles.

Les approches de la deuxième famille sont celles qui sont utilisées le plus souvent.

Les détecteurs des points d'intérêt les plus populaires sont le détecteur de Moravec [**Moravec 1980**], le détecteur de Harris et de Stephens [**Harris 1988**], le détecteur de SUSAN (Smallest Univalue Segment Assimilating Nucleus) [**Smith 1997**] et le détecteur SIFT (Scale-invariant feature transform) [**Lowe 2004**]. Les études réalisées par E. Pissaloux et Y. Chen [**Pissaloux 2010**] sur la comparaison des performances des détecteurs précédemment cités ont montré que, en conciliant les trois critères correspondants à la robustesse, la rapidité et la facilité d'implémentation, le détecteur d'Harris s'avère le plus approprié. De ce fait, nous avons choisi ce détecteur pour tester la première méthode de détection des flèches.

Une fois que les points d'intérêt sont bien détectés, on peut utiliser les informations extraites de chaque point pour faire la mise en correspondance entre l'image du template et l'image originale. Dans notre travail, nous avons utilisé un descripteur SURF pour extraire les informations de chaque point trouvé.

SURF (Speeded Up Robust Features), proposé par H.Bay, T.Tuytelaars et L.V.Gool en 2006 [**Bay 2006**] puis dans une version révisée en 2008 [**Bay 2008**], est un détecteur d'informations caractéristiques qui est utilisé souvent pour la détection d'objet. On peut dire que SURF est une version plus rapide et robuste du détecteur SIFT.

En pratique, nous avons appliqué cette méthode sur deux images astronomiques parmi celles que nous avons vues ci-dessus. L'image de template utilisée est représentée par la figure suivante :



Figure 4.20. Image de template utilisée pour l'image '4 saisons'

Nous appliquons le détecteur de Harris et le descripteur SURF sur les images originales et de template, et en même temps, pour vérifier si la méthode implémentée est

correcte, nous introduisons aussi une autre image de template qui comporte plus d'éléments (figure 4.21). Les résultats sont donnés dans les figures suivantes [**Chen 2015**] :

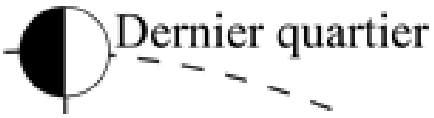


Figure 4.21. Image de template de vérification.

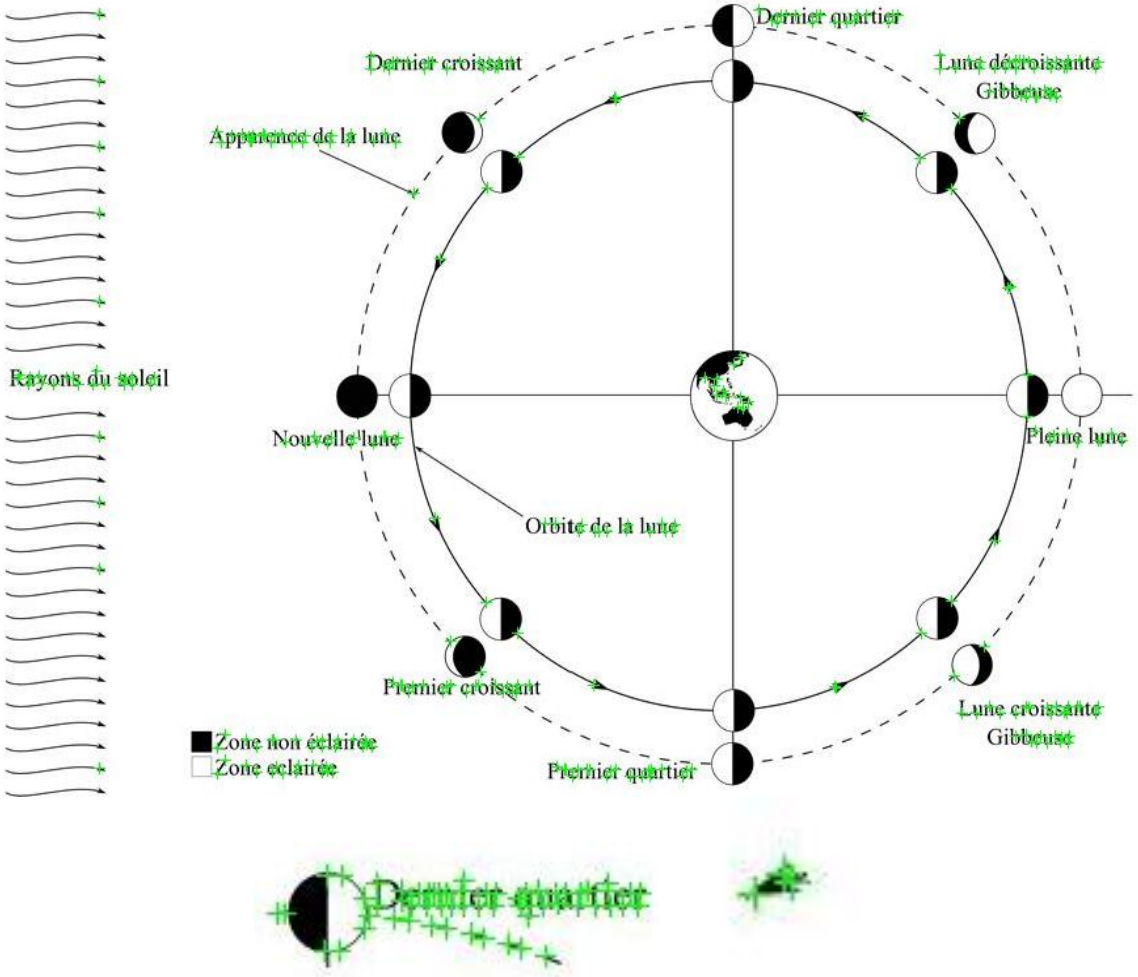


Figure 4.22. Points d'intérêt trouvés sur l'image originale et celles des templates.

En comparant les résultats (figure 4.23 et 4.24), la méthode a bien trouvé la position de l'image du template de vérification, mais n'a pas trouvé la position de la flèche. La raison de ce problème est que la flèche a une taille très petite. Le détecteur ne peut trouver que très peu de points d'intérêts (3), et ceux-ci ne sont pas suffisants pour localiser la

position de l'image template dans l'image originale. Si une image de template comporte suffisamment d'éléments et a aussi une taille assez grande pour que les détecteurs puissent trouver assez de points d'intérêt, cette méthode d'appariement donnera alors un très bon résultat.

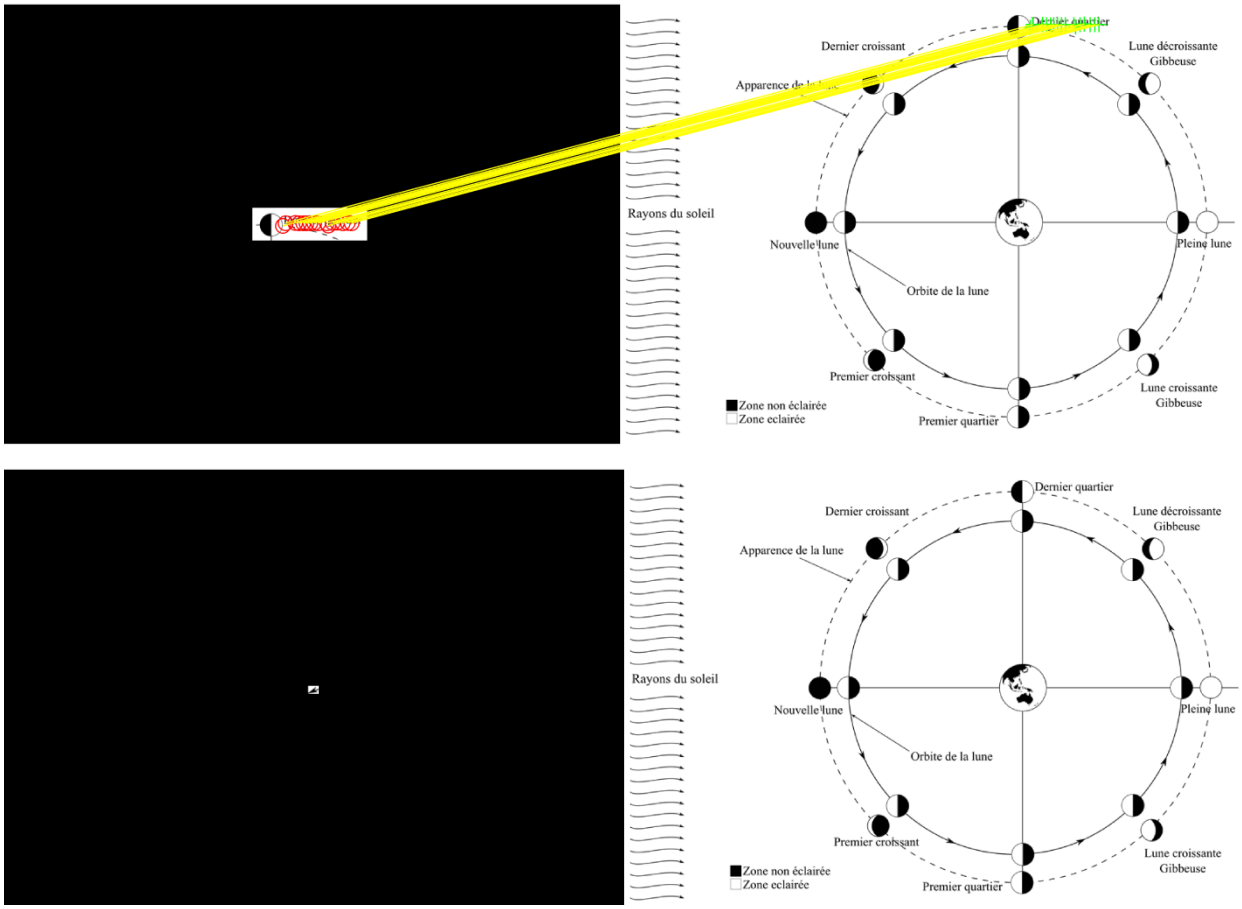


Figure 4.23. Exemple 1: Résultat de l'appariement

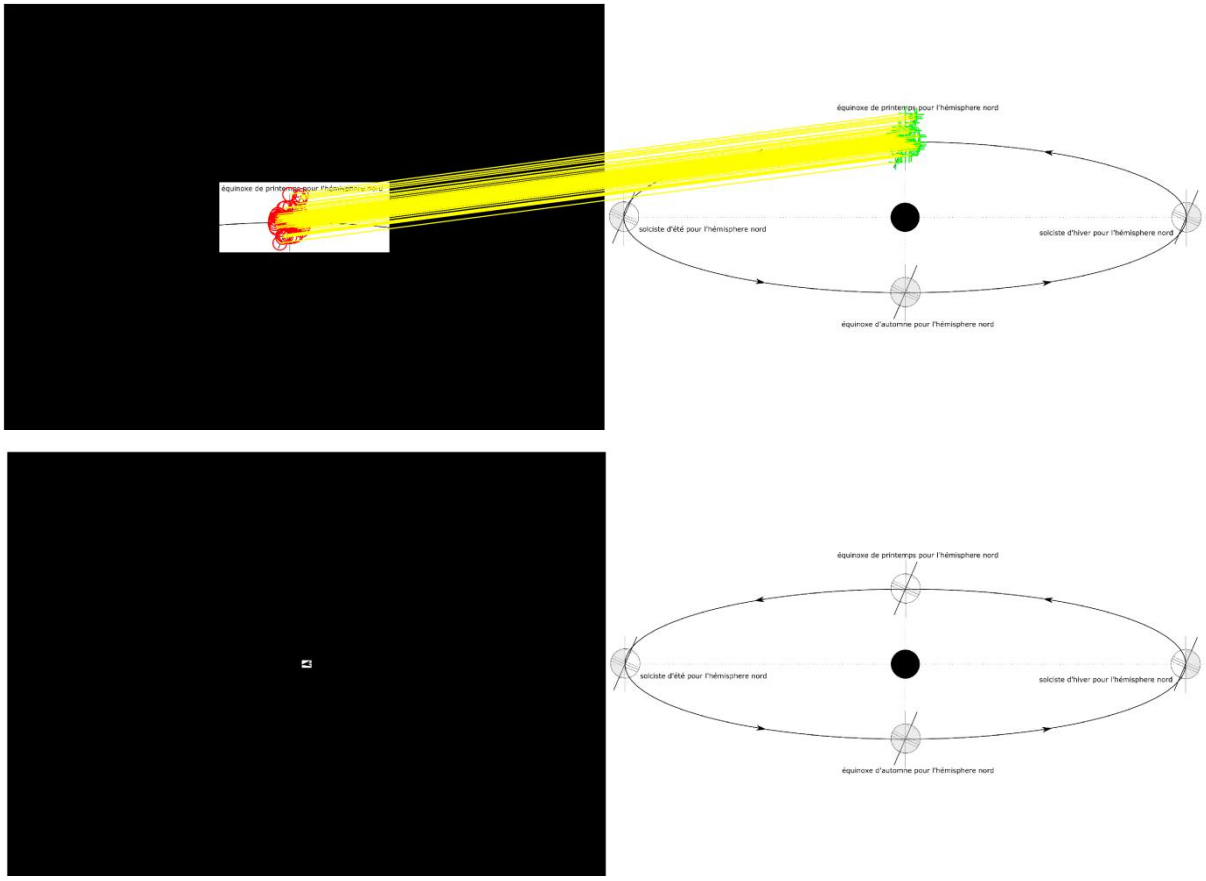


Figure 4.24. Exemple 2: Résultat de l'appariement

4.3.2.2 L'appariement des images via la corrélation

La corrélation est une mesure qui calcule le degré de ressemblance entre deux ensembles de données. Avant l'appariement, il faut déterminer la fenêtre de corrélation qui est, dans notre cas, une image de template. Ensuite, cette fenêtre de corrélation parcourt l'image originale. Chaque point de l'image originale obtient un score de corrélation. Après calcul, le point qui donne le plus grand score est considéré comme le point de correspondance de l'image du template dans l'image originale (figure 4.25).

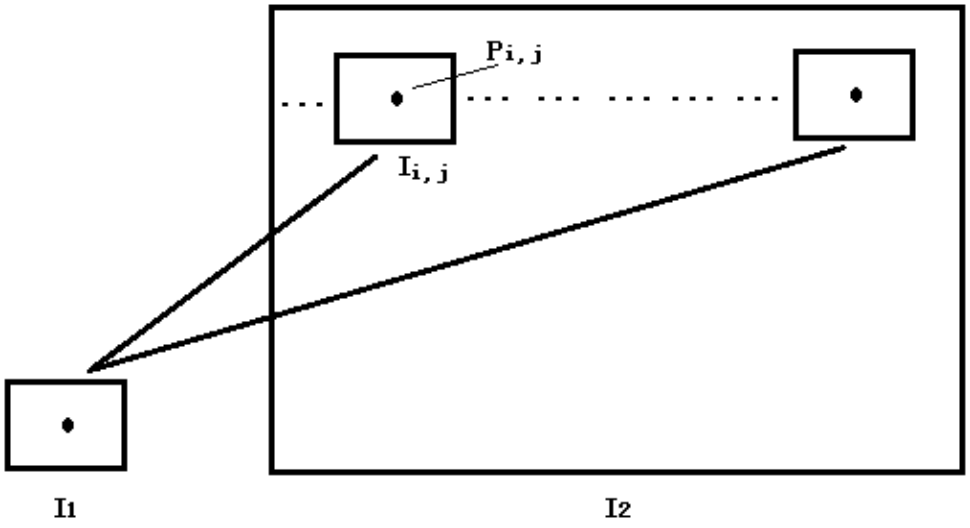


Figure 4.25. Corrélation entre le template I_1(m x n) et l'image originale I_2(M x N).

($I_{i,j}$: fenêtre à corréler de l'image originale)

($P_{i,j}$: point du centre de cette fenêtre. 0<i<=M, 0< j <= N.)

Dans ce qui suit, les termes f_1 et f_2 sont des vecteurs contenant respectivement les niveaux de gris des pixels du template I_1 et de la fenêtre à corréler de l'image originale $I_{i,j}$; on note $f_1 \cdot f_2$ le produit scalaire de f_1 par f_2, et $\bar{f}$ la moyenne de f :

$$f_1 \cdot f_2 = \sum_{p=1}^{m} \sum_{q=1}^{n} f_1^{p,q} f_2^{p,q} \qquad (36)$$

$$\bar{f} = \frac{1}{m \times n} \sum_{p=1}^{m} \sum_{q=1}^{n} f^{p,q} \qquad (37)$$

où $f_1 = \left(\cdots f_1^{p,q} \cdots \right)$ et $f_2 = \left(\cdots f_2^{p,q} \cdots \right)$.

On a, pour le point $P_{i,j,}$ le score de corrélation croisée:

$$CC(f_1, f_2) = f_1 \cdot f_2 \qquad (38)$$

Les valeurs de $CC(f_1, f_2)$ appartiennent donc à l'intervalle [0 ;+∞[.

Il existe plusieurs membres dans la famille des corrélations croisées :

- Corrélation croisée normalisée NCC (Normalized Cross-Correlation).

$$NCC(f_1, f_2) = \frac{f_1 \cdot f_2}{\|f_1\| \|f_2\|} \qquad (39)$$

où $\|f\|_P = \left(\sum_{p=1}^{m} \sum_{q=1}^{n} \left| f^{p,q} \right|^P \right)^{1/P}$ et $\|f\| = \|f\|_2$.

Les valeurs de cette mesure appartiennent à l'intervalle [0 ; 1].

- Corrélation croisée centrée et normalisée à moyenne 0 ou ZNCC (Zero mean Normalized Cross-Correlation).

$$ZNCC(f_1, f_2) = \frac{(f_1 - \bar{f}_1) \cdot (f_2 - \bar{f}_2)}{\|f_1 - \bar{f}_1\| \|f_2 - \bar{f}_2\|} \qquad (40)$$

Les valeurs de $ZNCC(f_1, f_2)$ appartiennent à l'intervalle [-1 ; +1].

En comparant les invariances des mesures, les études que nous avons réalisées ont donné le résultat (tableau 4.6) qui a montré que la mesure ZNCC possède plus d'invariance que les autres :

Tableau 4.6. Comparaison des invariances de certaines mesures de corrélation

(Extrait de [**Pissaloux 2010**])

		CC	NCC	ZNCC
Invariance de type gain+biais	M(a**f1** + b, c**f2** + d) = M(**f1**, **f2**)	-	-	X
	M(a**f1** + b, a**f2** + b) = M(**f1**, **f2**)	-	-	X
Invariance de type biais	M(**f1** + a**, f2**+b) = M(**f1 , f2**)	-	-	X
	M(**f1** + a**, f2**+a) = M(**f1 , f2**)	-	-	X
Invariance de type gain	M(a**f1**, b**f2**) = M(**f1**, **f2**)	-	X	X
	M(a**f1**, a**f2**) = M(**f1**, **f2**)	-	X	X

Les études comparatives réalisées par Ashwanden et Guggenbühl [**Aschwanden 1992**] ont prouvé que la ZNCC est la mesure la plus robuste aux changements linéaires de luminosité et au bruit. Dans notre étude de 2010 [**Pissaloux 2010**], nous avons aussi montré que la mesure ZNCC est une mesure très efficace pour une détection des points. De ce fait, la ZNCC est retenue comme mesure de notre travail.

Nous testons cette méthode sur les deux images. Dans l'image originale, les flèches sont de différents types et directions. Pour que toutes ces flèches puissent être détectées, nous avons appliqué une base de données de flèches de template pour prendre en compte les différents styles et appliqué à ces flèches de template des rotations dans l'intervalle [0,360 [avec un petit pas de 5 ° pour prendre en compte toutes les directions. Les résultats obtenus sont données dans les figures suivantes :

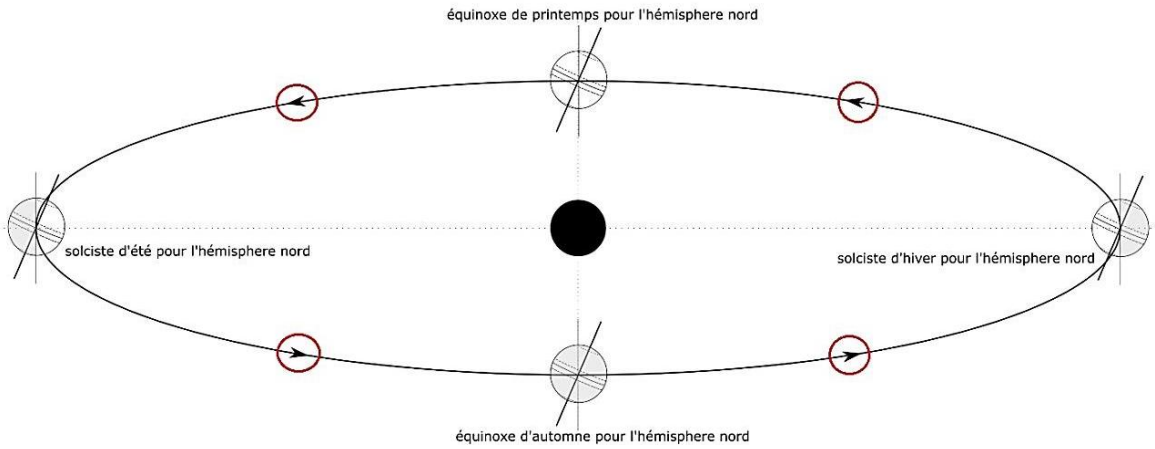


Figure 4.26. Les flèches détectées avec la corrélation.

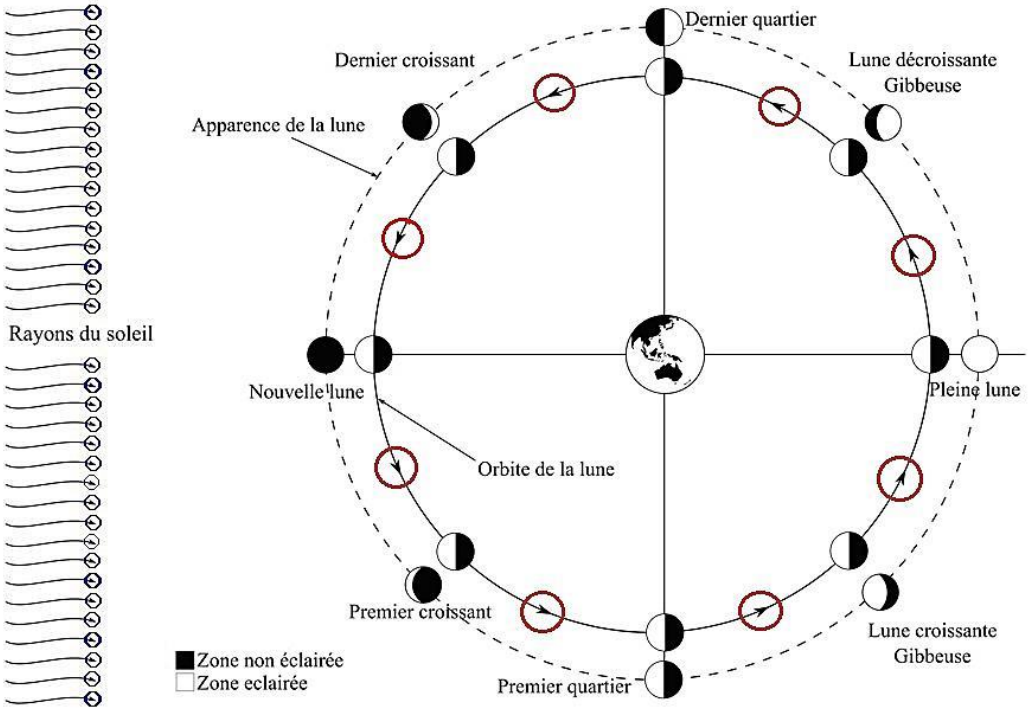


Figure 4.27. Autre exemple de détection de flèches avec la corrélation.

Les résultats ci-dessus ont montré qu'avec la corrélation, les positions des flèches peuvent être efficacement trouvées dans l'image originale [**Chen 2015**]. C'est une méthode possible pour résoudre la problématique proposée, mais pendant les tests que nous avons réalisés, nous nous sommes rendu compte que dans le cas de détection de plusieurs flèches

dans une image, il faut utiliser un seuil de corrélation. Il est évident que pour avoir un bon résultat, le seuil doit être bien défini dès le départ. Pour l'instant, le seuil est défini manuellement sur la base des tests que nous avons réalisés. Par la suite, nous chercherons à développer une méthode automatique pour la détermination de ce seuil.

Une fois les flèches bien détectées, nous agrandissons leur taille pour les rendre plus lisibles ou détectables tactilement (figure 4.28 et 4.29) [**Chen 2015**].

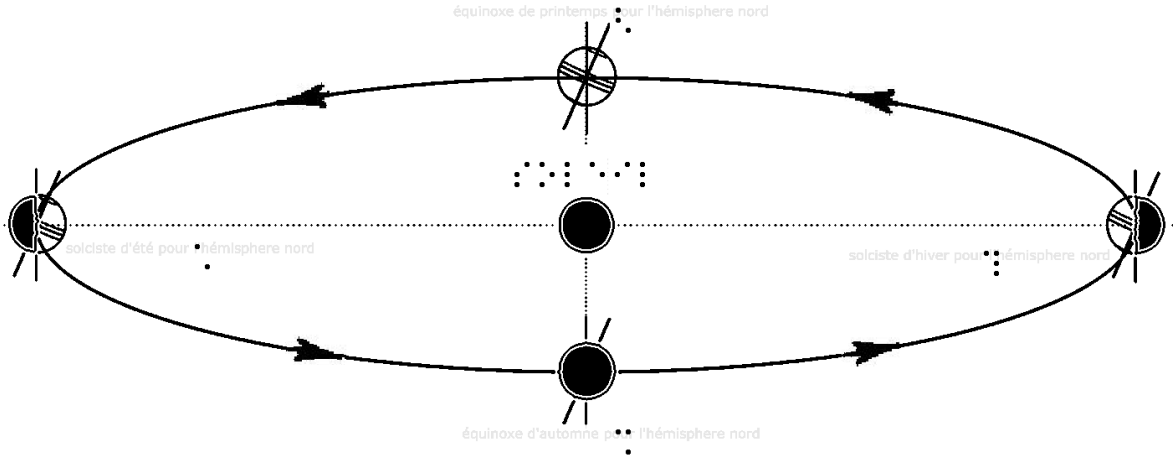


Figure 4.28. Exemple 1 d'une image tactile contenant des flèches accessibles.

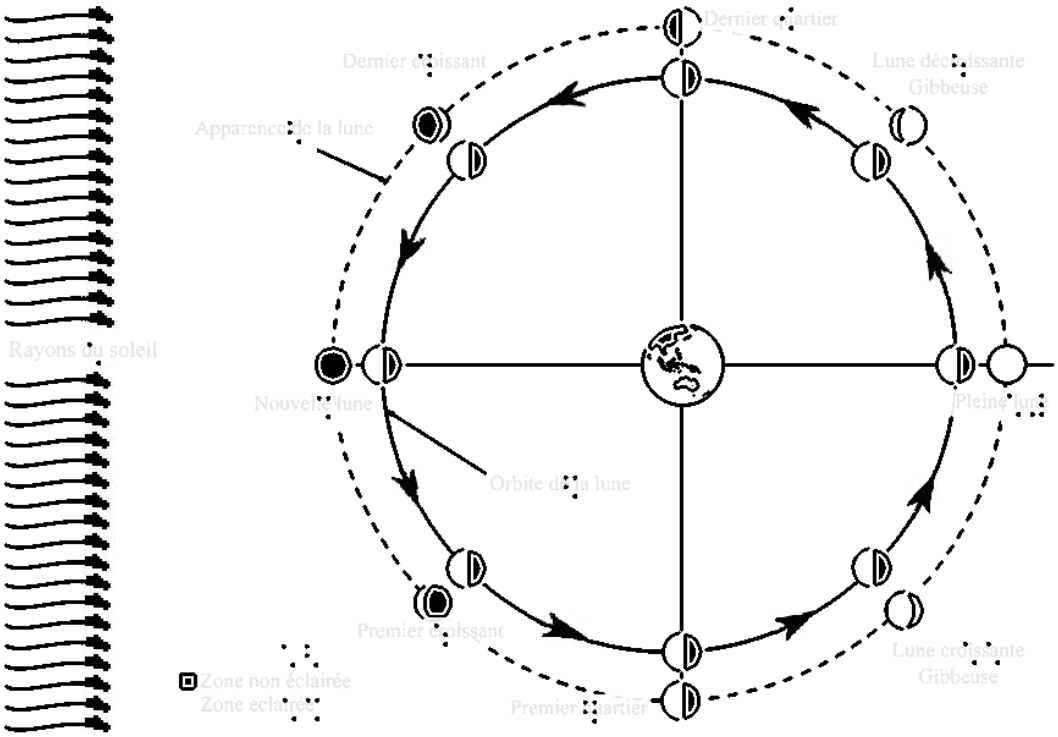


Figure 4.29. Exemple 2 d'une image tactile contenant des flèches accessibles.

4.4 Étape de la reconnaissance de texte

Les textes dans une image nous apportent des informations très importantes. Les textes décrivent les différents objets ou donnent une information complète de l'image. Ils sont les éléments élémentaires et nécessaires pour qu'une image puisse être facile à comprendre. Dans notre travail, une partie très importante est d'effectuer une transcription automatique des textes vers le Braille. La méthodologie adoptée pour effectuer cette transcription se décompose en 3 étapes : la localisation des zones de texte, la reconnaissance du texte par un outil ROC (Reconnaissance optique de caractères), la transformation du texte en braille.

4.4.1 Introduction de la ROC

La reconnaissance optique de caractère (ROC, OCR en anglais- Optical Character Recognition) est une technologie informatique qui peut réaliser la traduction d'image de textes imprimés ou dactylographiés en fichiers de texte.

La reconnaissance de caractère (du texte) est une partie du domaine de la reconnaissance des formes. La première machine de lecture ROC a été installée chez Reader's Digest en 1954 pour convertir les rapports de ventes dactylographiées en carte perforée pour l'introduire dans l'ordinateur. Les systèmes ROC commerciaux apparaissant sur la période 1960-1965 peuvent être appelés la première génération de la ROC. Les machines de la deuxième génération sont apparues dans le milieu des années 1960 et au début des années 1970. Ces systèmes ont été capables de reconnaître régulièrement les caractères imprimés. Pour la troisième génération de systèmes, apparaissant au milieu des années 1970, les documents de mauvaise qualité et les caractères écrits à la main constituaient un réel défi. Aujourd'hui, la ROC est devenu un outil largement utilisé dans les travaux de bureautique. Mais jusqu'à une date récente, le fonctionnement des systèmes ROC les plus performants étaient peu connus à cause de leur protection par le secret industriel.

Un système ROC typique est constitué de plusieurs composants :

- le prétraitement d'image ;

  Le but de cette étape est d'améliorer la qualité de l'image.

- la localisation et la segmentation ;

  Cette étape est un processus nécessaire qui détermine le contenu d'une image. Il permet de localiser et de séparer les zones qui contiennent des textes des autres éléments de l'image.

- la reconnaissance des caractères ;

  Cette étape a deux phases : la première phase est l'extraction de caractéristiques ; la seconde est basée sur les caractéristiques trouvées, et consiste à classer, à identifier chaque caractère et à le mettre dans une bonne classe.

- Post-traitement ;

  La dernière étape est d'utiliser les méthodes linguistiques et contextuelles pour minimiser le nombre d'erreurs de reconnaissance.

Nous orientons notre travail sur l'étape de localisation parce que les images qui seront traitées comportent peu de texte par rapport aux autres éléments (les régions, les formes, etc.). Les éléments non-textuels ont trop d'influence sur les résultats de reconnaissance si on applique directement une ROC sur l'ensemble d'une image. Pour améliorer les résultats, nous avons décidé de localiser et d'extraire les zones de texte avant d'utiliser une ROC. Les méthodes de localisation de texte sont présentées ci-dessous.

4.4.2 Méthodes testées

4.4.2.1 Méthode basée sur les ondelettes

En 1983, Jean Morlet a proposé pour la première fois le nom 'ondelette' ('wavelet' en anglais). «La transformée en ondelette est un outil qui découpe les données, les fonctions ou les opérateurs en composantes fréquentielles suivant une résolution adaptée à l'échelle. »[**Truchetet 1998**].

De manière plus générale, si $f(x)$ est une fonction réelle de variable réelle, sa transformée en ondelette $F(a,b)$ peut être exprimée avec l'équation suivante :

$$F(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \psi^*_{(a,b)}(x) dx \qquad (41)$$

où ψ^* est le conjugué complexe de la fonction donnée $\psi_{(a,b)}(x)$ qui est obtenue par translation et dilatation d'une fonction appelée ondelette mère $\Psi(x)$:

$$\psi_{(a,b)}(x) = \Psi\left(\frac{x-b}{a}\right) \qquad (42)$$

où a donne l'échelle et b est le paramètre de translation.

Il existe plusieurs façons de classer les types de transformation en ondelette, une d'entre elles est basée sur l'orthogonalité des ondelettes. La transformée en ondelette orthogonale est notamment utilisée dans la compression des données numériques avec ou sans perte. Pour la transformation continue, les ondelettes utilisées ne sont pas orthogonales et les données obtenues sont hautement corrélées.

Dans le domaine du traitement d'image numérique, les applications des ondelettes sont variées : elles peuvent être utilisées pour décompresser et compresser les images [**Villasenor 1995**] (figure 4.30), ou restaurer des images détériorées par le bruit (figure

4.31). Dans le travail de M. H. Sahbani et K. Hamrouni, l'image est segmentée à l'aide de la transformée en ondelette et de l'algorithme des C-moyennes floues [**Sahbani 2005**].



(a) Image originale : «Lena» (1024x1024, 306KB)



(b) Compression de l'image (1024x1024, 101KB)

Figure 4.30. Compression d'image par la transformée en ondelette.



(a) Image bruitée



(b) Image débruitée (niveau 3)

Figure 4.31. Débruitage d'une image par la transformée en ondelette

Pour la détection et la localisation du texte, Gllavata et Ewerth [**Gllavata 2004**] ont proposé dans leur travail l'utilisation d'un coefficient de haute fréquence d'ondelette. Li,

Doermann et Kia [**Li 2000**] ont aussi montré une application de la transformée en ondelette pour la détection de texte dans une vidéo.

Selon la méthode utilisée par Gllavata, l'application du filtrage par ondelette sur une image consiste en un processus de filtrage horizontal, et ultérieurement de filtrage vertical. Les zones de texte sont généralement caractérisées par des contours à fort contraste ; les coefficients de grande valeur se trouvent dans les sous-bandes de haute fréquence. Le coefficient choisi par Gllavata et Ewerth est la déviation standard de l'histogramme respectif de chaque sous-bande image. Il est présenté par l'équation suivante :

$$stdev_{block} = \sqrt{\sum_{i=1}^{K} (f(i) - avg_{f(I)})^2 \Big/ K - 1}, \begin{cases} si & H_i > 0 : f(i) = i \\ & else : f(i) = 0 \end{cases} \qquad (43)$$

où H_i est la valeur de histogramme, $avg_{f(I)}$ est la valeur moyenne de $f(i)$, K est le nombre de i.

Le choix de ce paramètre caractéristique est basé sur l'étude de Lia et Gray [**Lia 1998**] qui ont montré que pour les zones de non-texte, les coefficients d'ondelette suivent une distribution de Laplace, tandis que pour les zones texte, ces coefficients sont dispersés et concentrés sur un petit nombre de valeurs discrètes (figure 4.32).

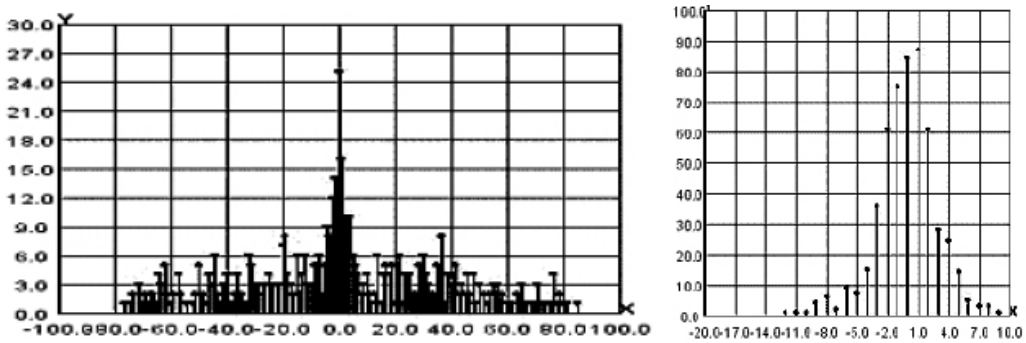


Figure 4.32. Histogramme de la zone de texte et la zone de non texte

(Extraite de [**Gllavata 2004**])

Selon le travail réalisé par J. Gllavata et R. Ewerth, nous avons choisi un filtre d'ondelette de LeGall 5/3 pour tester la méthode d'ondelette. Le filtrage LeGall 5/3 se compose d'un filtre passe-haut HiD et d'un filtre passe-bas LoD :

$$HiD = [0 \quad -1/2 \quad 1 \quad -1/2 \quad 0]$$
$$LoD = [-1/8 \quad 2/8 \quad 6/8 \quad 2/8 \quad -1/8] \qquad (44)$$

Pour tester cette méthode mentionnées ci-dessus, nous avons choisi deux images du type que nous devons traiter : une image géographique (carte de l'Asie, figure 4.33) et une image astronomique (explication des phases de la lune, figure 4.34).



Figure 4.33. Carte de l'Asie

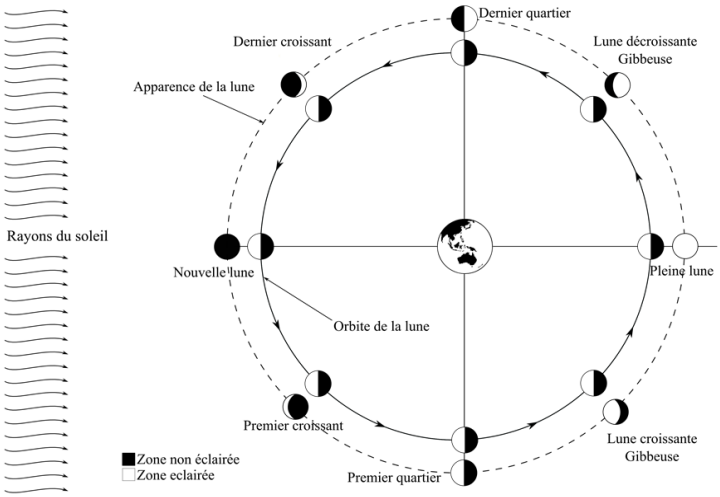


Figure 4.34. Explication des phases de la lune

Dans les figures suivantes, on montre les résultats de la localisation du texte.

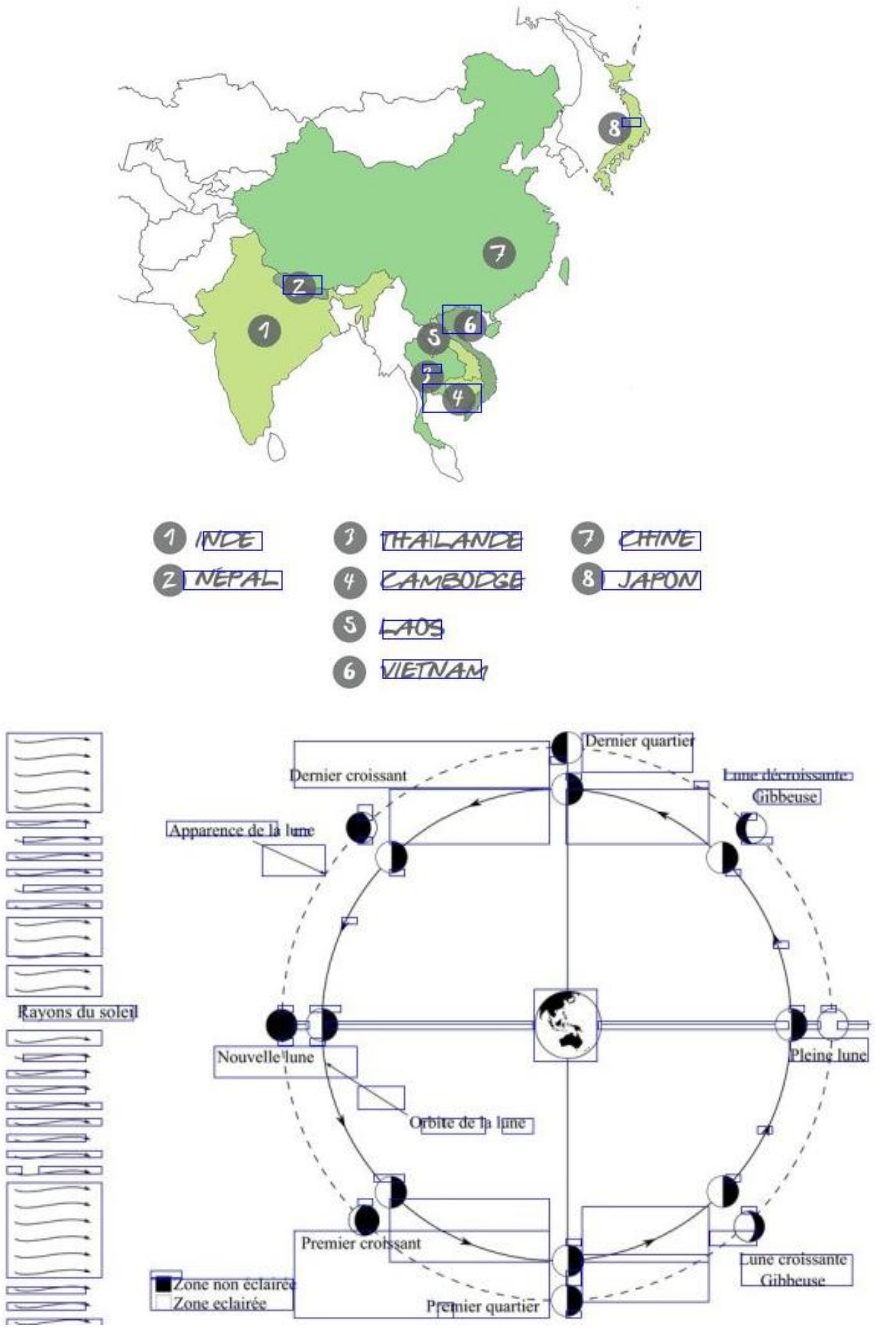


Figure 4.35. Les résultats de la localisation du texte par l'ondelette

Pour la carte d'Asie, les régions du texte sont bien détectées par l'ondelette, mais pour l'image astronomiques, nous trouvons beaucoup d'erreurs. C'est parce que, dans la carte d'Asie, les textes sont loin des autres éléments. Mais dans l'image des phases de lune,

les textes se trouvent près des autres traits et courbes. Les régions qui contiennent des traits ont presque la même distribution de coefficients d'ondelette que les régions du texte. Pour les séparer, il faut donc définir un seuil assez précis, ce qui n'est pas facile à faire d'une manière automatique.

4.4.2.2 Méthode basée sur RLSA

J. Lopez Krahe et P. Chauvet [**Chauvet 1993**] ont proposé dans leur travail d'utiliser un système intelligent du traitement de documents. Ce système est basé sur la méthode de RLSA (Run Length Smoothing Algorithm en anglais). Le RLSA [**Wahl 1981**] [**Wahl 1982**] est une méthode de segmentation capable de différencier les blocs de texte et de non-texte [**Rege 2012**]. Cette méthode se compose de deux étapes. Tout d'abord, une procédure de segmentation divise le document en différentes régions. Chacune d'elles doit contenir un seul type de données (texte, graphique, etc.). Ensuite, certaines caractéristiques de base de ces régions sont calculées pour classer ces régions.

Le RLSA est appliqué à une image en binaire dans laquelle les pixels blancs sont représentés par 0 et les pixels noirs par 1. L'algorithme transforme une séquence d'entrée x en une séquence de sortie y selon les règles suivantes :

- les valeurs 0 sont changées en 1 si le nombre successif de 0 est inférieur ou égal à une limite prédéfinie C.
- les valeurs 1 de x sont inchangées.

Par exemple, avec C= 4, x est l'entrée, avec le RLSA, on obtient la sortie y :

x : 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0

y : 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1

Le RLSA est appliqué ligne par ligne, ainsi que colonne par colonne sur une image. Les différentes valeurs de C sont utilisées pour les lignes et les colonnes. Ensuite, les résultats sont combinés pour obtenir une segmentation finale de ce document. À la fin,

pour chaque bloc les différentes caractéristiques calculées (la hauteur, l'excentricité du rectangle entourant, le rapport entre le nombre de pixels noirs et le nombre de pixels totaux du rectangle entourant) permettent de classer ces régions en différentes catégories (figure 4.36).



Figure 4.36. Détection du texte par le RLSA

(Extraite de [**Rege 2012**])

Les mêmes images sont utilisées pour tester cet algorithme, nous avons obtenu les résultats qui sont présentés dans les figures suivantes :

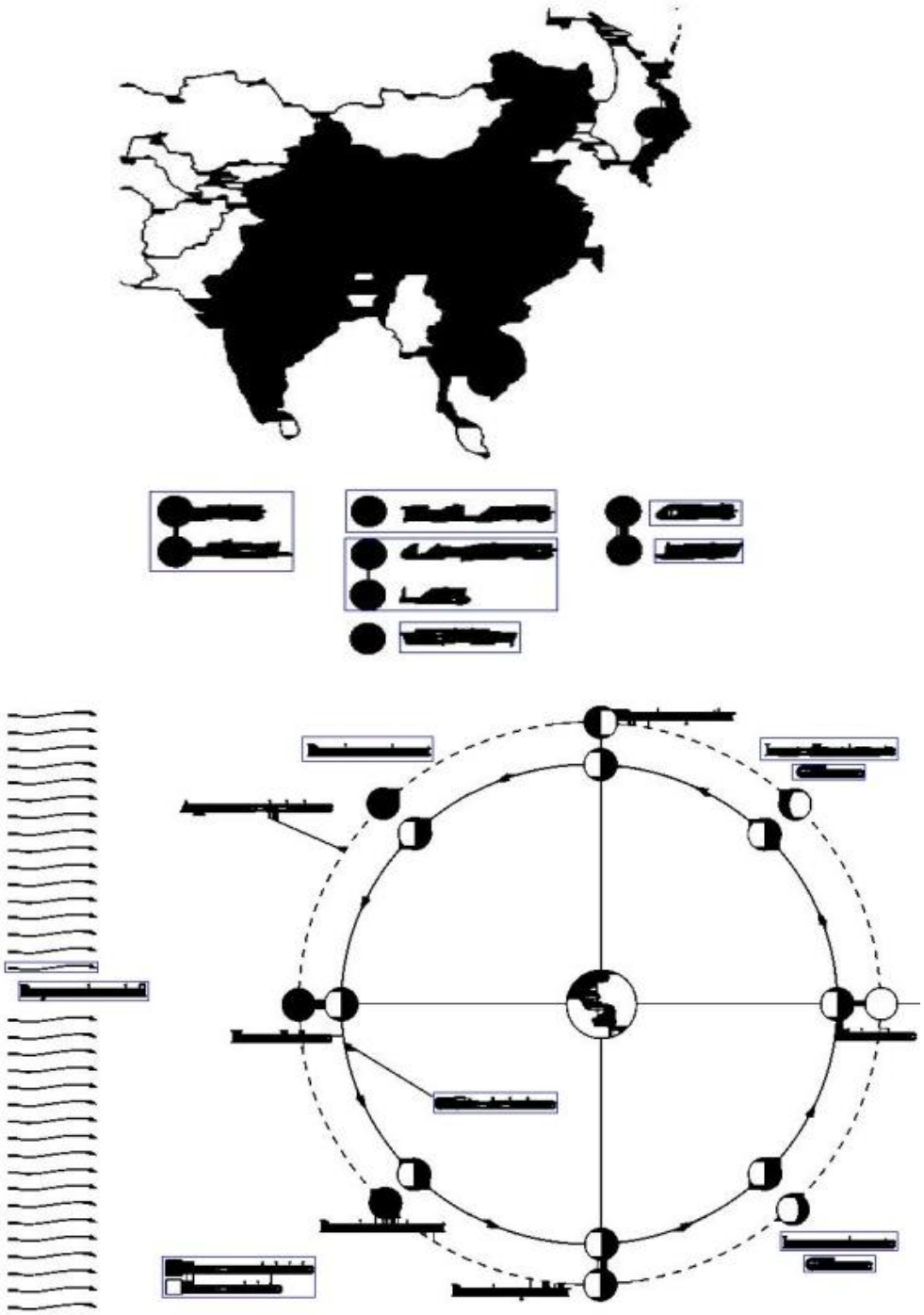


Figure 4.37. Les résultats de la localisation du texte par RLSA

La méthode RLSA donne un bon résultat de la localisation du texte uniquement avec une bonne définition des seuils, mais un bon seuil est variable selon les différents documents d'entrée et selon les tailles du texte à détecter. Si les lettres sont grandes, les seuils aussi doivent être élevés ; si les lettres sont petites, les seuils aussi doivent être bas ; si les seuils sont très grands, les zones du texte risquent d'être collées avec les autres éléments (par exemple, pour le résultat sur l'image de phase de lune); si les seuils sont très petits, les zones de texte ne seront pas bien formées. Le fait que ces seuils ne peuvent pas être définis de façon automatique pour les différentes images d'entrée est un inconvénient majeur de cette méthode.

4.4.2.3 Méthode basée sur SWT

L'algorithme SWT (Stroke Width Transform) est une opération qui transforme les valeurs de couleur contenues par chaque pixel en valeurs de largeur du trait le plus probable, caractéristique assez sûre qui permet de séparer le texte des autres éléments de l'image.

Cet algorithme est proposé par C. Jung, Q. Liu et J. Kim [**Jung 2008**] [**Jung 2009**]. Dans leur travail, les auteurs ont montré que sur les images vidéo, les résultats de la localisation du texte sont très performants en utilisant une nouvelle caractéristique - l'épaisseur du trait des lettres. Basé sur leur recherche, B. Epshtein et E.Ofek [**Epshtein 2010**] ont réalisé un autre travail sur la détection du texte dans les images naturelles.

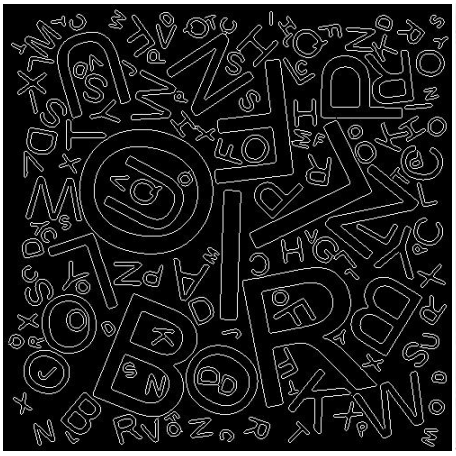
Avec SWT, une image avec la valeur d'intensité de chaque pixel est transformée en image qui ne contient que les valeurs de l'épaisseur du trait des lettres (figure 4.38). Pour réaliser cette transformation et ensuite localiser les zones du texte, basé sur le travail d'Epshtein, nous proposons les étapes suivantes :

- Détecter les contours de l'image (figure 4.38b).
- Calculer la direction de gradient de chaque pixel du contour.
- Calculer la largeur du trait selon la direction de gradient.

- Remplir les pixels dans le trait par la valeur de largeur (figure 4.38c).
- Regrouper les pixels par une analyse de composantes connexes [**Bouman 2012**].
- Filtrer les éléments non-lettres selon les règles suivantes :
  - rejeter les éléments dont l'excentricité est inférieure à 0.1 ou supérieure à 10 ;
  - rejeter les éléments qui ont une très grande surface (plus que 300 pixels) ou très petite (moins que 10 pixels) ;
  - rejeter les éléments qui ont un diamètre très grand par rapport à la valeur moyenne de largeur du trait (si le diamètre est plus que 10 fois la valeur moyenne).
- Grouper les lettres en mots selon les règles prédéfinies (figure 4.39) :
  - les lettres dans un mot doivent avoir la même valeur de largeur du trait ;
  - les lettres dans un mot doivent être sur le même niveau horizontal (en cas de mot horizontal), ou vertical (en cas de mot vertical) ;
  - dans un mot, l'espace entre deux lettres doit être inférieur à un seuil.
- Construire les zones du texte.



(a) Image entrée





(b) Image de contours (c) Image segmentée par SWT

Figure 4.38. Segmentation d'image avec SWT : les lettres qui ont la même valeur de largeur du trait sont représentées par la même intensité de gris.

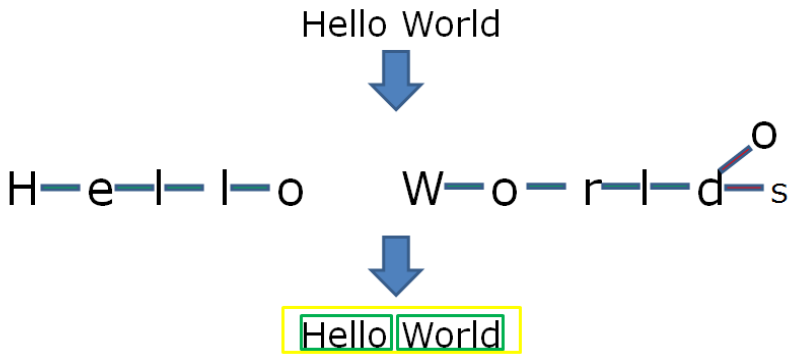


Figure 4.39. La construction de la zone du texte.

(la lettre "o" n'est pas sur le même niveau que la lettre "d", la lettre "s" n'a pas la même largeur du trait que la lettre "d", ces deux lettres ne peuvent pas être considérées comme les lettres du mot "World")

Avec le test effectué, nous avons les résultats de la location du texte par SWT, ils sont présentés dans les figures suivantes :

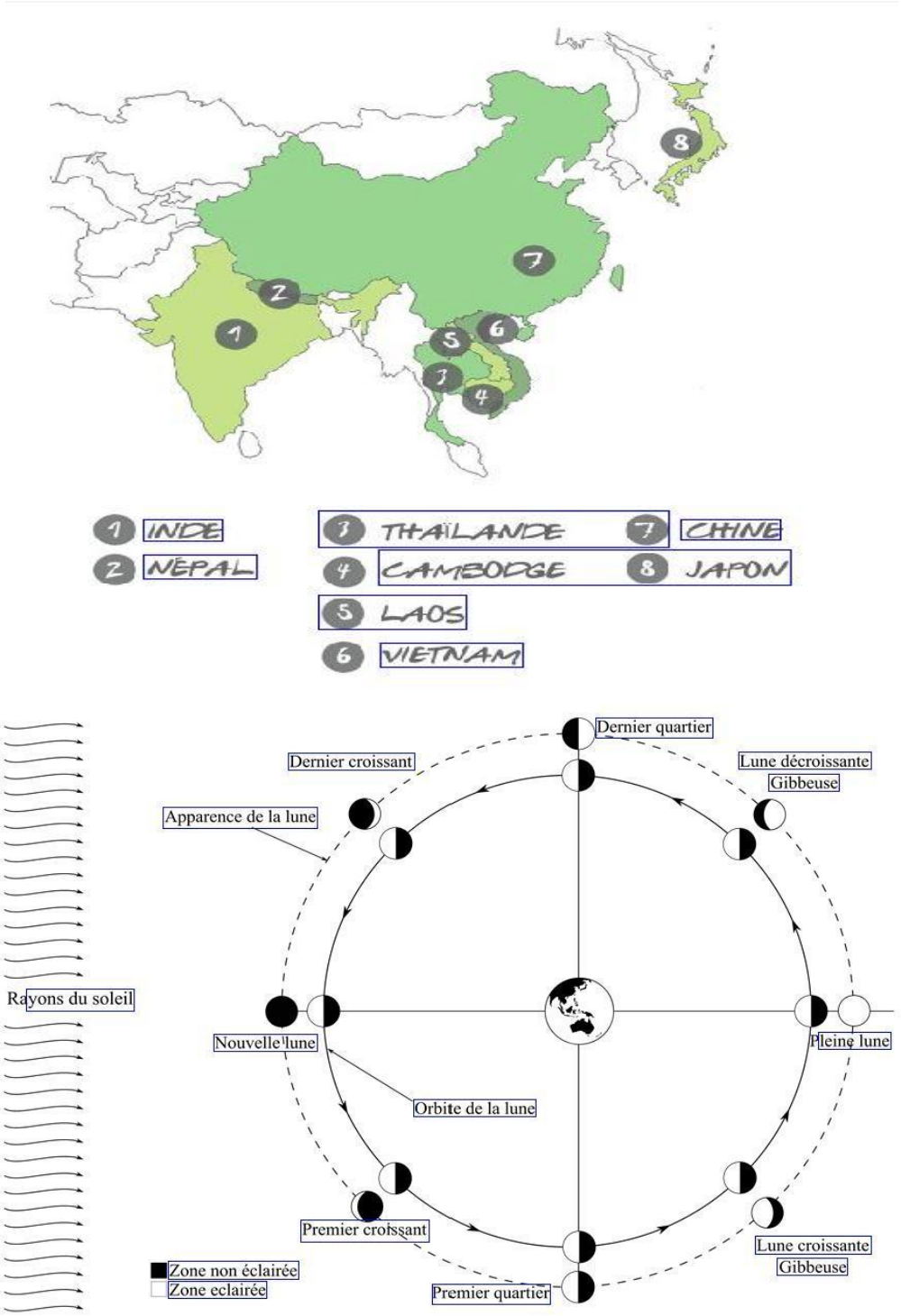


Figure 4.40. Les résultats de la localisation du texte par SWT

La méthode SWT donne quant à elle de très bons résultats sur les deux types d'images testées. De plus, cette méthode ne nécessite pas de définir un seuil. En effet, avec les règles utilisées, ce système peut facilement récupérer les régions qui sont susceptibles de contenir du texte, quelle que soit l'ampleur, la direction, la police et la langue du texte.

4.4.3 Conclusion

Avec les tests que nous avons effectués, nous observons que les méthodes de l'ondelette et de RLSA ont des inconvénients majeurs. Il est difficile de déterminer les bons paramètres de ces deux méthodes de manière automatique. Mais la méthode basée sur SWT n'a pas ce type de problème. Avec les règles prédéfinies, les régions du texte peuvent être détectées automatiquement et correctement. Nous avons donc décidé d'utiliser le SWT comme méthode de localisation du texte dans notre travail.

4.5 Conclusion

Jusqu'à présent, nous avons présenté toutes les méthodes utilisées dans notre travail, y compris les traitements de base et les traitements haut niveaux. Afin de réaliser une procédure plus automatique et plus simple de transcription d'une image en une carte tactile, les choix et les implémentations des algorithmes de traitement d'images numériques ont constitué la partie la plus importante de notre travail. Pour chaque étape de la procédure proposée, nous avons implémenté et testé au moins trois méthodes différentes. En comparant les différents résultats, nous avons choisi les méthodes les plus adaptées à notre problématique. Pour certains problèmes plus complexes, nous nous sommes basés sur des solutions existantes, et nous avons proposé des améliorations de ces méthodes qui ont donné de bons résultats.

Pour transcrire une image en image tactile, il faut suivre toutes les étapes présentées dans ce chapitre. La suite de nos travaux consiste donc à concevoir une IHM (Interface Homme-Machine) qui propose d'effectuer chacune de ces étapes, automatiquement.

CHAPITRE 5

Développement de l'interface et analyses expérimentales

Sommaire :

5.1 Introduction 134
5.2 Interface réalisée 134
5.2.1 Étape du prétraitement d'image 135
5.2.2 Étape de validation 137
5.2.3 Étape de segmentation 137
5.2.4 Étape de reconnaissance de formes 140
5.2.5 Étape de la transcription du texte en Braille 140
5.3 Image en relief imprimée 149
5.4 Test avec les utilisateurs finaux et résultats expérimentaux 150
5.4.1 Protocole 150
5.4.2 Procédure expérimentale 155
5.4.3 Résultats et perspectives 169

5.1 Introduction

Le but de cette partie du travail est, tout d'abord, de construire une interface qui peut réaliser toutes les tâches présentées dans le chapitre précédent. Ensuite, un test expérimental, avec des personnes non-voyantes, est mis en place pour évaluer les images en relief créées par cette interface.

5.2 Interface réalisée

Avec MatLab, nous avons testé tous les algorithmes proposés. Nous avons aussi créé une interface qui peut être utilisée par l'opérateur. Cette interface se compose de zones d'affichage, de boutons de fonctionnalité, de listes de paramètres, etc. Avec ces composantes, cette interface peut réaliser certaines tâches de manière automatique ou au moins semi-automatique.

La figure 5.1 montre une conception générale de l'interface. Concrètement, la procédure générale de mise en œuvre comprend plusieurs étapes, et l'allure de cette interface varie selon les différentes étapes. Nous présenterons dans ce qui suit, les interfaces réalisées et les résultats obtenus à chaque étape.

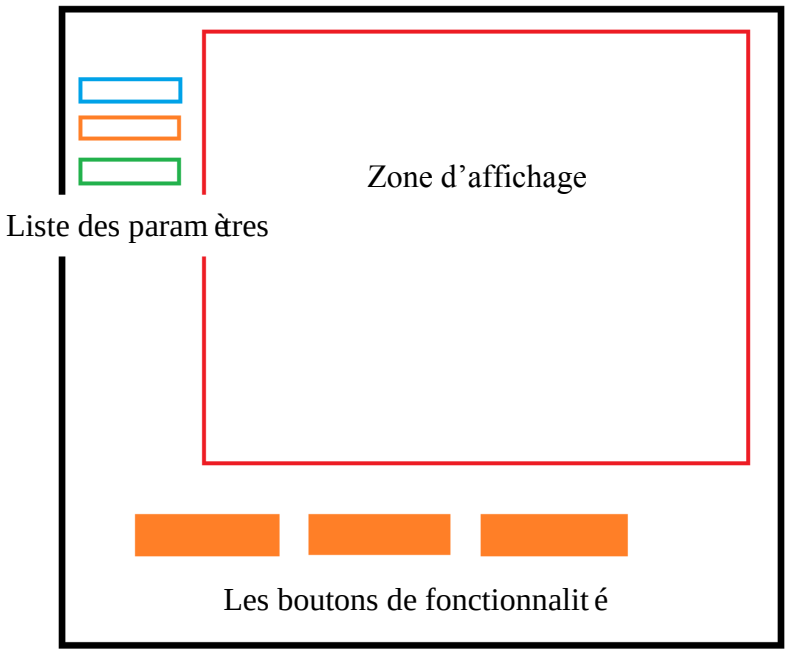
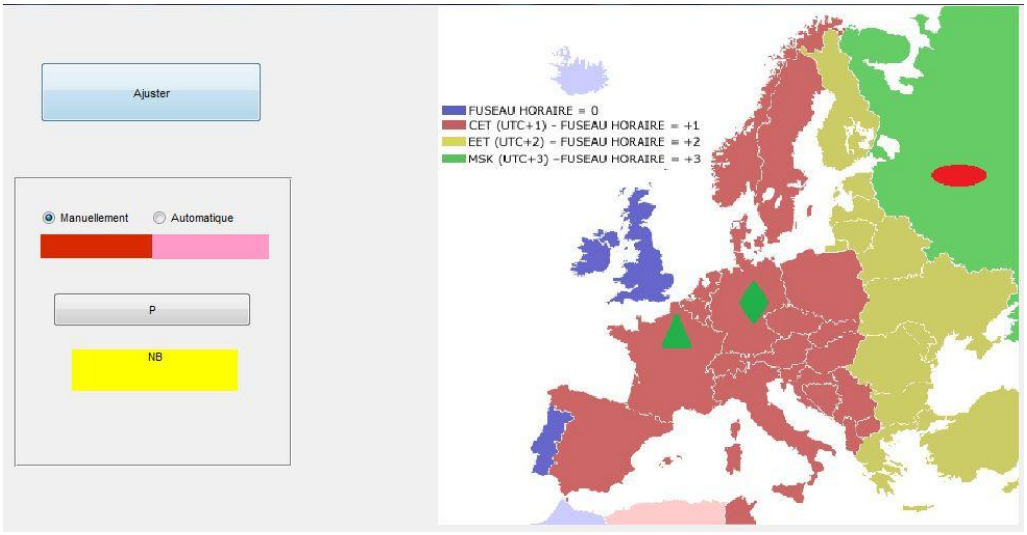


Figure 5.1. Conception d'une interface simple

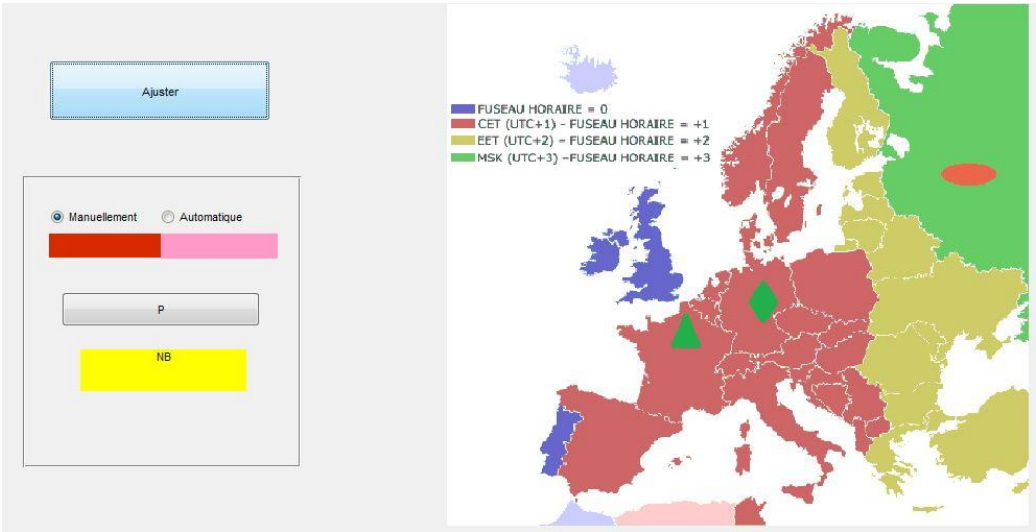
5.2.1 Étape du prétraitement d'image

Avant de commencer les tâches principales, nous avons appliqué une étape du prétraitement qui a comme but de faciliter les tâches suivantes en transformant les images couleurs en binaire ou en supprimant les éléments et les points inutiles des images.

Pour les images en couleur, le prétraitement comprend d'abord une étape d'ajustement pour enlever les points isolés. Cette étape est basée sur une analyse de l'histogramme de l'image. Elle permet ainsi de réduire le nombre de couleurs de l'image et de faciliter ainsi la segmentation (figure 5.2).



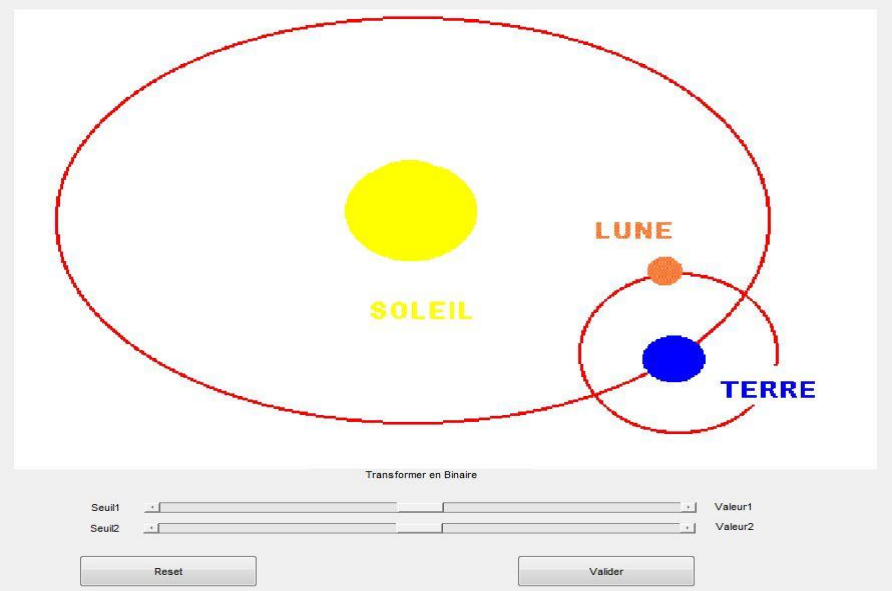
(a) avant ajustement



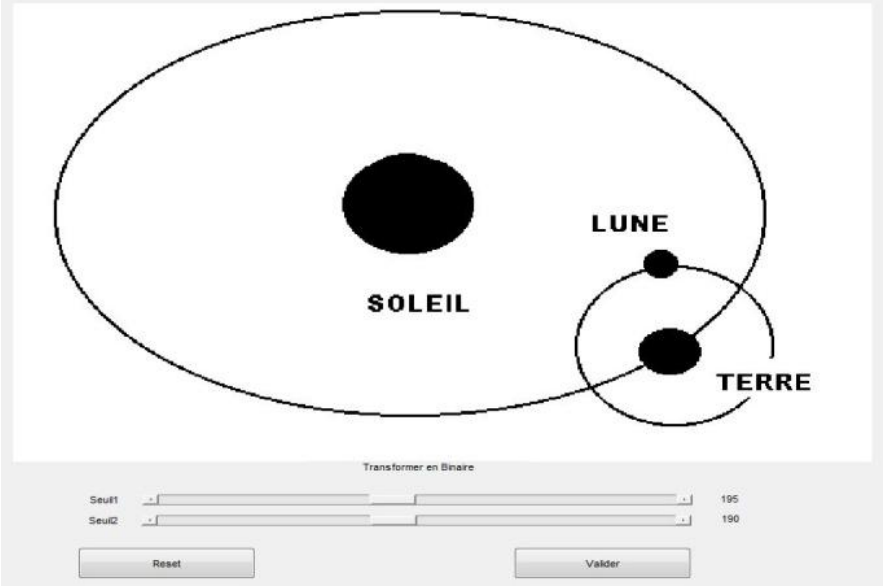
(b) après ajustement

Figure 5.2. Ajustement des images

La deuxième étape du prétraitement consiste à appliquer une étape de binarisation pour transcrire les images couleurs en images binaires. Pour réaliser la binarisation, nous avons utilisé une méthode de seuillage local afin de transformer les images en noir et blanc en conservant le maximum de zones de texte (figure 5.3).



(a) image en couleur



(b) image en binaire

Figure 5.3. Binarisation de l'image

Il est nécessaire d'appliquer toutes ces tâches de prétraitement pour améliorer l'image d'entrée au traitement, pour que les résultats du traitement par les étapes suivantes soient les meilleurs possibles.

5.2.2 Étape de validation

Sachant qu'une méthode automatique peut avoir un résultat incorrect (par exemple, la technique ROC est déjà très développée aujourd'hui, mais même pour les ROC les plus performantes, le taux de reconnaissance n'est pas toujours 100%), une phase de validation de certaines étapes est nécessairement ajoutée pour l'instant. Dans cette phase, l'utilisateur (créateur d'image en relief) peut contrôler les résultats obtenus et faire les modifications dans le cas où il y aurait des fautes.

Nous proposons dans le carde de travaux futurs une phase d'apprentissage pour rendre ce système plus intelligent. Avec cet apprentissage, les paramètres peuvent être définis plus automatiquement et les fautes peuvent aussi diminuer.

5.2.3 Étape de segmentation

Cette étape segmente l'image en différentes classes selon ses couleurs. Ici, le nombre de classes *N* peut être estimé automatiquement par une analyse de l'histogramme. La valeur de *N* peut être aussi donnée par l'utilisateur, pour avoir une segmentation plus correcte qu'avec le nombre estimé automatiquement. Une fois que *N* est défini, l'algorithme FCM est appliqué pour réaliser la segmentation. Cette procédure est répétée 3 fois, avec 3 valeurs *: N-1, N, N+1*, et l'utilisateur peut choisir visuellement la meilleure segmentation en analysant les 3 résultats (figure 5.4).

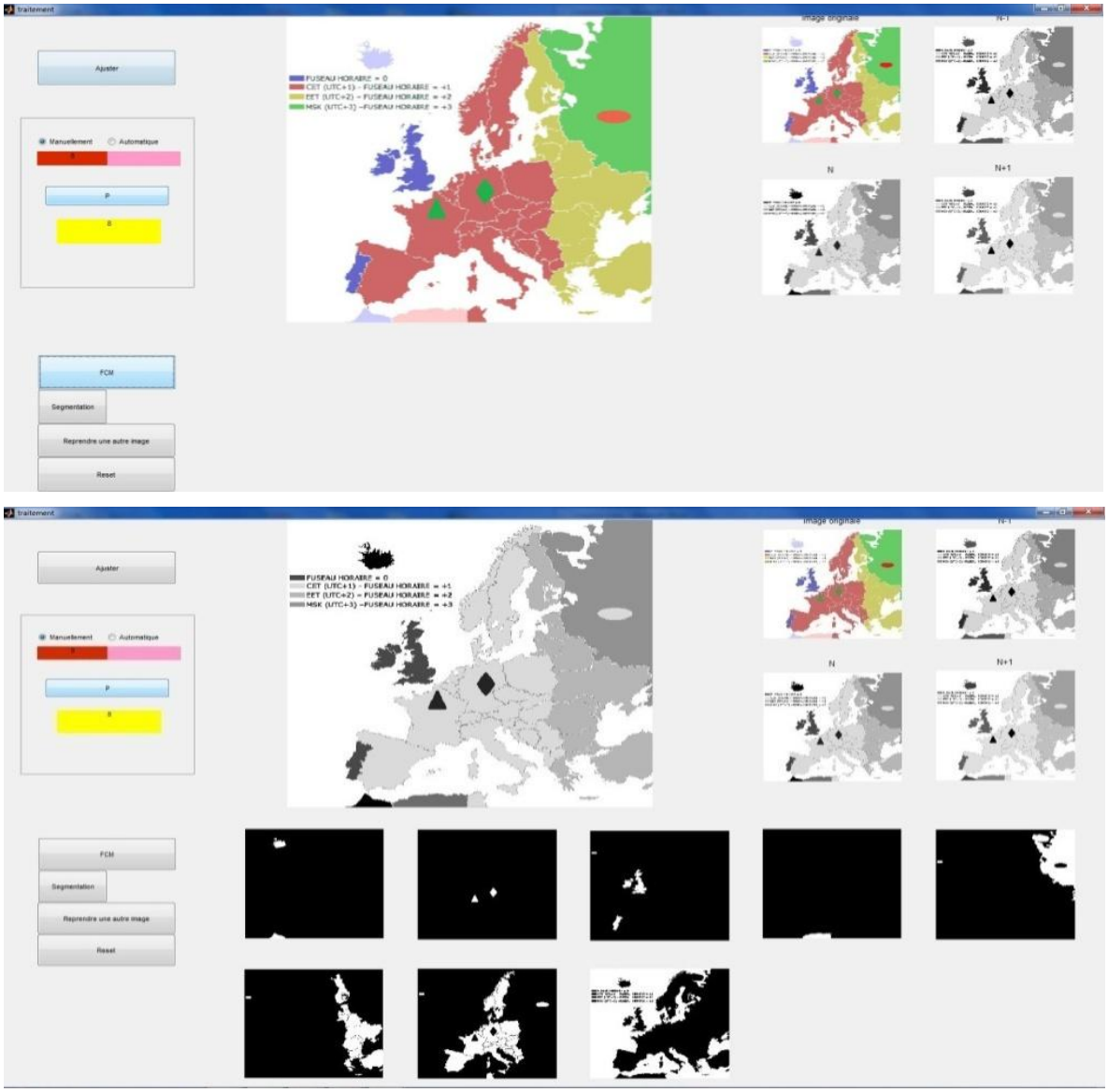


Figure 5.4. Segmentation de l'image

Une fois l'image bien segmentée, la procédure passe à l'étape de post-traitement de chaque classe. Dans cette étape, nous détectons les contours de chaque classe, puis les simplifions et les renforçons. À la fin, les classes vont être remplies par une texture adaptée. Une liste des différentes textures est donnée par le système pour que les utilisateurs puissent faire le choix. Pour chaque texture, la largeur du bas et la valeur de la graisse sont modifiables (figure 5.5). Nous donnons des explications correspondant aux textures appliquées dans la carte de légende (figure 5.6).

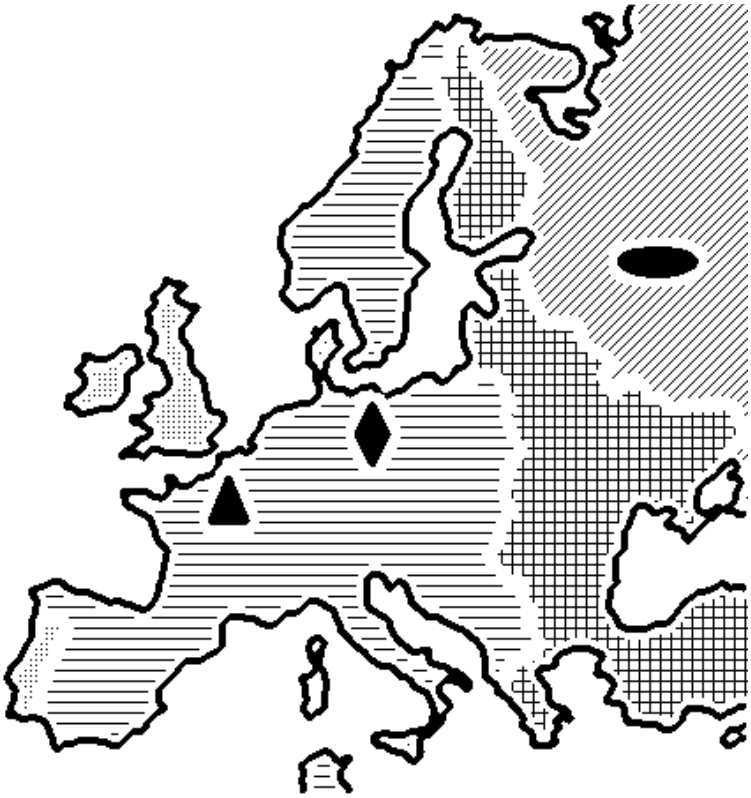


Figure 5.5. Image avec les textures adaptées

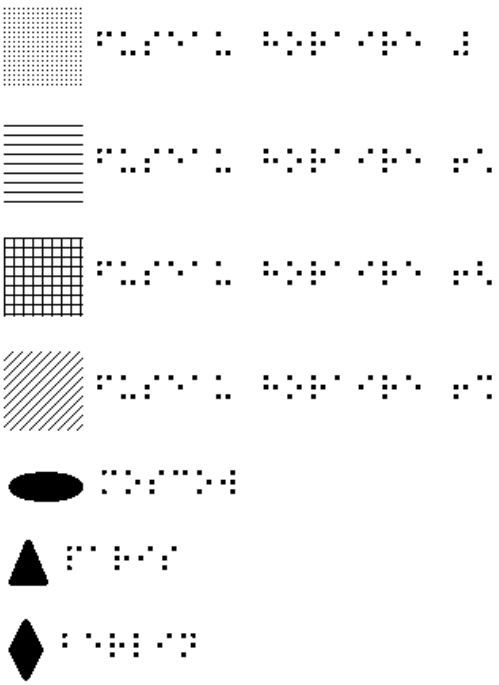


Figure 5.6. Carte de légende.

5.2.4 Étape de reconnaissance de formes

S'il y a des formes dans les classes, nous appliquons la méthode de la logique floue pour les détecter et les identifier (figure 5.7). Une fois les formes identifiées, un symbole de chaque forme est donné automatiquement dans la légende, suivi par une explication précise (figure 5.6).

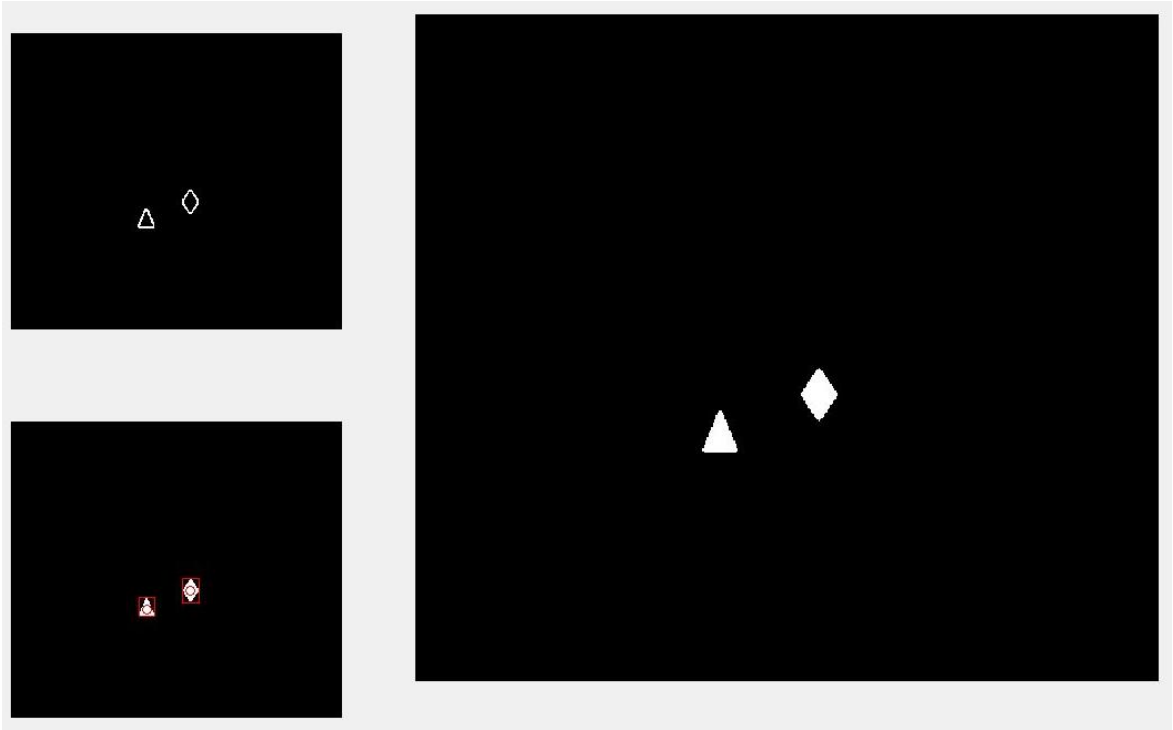


Figure 5.7. Les formes dans une classe

S'il y a des formes de flèche sur l'image, on ajoute en plus une phrase de détection et de modification pour rendre ces flèches plus accessibles aux non-voyants (voir chapitre 4.4).

5.2.5 Étape de la transcription du texte en Braille

Pour transcrire les textes de l'image en Braille, on applique la technique ROC. Sur les images déjà prétraitées, une procédure de localisation des zones du texte est d'abord mise en place dans un premier temps (figure 5.8). Cette procédure est nécessaire pour améliorer les résultats de la ROC qui correspondraient à l'étape suivante. En effet, les

images, contrairement aux documents du texte simple, ont toujours des éléments complexes (les traits, les textures, les contours, les formes, etc.) qui influencent les résultats de la ROC si on l'applique directement sur l'ensemble de l'image. Une procédure de localisation des zones du texte peut diminuer largement ces influences, ce qui peut augmenter la qualité et l'efficacité de la ROC (figure 5.9).

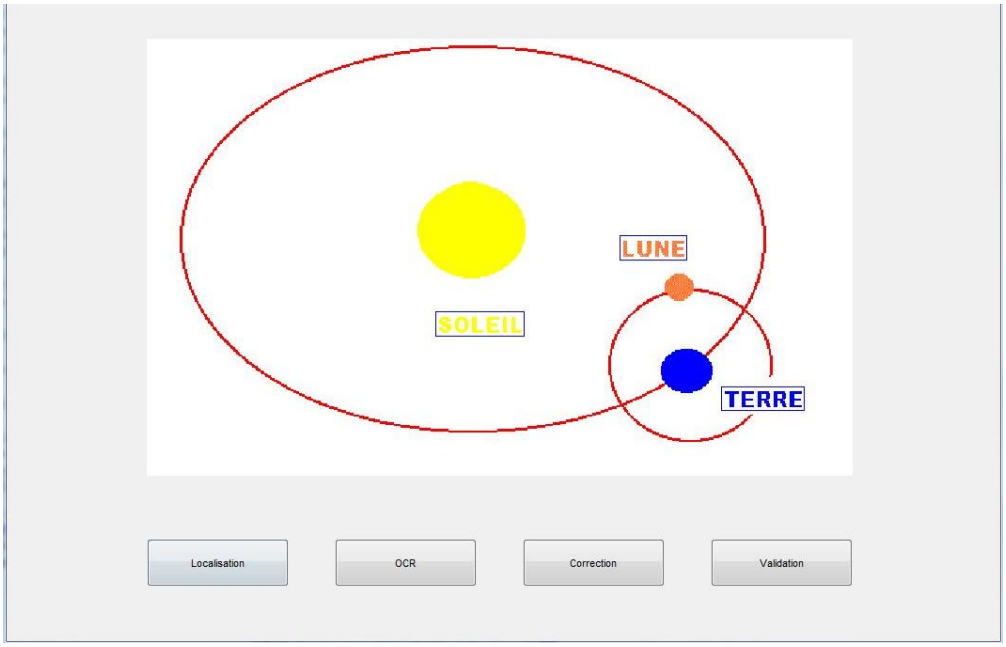


Figure 5.8. Localisation des zones du texte

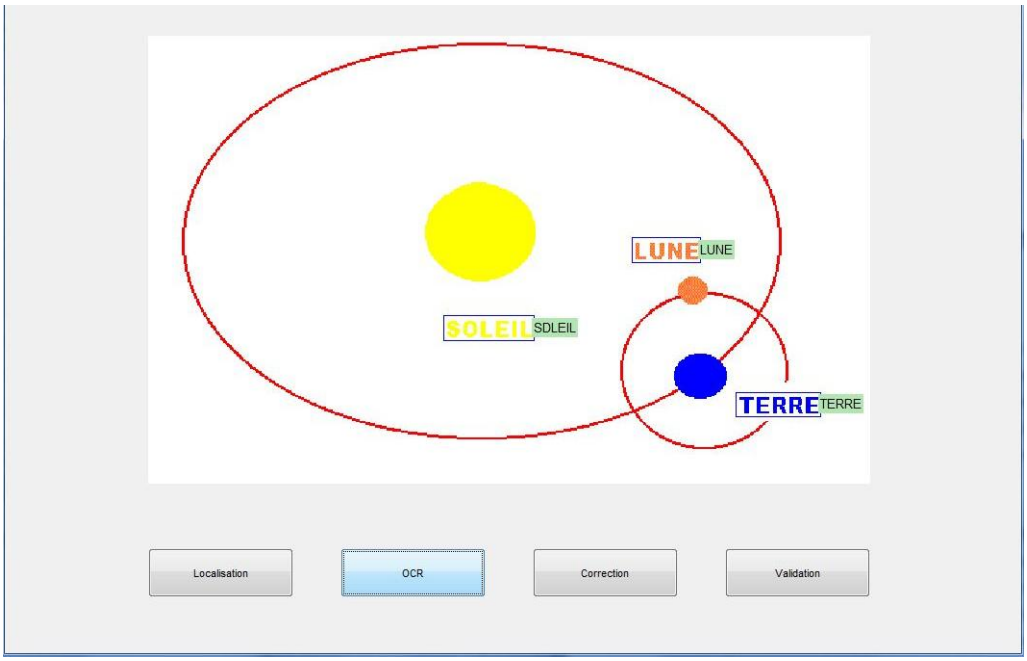


Figure 5.9. Résultats de la ROC

Comme il a été vu dans les chapitres précédents, une étape de correction manuelle est souvent introduite après la ROC pour corriger les résultats de la reconnaissance du texte (figure 5.10). Dans cette étape, l'interface propose plusieurs fonctions pour réaliser les corrections. Pour les textes trouvés et reconnus, s'il y a des erreurs, l'utilisateur peut cliquer avec le bouton gauche de la souris sur les textes pour procéder à une correction d'orthographe (figure 5.11), ou corriger directement les textes, en cliquant avec le bouton droit de la souris (figure 5.12). Il est possible aussi de supprimer, fusionner, ajouter des zones de texte selon les besoins de l'utilisateur (figure 5.13).

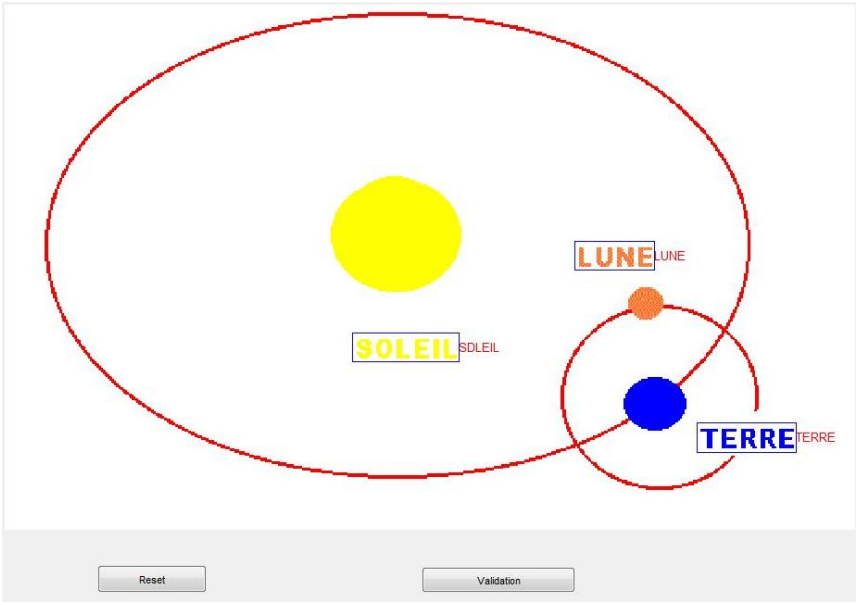


Figure 5.10. Étape de correction

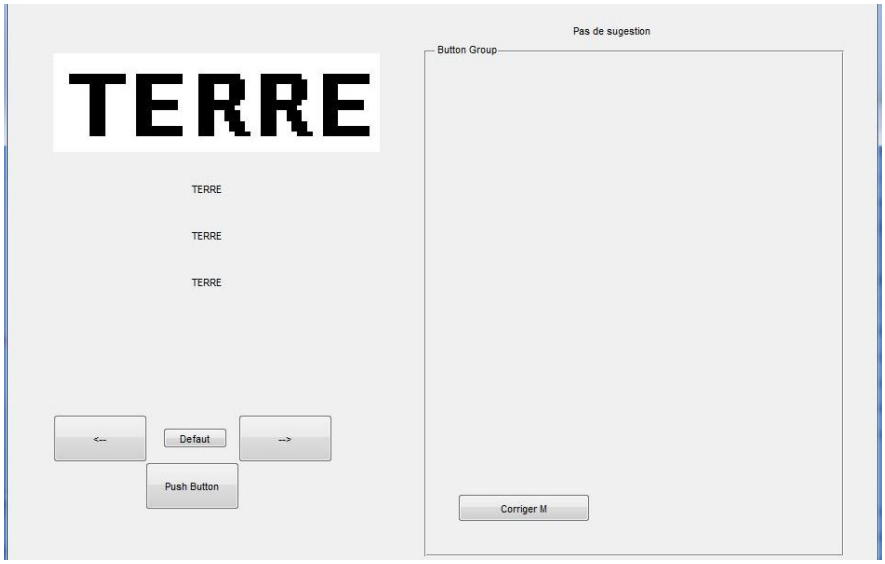


Figure 5.11. Correction d'orthographe

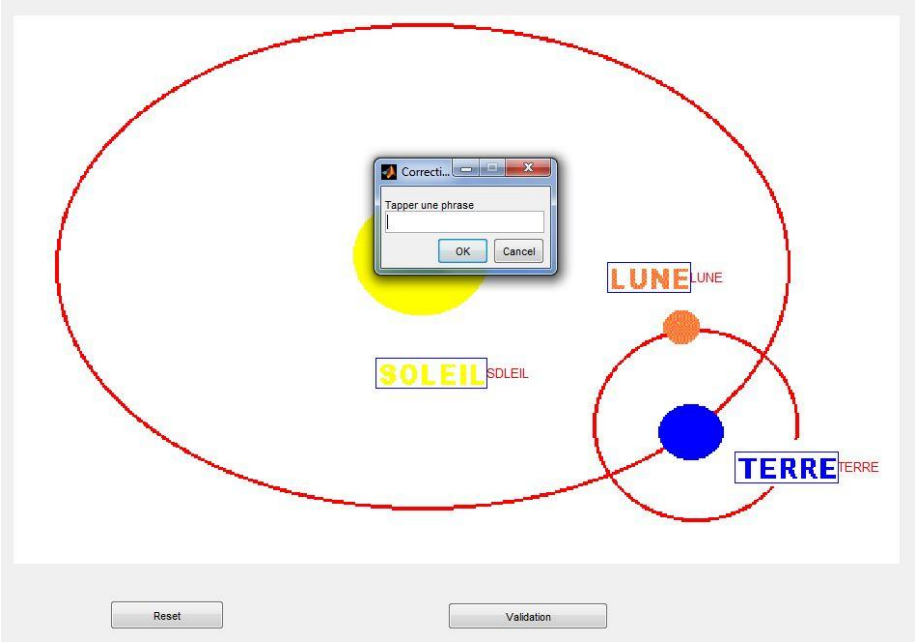


Figure 5.12. Correction directe manuelle

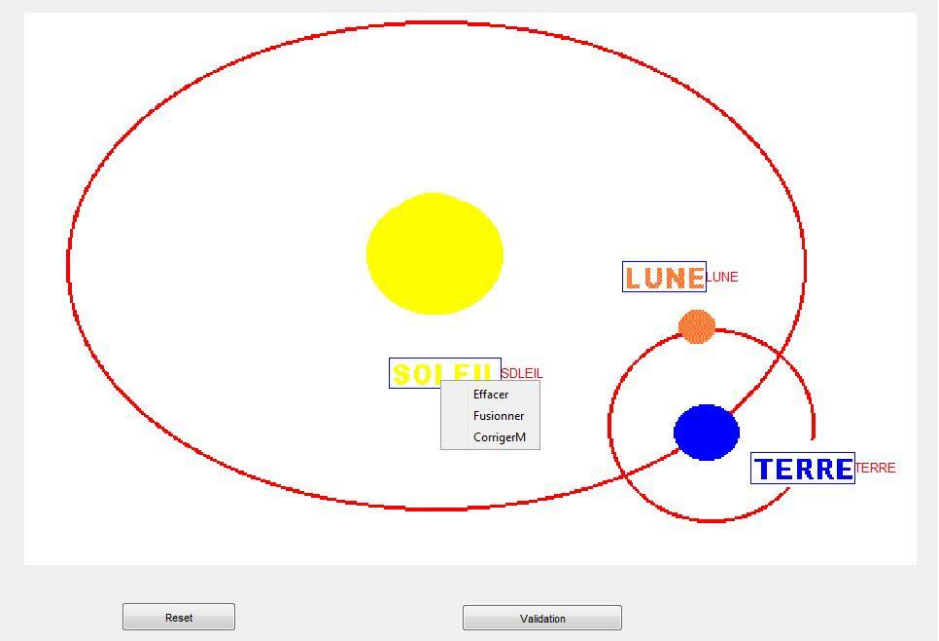
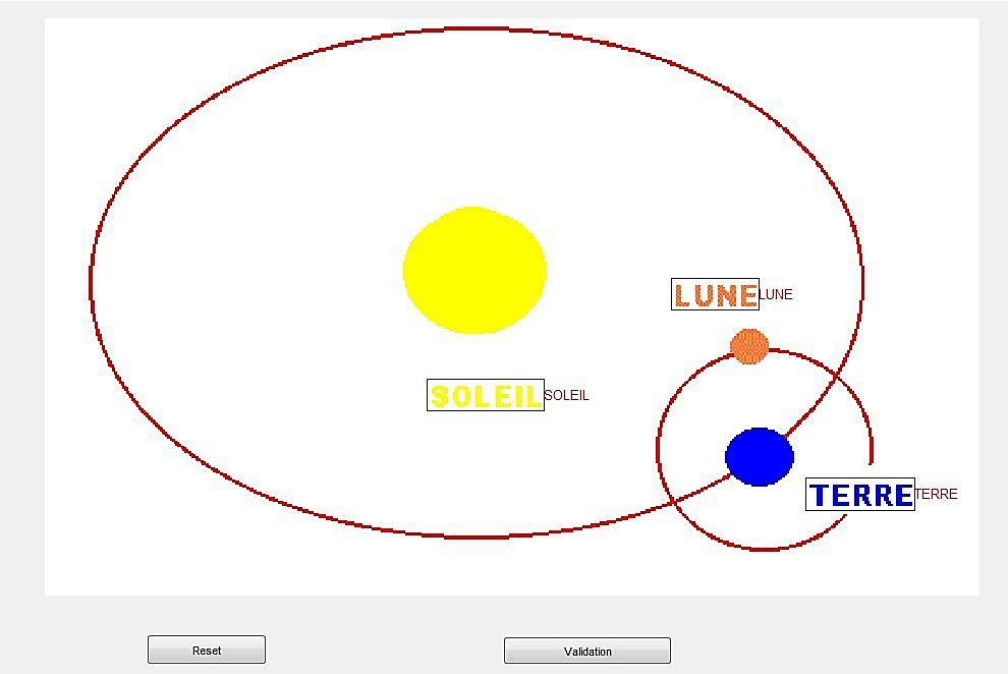


Figure 5.13. Différentes fonctions proposées sur le clic droit de la souris

Une fois que les textes reconnus sont tous corrects, une étape de transcription automatique en Braille est appliquée. Pour cela, on introduit sur l'image, une étiquette en Braille qui remplace chaque zone de texte et on lui affilie automatiquement une explication en Braille dans la légende de l'image (figure 5.14).



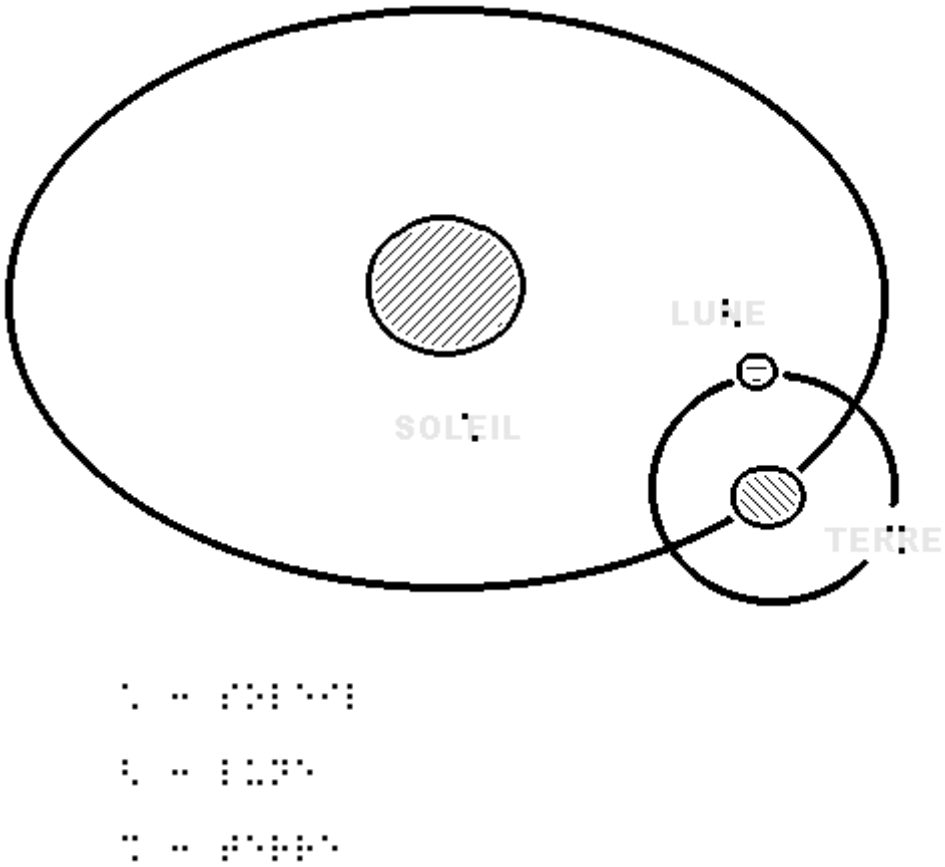


Figure 5.14. Image avec les textes corrigés et sa version en relief avec la légende en Braille

Le logiciel de ROC que nous utilisons est Tesseract. Ses caractéristiques sont présentées ci-dessous :

> "Tesseract est un logiciel de reconnaissance optique de caractères (ROC) qui a été conçu par les ingénieurs de HP de 1985 à 1995. En 2005, son code a été rendu public, et Google a repris son développement. La dernière version actuelle est v3.02 qui a été libérée en 2012, sous licence Apache. Pour l'instant, Tesseract est l'un des moteurs de ROC gratuits les plus performants. Il peut détecter maintenant 40 langues."

Tesseract est un logiciel open-source très performant, avec lequel notre système a réussi à reconnaître les textes dans les images. Pour augmenter le taux de correction des résultats et l'automatisation du système, nous pourrions utiliser un autre ROC plus performant mais payant (au moins 3000 euros pour le SDK de OminiPage, par exemple).

Il existe plusieurs moteurs de ROC sur le marché, ces moteurs ne sont pas en open-source mais sont plus performants que Tesseract. Dans le tableau suivant, nous présentons quelques moteurs ROC connus avec leurs caractéristiques :

Tableau 5.1. Comparaison des moteurs ROC

(-: Pas d'informations précises, W : Windows, M : Mac OS, L : Linux)

Nom	***L'année de création***	***La dernière version***	***Type de Licence***	***Systèmes supportés***	***SDK***	***Langages supportées***
Tesseract	1985	3.02	Apache	W/M/L	oui	35
OminiPage	1970	19	Propriété	W/M/L	oui	125
Microsoft office Document imagning	-	2007	Propriété	W	-	-
OCRFeeder	-	0.7.11	GPL	L	-	-
FreeOCR	-	4.2	Propriété	W	-	-
ABBYY FineReader	1989	11	Propriété	W/M/L	oui	198
Dynamesoft OCR SDK	2003	8.2	Propriété	W	oui	40

Dans le tableau, nous voyons bien que les moteurs OminiPage et ABBYY FineReader sont les ROC les plus performants : ils supportent respectivement 125 langages et 198 langages. Pour chacun d'entre eux, leur version SDK peut être facilement adaptée à notre système. Dans le travail futur, nous pourrons utiliser ces deux ROC pour améliorer les résultats de reconnaissance de texte.

Nous utilisons les étiquettes en Braille (les chiffres en Braille) pour remplacer les textes originaux dans l'image quand le système fait la transcription du textes en Braille. Les étiquettes sont utilisées pour référencer les éléments dans l'image et faire le lien entre ces éléments et la légende. La position des étiquettes doit être proche des éléments étiquetés pour que les non-voyants puissent facilement les détecter. Notre méthode fixe automatiquement la position des étiquettes au milieu des textes originaux (figure 5.15).

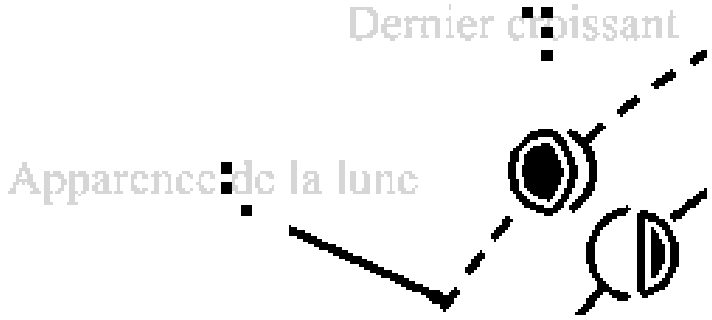


Figure 5.15. Positions correctes des étiquettes

Dans certains cas, la position des étiquettes est bien placée. Elle est proche des éléments (les deux étiquettes dans la figure 5.15). Mais comme la position et la longueur du texte dans l'image sont très variables, la place des étiquettes peut aussi être incorrecte (figure 5.16).

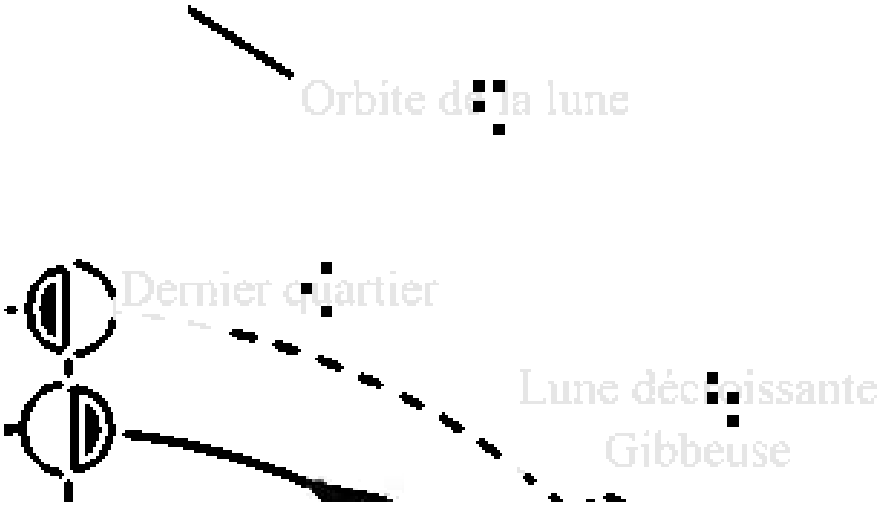


Figure 5.16. Positions incorrectes des étiquettes

Selon les différentes positions relatives entre les textes et les éléments ainsi que la longueur des textes, nous proposons dans notre travail à venir d'établir des règles adaptatives pour trouver automatiquement la bonne position de chaque étiquette.

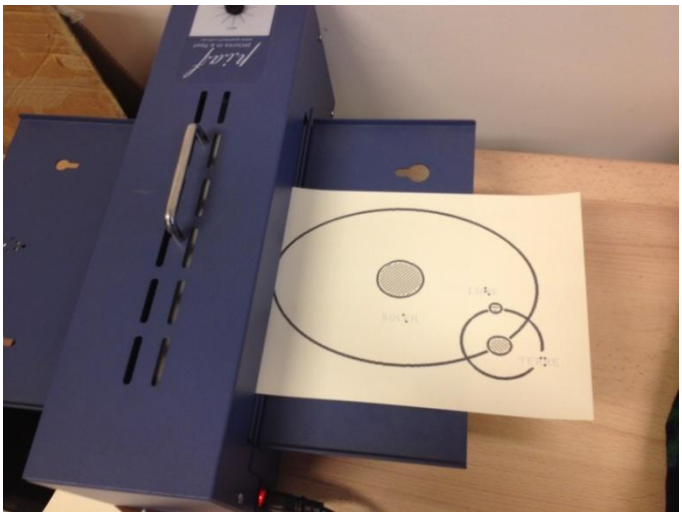
5.3 Image en relief imprimée

Après avoir appliqué toutes les étapes vues précédemment, une version numérique de l'image en relief et sa carte légende sont créées. Pour réaliser sa version physique, il reste la dernière étape - l'impression.

On imprime tout abord les images sur un papier multicouche (qui est recouvert d'une couche de mouse composée de microcapsules d'air qui se gonflent sous la chaleur) avec une imprimante normale, ensuite on passe ce papier dans le four pour le chauffer, ce qui va produire au final une carte en relief (figure 5.17).



(a) Image imprimée avec l'imprimante : HP LaserJet Pro-CM1410



(b) Image en relief imprimée avec le four : PIAF- Four à Thermorelief

Figure 5.17. Impression de l'image en relief.

5.4 Test avec les utilisateurs finaux et résultats expérimentaux

Avec la nouvelle méthode proposée, nous avons créé des images en relief en quatre catégories : image d'astronomie, image géographique, image d'histogramme, image de camembert (les cartes en relief sont données dans l'Annexe).

Pour savoir, si les images fabriquées par notre méthode sont bien opérationnelles pour les utilisateurs (les non-voyants), nous avons organisé un test expérimental sur ces images avec 8 personnes non-voyantes.

Dans ce chapitre, nous présentons ce test expérimental avec les résultats obtenus et des analyses de ces résultats.

5.4.1 Protocole

Avant de commencer le test, nous avons préparé un protocole. Dans celui-ci, nous avons défini d'abord la méthode du test, ensuite, le matériel nécessaire avec une grille d'analyse, ainsi que les tâches et les consignes.

5.4.1.1 La méthode

Nous avons préparé 8 images divisées en 4 catégories. Nous invitons 8 personnes non-voyantes adultes à participer à ce test. Pour chaque participant, nous lui donnons 8 images en relief faites selon notre méthode. Les 8 images sont toutes différentes. Pour bien organiser ce test, nous préparons un tableau de distribution (tableau 5.2).

Tableau 5.2. Tableau de la distribution

(P1-P8 : 8 participants, Pl1-Pl8 : 8 images en relief, a-h : numéro des images)

	Pl1	Pl2	Pl3	Pl4	Pl5	Pl6	Pl7	Pl8
P1	a	b	c	d	e	f	g	h
P2	b	c	d	e	f	g	h	a
P3	c	d	e	f	g	h	a	b
P4	d	e	f	g	h	a	b	c
P5	e	f	g	h	a	b	c	d
P6	f	g	h	a	b	c	d	e
P7	g	h	a	b	c	d	e	f
P8	h	a	b	c	d	e	f	g

5.4.1.2 Le matériel

Le matériel comprend le matériel à évaluer et le matériel d'évaluation.

a) Le matériel à évaluer est composé des images en relief faites par notre méthode ;

b) Le matériel d'évaluation est composé du chronomètre pour mesurer le temps, d'exploration par un participant de chaque image en relief, et d'une questionnaire sur la compréhension de chaque image.

5.4.1.3 La grille d'analyse

Le but de ce test expérimental est d'évaluer les images en relief faites par la méthode proposée via des utilisateurs finaux - les non-voyants. Au point de vue de l'analyse, nous devons noter la durée d'exploration sur chaque image par chaque participant, et aussi le taux de compréhension de cette image.

Les questionnaires sont préparés selon 4 catégories différentes :

a) Image d'astronomie (image a et image e) :

Q1 : Quel type de graphique avez-vous sous les doigts ?

QCM: Est-ce un/une ?

R : a) camembert ?

b) carte ?

c) histogramme ?

d) schéma ?

Q2 : Ce schéma s'agit-il d'un schéma de ?

R : a) physique ?

b) chimie ?

c) astronomie ?

d) problématique de billard ?

Q3 : De quelle planète s'agit-il ?

R : a) La terre ?

b) La lune ?

c) Jupiter ?

d) Mercure ?

Q4 : Comment pouvez-vous expliquer le mécanisme de rotation des planètes dans ce schéma ?

b) Carte géographique (image b et image f) :

Q1 : Quel type de graphique avez-vous sous les doigts ?

QCM : Est-ce un/une ?

R : a) camembert ?

b) carte géographique ?

c) histogramme ?

d) schéma ?

Q2 : C'est une carte de géographie. De quel continent s'agit-il ?

R : a) de l'Afrique ?

b) de l'Asie ?

c) de l'Australie ?

d) de l'Europe ?

Q3 : Les textures remplacent les couleurs. Combien y a-t-il de textures dans l'image ?

R : a) 3

b) 4

c) 5

d) 6

Q4: Que représentent les textures, et quel est le but de cette carte ou comment la comprenez-vous ?

c) Image d'un histogramme (image c et image g)

Q1 : Quel type de graphique avez-vous sous les doigts ?

QCM : Est-ce un/une ?

R : a) camembert ?

b) courbe ?

c) histogramme à barre ?

d) histogramme à rectangle ?

Q2 : Combien y a-t-il de rectangles représentés ?

R : a) 3

b) 4

c) 5

d) 6

Q3 : Que représente chacun des axes ?

Q4 : Quelle information vous donne cet histogramme ?

d) Image de camembert

Q1 : Quel type de graphique avez-vous sous les doigts ?

QCM : Est-ce un/une ?

R : a) planète ?

b) camembert ?

c) cadre d'horloge ?

d) symbole d'une ampoule ?

Q2 : Combien y a t-il de parts dans le camembert?

R : a) 3

b) 4

c) 5

d) 6

Q3 : Combien représente environ la part de ceux qui mettent le pied dessus ?

Q4 : Quelle est la part de chacun ?

5.4.1.4 Les tâches et les consignes

a) Les tâches :

Les tâches consistent à explorer chacun des graphiques qui sont présentés à chaque participant selon la distribution du tableau 5.2.

b) Consigne générale :

"Je vais vous laisser explorer 8 supports en relief. Pour chacun d'entre eux, vous les explorez pour comprendre les contenus. Dès que vous avez compris de quoi il s'agit et bien compris son contenu, vous me le dites sans vous stresser. Le temps d'exploration sera estimé. Un petit questionnaire de 4 questions sur les contenus vous sera demandé à la fin de l'exploration. Ce qui est l'objet d'évaluation, ce n'est pas vous, mais à travers vous les supports qui vous sont présentés".

5.4.2 Procédure expérimentale

Nous invitons 8 personnes non-voyantes braillistes pour tester les images. Avant de commencer chaque test, nous posons à chaque participant des questions concernant certaines informations personnelles qui sont présentées dans le tableau 5.3 :

Tableau 5.3. Informations personnelles

	Age	Age de perdu la vu	Niveau d'étude	Type de milieu de scolarisation	Profession
P1	50	9	Bac+5	normale	Responsable d'une association
P2	28	2	BEP	normale et spécialisé	Vendeuse
P3	34	10	Bac+5	normale et spécialisé	Télé conseille
P4	36	15	Bac+5	normale et spécialisé	Pas de travail
P5	44	17	troisième avant Bac	normale	Pas de travail
P6	35	5	Bac + 5	normale et spécialisé	Informaticien
P7	29	non-voyant né	Bac + 5	normale	Télé conseille
P8	41	7 ou 8	Bac + 5	spécialisé	Responsable de la mission lecture et handicap de la BPI

Ensuite, nous laissons les personnes explorer chaque image. Nous notons le temps d'exploration. Lorsque le participant a fini l'exploration, nous posons les questions et notons ses réponses. Chaque question porte sur différents points. Si la personne donne la bonne réponse à la question 1, nous notons 2 points, sinon, nous proposons un QCM (Questionnaire à Choix Multiples) qui contient 4 choix. Si la personne fait le bon choix à la question 2, nous notons 1 point. La question 3 et la question 4 prennent 2 points chacune. La dernière question prend 4 points. Si toutes les questions ont obtenu une bonne réponse, la personne obtient 10 points.

Les réponses de chaque participant sont données dans les tableaux suivants (tableau 5.4 -5.11) :

Tableau 5.4. Carte a (astronomique)

	Q1	QCM	Q2	Q3	Q4
P1		d)			
P2		d)	c)	a)	Non, les symboles de la terre ne sont pas le même, il faut doubler la taille des flèches.
P3		d)	c)	a)	La terre tourne autour du soleil. Les flèches sont lisibles mais la taille est un peu petite.
P4	astronomie		c)	a)	La terre tourne autour du soleil. La direction n'est pas connue avec les flèches.
P5		d)		a)	La terre tourne autour du soleil. Impossible de connaitre les flèches.
P6		d)	c)	a)	La terre tourne autour du soleil. Les flèches sont lisibles pour connaitre la direction.
P7		d)	c)	a)	Le soleil est au milieu, la terre tourne. La direction donnée par les flèches.
P8		d)	c)	a)	La terre tourne autour du soleil qui est au milieu, les flèches sont un peu petites.

Tableau 5.5. Carte b (géographique)

	Q1	QCM	Q2	Q3	Q4
P1		b)	d)	a)	Les textures représentent les fuseaux horaires.
P2	carte géographique		d)	b)	Les fuseaux horaires des pays de l'Europe.
P3	carte géographique		d)	a)	Les textures représentent les fuseaux horaires.
P4	carte géographique		d)	b)	Les fuseaux horaires des pays de l'Europe. Les textures représentent les différents fuseaux horaires.
P5		b)	d)	b)	Les textures représentent les fuseaux horaires de l'Europe.
P6	carte géographique		d)	b)	Les fuseaux horaires de l'Europe.
P7	carte géographique		d)	b)	Les fuseaux horaires.
P8		b)	d)	b)	Fuseaux horaires de l'Europ, et des villes.

Tableau 5.6. Carte c (histogramme)

	Q1	QCM	Q2	Q3	Q4
P1		d)	c)	Nombre des étudiants pour axe vertical, note d'examen pour axe horizontale.	Combien d'étudiant ont eu telle note.
P2	Graphique pour des statistiques		c)	Nombre d'étudiant, note d'examen.	Nombre d'étudiant selon note d'examen.
P3		d)	c)	Vertical : nombre d'étudiant; horizontale : note d'examen final.	40 étudiants entre 60 et 80, 4 étudiants entre 0 et 10.
P4		d)	c)	Vertical : nombre d'étudiant; horizontal : note d'examen.	
P5		d)	c)	Vertical : nombre d'étudiant; horizontal : note d'examen.	40 étudiants entre 60 et 80.
P6		d)	c)	Vertical : nombre d'étudiant; horizontal : note d'examen.	Nombre d'étudiant selon les notes d'examen final.
P7		d)	c)	Vertical : nombre d'étudiant; horizontal : note d'examen.	Entre 60 et 80, il y a 40 étudiants.
P8		d)	c)	Vertical : nombre d'étudiant; horizontal : note d'examen.	Combien d'élèves ont telle notes.

Tableau 5.7. Carte d (camembert)

	Q1	QCM	Q2	Q3	Q4
P1			b)	Plus petite	Ramasse 1/3, le plus grande, donne un coup de pied le seconde, regarde sans prendre le troisième.
P2		b)	b)	Plus petite, 5%	Le plus grande : essayer de ramasser. Deuxième : donner un coup de pied. Troisième : regarde sans prendre.
P3	camembert		b)	La plus faible.	Le plus grande : essayer de ramasser. Deuxième : donner un coup de pied. Troisième : regarde sans prendre.
P4		b)	b)	5%	5% - met le pied; 20% - regarde sans prendre 30% -donne un coup de pied 45% - essaye de ramasser
P5	camembert		b)	Plus petite - 10%	Le plus grande : essayer de ramasser. Deuxième : donner un coup de pied. Troisième : regarde sans prendre.
P6	camembert		b)	5%	Le plus grande : essayer de ramasser. Deuxième : donner un coup de pied. Troisième : regarde sans prendre.
P7	camembert		b)	10%	La plupart essayent de ramasser
P8	camembert		b)	10%	Le plus grande : essayer de ramasser. Deuxième : donner un coup de pied. Troisième : regarde sans prendre.

Tableau 5.8. Carte e (astronomique)

	Q1	QCM	Q2	Q3	Q4
P1		c)	c)	a) b)	La lune tourne autour de la terre, la terre tourne autour de la soleil sur le grand cercle.
P2	système astronomique		c)	a)b)	La terre tourne autour du soleil, la lune tourne autour de la terre.
P3	spatial, astronomie.		c)	a)b)	
P4	carte spatiale		c)	a)b)	La terre tourne autour du soleil, la lune tourne autour de la terre.
P5	astronomie		c)	a)b)	
P6	schéma		c)	a)b)	La terre tourne autour du soleil, la lune tourne autour de la terre.
P7		c)	c)	a)b)	La terre tourne autour du soleil
P8		c)	c)	a)b)	La terre tourne autour du soleil, la lune tourne autour de la terre

Tableau 5.9. Carte f (géographique)

	Q1	QCM	Q2	Q3	Q4
P1	carte géographique		a)	b)	Identifier Sahara, Maroc et Mauritanie.
P2		b)	a)	b)	Identifier Sahara, Algérie et Mauritanie.
P3	Carte géographique		a)	b)	Identifier Sahara, Maroc, Algérie, Mali et Mauritanie.
P4	Carte géographique		a)	b)	Identifier Sahara, Maroc, Algérie, Mali et Mauritanie.
P5	géographique		a)	b)	Identifier Sahara, Maroc, Algérie, Mali et Mauritanie.
P6	géographique		a)	b)	Identifier Sahara, Maroc, Algérie, Mali et Mauritanie.
P7		b)	a)	b)	Identifier Sahara, Maroc, Algérie, Mali et Mauritanie.
P8	géographique		a)	b)	Identifier Sahara, Maroc, Algérie, Mali et Mauritanie.

Tableau 5.10. Carte g (histogramme)

	Q1	QCM	Q2	Q3	Q4
P1		d)	d)	Vertical : le nombre de personnes, horizontal : l'âge	Le nombre de personnes selon l'âge dans un concert.
P2		d)	d)	Vertical : le nombre de personnes, horizontal : l'âge	Le nombre de personnes dans un concert.
P3		d)	d)	Vertical : le nombre de personnes, horizontal : l'âge	Le plus grand partie est: 24 personnes entre 20 ans et 25 ans, le plus petite est 7 personnes entre 5 ans et 7 ans.
P4	Histogramme en rectangle		d)	Vertical : le nombre de personnes, horizontal : l'âge	Dans un concert, combien de personnes ont telle âge.
P5		d)	d)	Vertical : le nombre de personnes, horizontal : âge	Nombre de personne dans un concert selon l'âge
P6		d)	d)	Vertical : le nombre de personnes, horizontal : l'âge	24 personnes entre 20 ans et 25 ans
P7	Histogramme en rectangle		d)	Vertical : le nombre de personnes, horizontal : l'âge	Proportionnalité entre l'âge et le nombre de personnes.
P8		d)	d)	Vertical : le nombre de personnes, horizontal : l'âge	Dans un concert, combien de personnes ont telle âge.

Tableau 5.11. Carte h (camembert)

	Q1	QCM	Q2	Q3	Q4
P1	camembert		b)	30%	Hiver 30%, l'été 45%, printemps 10%, automne 15%.
P2	camembert		b)	30%	L'été 45%, printemps 10%, automne 15%.
P3	camembert		b)	30%	L'été 45%, printemps 10%, automne 15%.
P4	camembert		b)	30%	L'été 45%, printemps 10%, automne 15%.
P5	camembert		b)	30%	L'été 45%, printemps 10%, automne 15%.
P6	camembert		b)	30%	L'été 45%, hiver 30%, automne 15%, printemps 10%.
P7		b)	b)	30%	L'été 45%, hiver 30%, automne 15%, printemps 10%.
P8	camembert		b)	30%	L'été 45%, printemps 10%, automne 15%.

Voici les résultats obtenus pour chaque image :

Tableau 5.12. Notes et temps d'exploration de chaque image selon les participants.

(N : note, T : temps en second de l'exploration, **M** : valeur moyenne, σ : écart type)

		P1	P2	P3	P4	P5	P6	P7	P8	**M**	σ
a	N	1.0	6.5	9.0	9.0	6.0	9.0	9.0	9.0	7.31	2.84
	T	642	315	600	97	436	133	210	234	333.38	206.44
b	N	8.0	10	8.0	10	7.0	10	10	9.0	9.00	1.20
	T	368	188	398	198	463	188	220	126	268.63	122.53
c	N	9.0	10	9.0	9.0	9.0	9.0	9.0	9.0	9.13	0.35
	T	478	230	360	157	405	178	134	106	256.00	139.51
d	N	8.0	9.0	10	9.0	10	10	10	10	9.50	0.76
	T	215	120	210	70	255	180	222	106	172.25	65.69
e	N	9.0	10	6.0	10	6.0	10	7.0	9.0	8.38	1.77
	T	188	153	180	31	215	60	130	115	134.00	63.71
f	N	9.0	8.0	10	10	10	10	9.0	10	9.50	0.76
	T	180	161	145	60	192	180	195	108	152.63	47.12
g	N	9.0	9.0	9.0	10	9.0	9.0	10	9.0	9.25	0.46
	T	370	120	120	120	260	264	205	150	201.13	91.27
h	N	10	10	10	10	10	10	9.0	10	9.88	0.35
	T	119	120	140	139	165	110	115	78	123.25	25.61

Dans la figure 5.18, nous montrons les notes obtenues par chaque participant, et le note moyenne de chaque image testée. Les écarts type de la Note σ_N et du temps σ_T sont calculés selon les équations suivantes :

$$\sigma_N = \sqrt{\frac{\sum_{i=1}^{n}(x_{Ni} - M_N)^2}{n-1}}\ ,\ \sigma_T = \sqrt{\frac{\sum_{i=1}^{n}(x_{Ti} - M_T)^2}{n-1}} \qquad (45)$$

où n est le nombre de données (n = 8), M_N est la valeur moyenne de la Note, M_T est la valeur moyenne du temps d'exploration,.

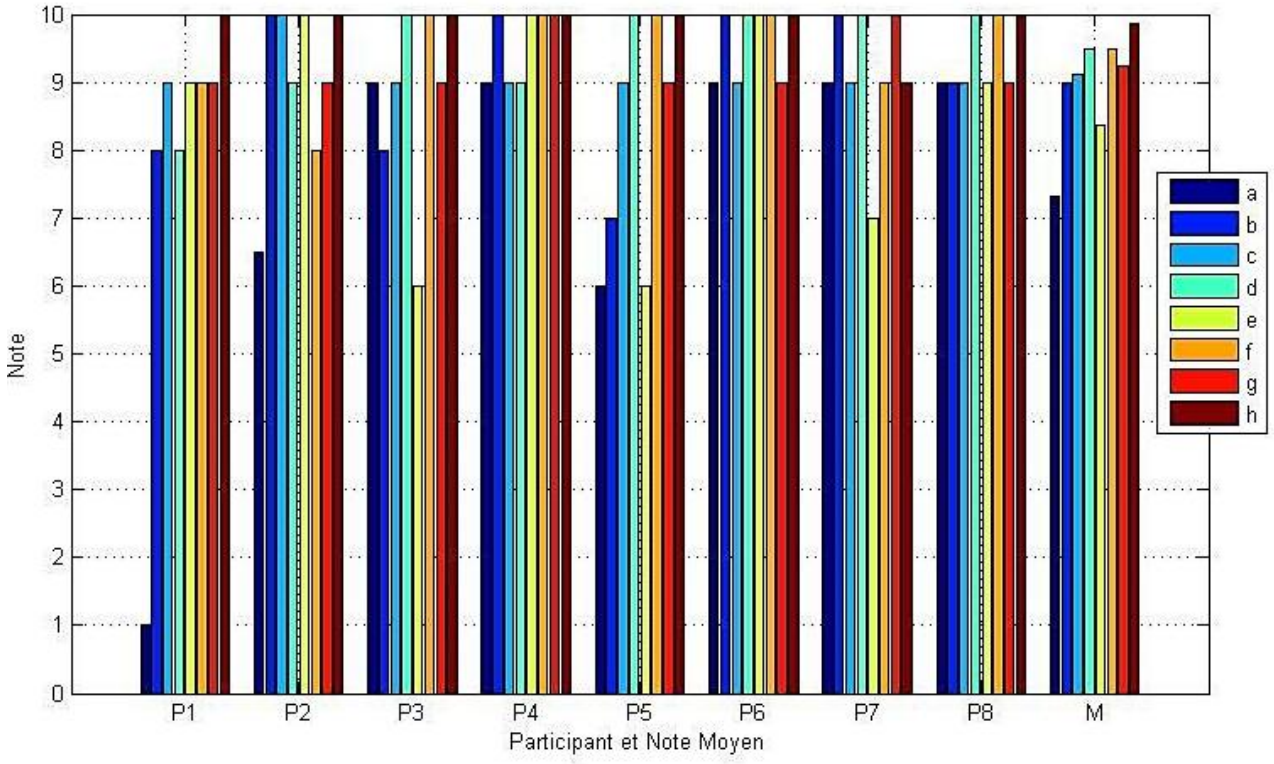


Figure 5.18. Note obtenue par chaque participant et la note moyenne de chaque image

Avec les écarts types calculés, nous pouvons estimer l'intervalle de confiance du temps d'exploration selon l'étude de *la Loi t de Student* :

$$\left[M_T - t\frac{\sigma_T}{\sqrt{n}}, M_T + t\frac{\sigma_T}{\sqrt{n}} \right] \qquad (46)$$

où M_T est la valeur moyenne du temps d'exploration, *t* est le quantile de *la Loi de Student* à *k (7)* degrés de liberté, σ_T étant l'écart type du temps d'exploration.

Nous prenons le seuil de confiance $\alpha = 0.025$ et la valeur de *t* = 2.365, donc pour des intervalles de pari bilatéral à 95%, les résultats obtenus sont donnés dans le tableau suivant (tableau 5.13) :

Tableau 5.13. Intervalle de confiance estimé du temps d'exploration de chaque image testée

(avec $k = n - 1 = 7, \alpha = 0.025, t = 2.365$)

	Temps d'exploration (en seconde)
a	[161, 506]
b	[166, 371]
c	[139, 373]
d	[117, 227]
e	[81, 187]
f	[113, 192]
g	[125, 277]
h	[102, 145]

Pour calculer l'intervalle de confiance de la note, sachant que la borne supérieure est fixée (la note ne peut pas dépasser 10), il est plus intéressant de chercher seulement la borne inférieure de l'intervalle. Nous pouvons calculer l'intervalle unilatéral de la note selon l'équation :

$$\left[M_N - t\frac{\sigma_N}{\sqrt{n}}, 10 \right] \qquad (47)$$

où M_N est la valeur moyenne de la Note, *t* est le quantile de *la Loi de Student* à *k (7)* degrés de liberté, σ_N étant l'écart type de la Note.

Nous prenons le seuil de confiance $\alpha = 0.05$ et la valeur de t = 1.895, donc pour un intervalle unilatéral à 95%, les résultats obtenus sont donnés dans le tableau suivant (tableau 5.14) :

Tableau 5.14. Intervalle de confiance estimé de la note de chaque image testée

(avec $k = n - 1 = 7, \alpha = 0.05, t = 1.895$)

	Note
a	[5.4, 10]
b	[8.2, 10]
c	[8.9, 10]
d	[9.0, 10]
e	[7.2, 10]
f	[9.0, 10]
g	[8.9, 10]
h	[9.6, 10]

Les figures 5.19 et 5.20 montrent la relation entre les valeurs moyennes et les intervalles de confiance estimés :

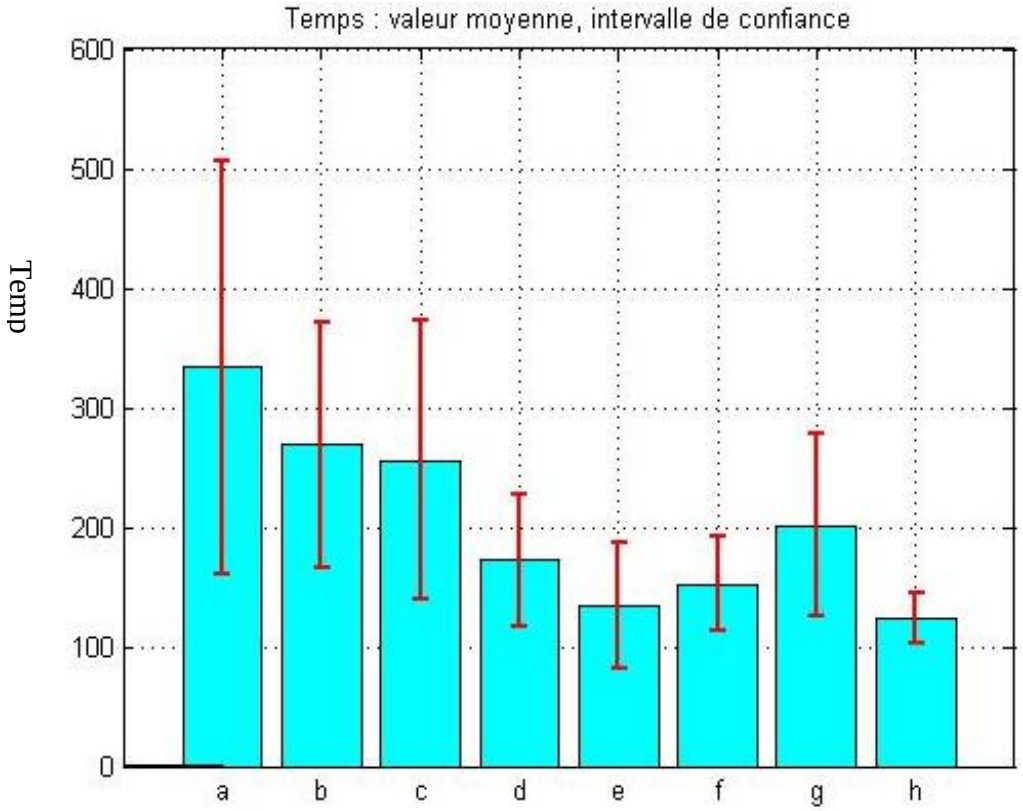


Figure 5.19. Temps d'exploration moyen de chaque carte et intervalle de confiance estimé

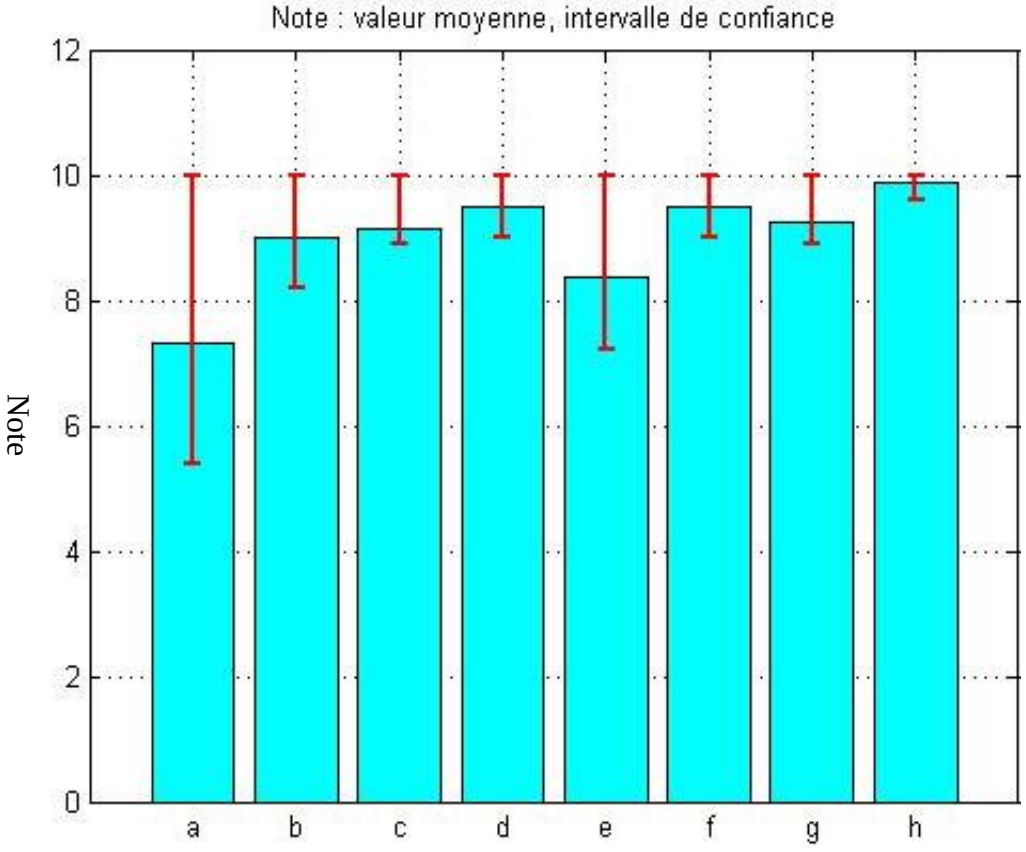


Figure 5.20. Note moyenne de chaque image et intervalle de confiance estimé

5.4.3 Résultats et perspectives

Les résultats montrent que, pour la même image, le temps d'exploration peut être très variable selon les différents participants. Par exemple, pour l'image a, le participant P1 a utilisé 642 secondes, mais le participant P6 a utilisé seulement 133 secondes. Cela peut s'expliquer par le fait que les participants n'ont pas les mêmes expériences antérieures. Les personnes qui ont déjà travaillé avec ce genre d'images en relief prennent moins de temps par comparaison avec les personnes qui n'ont aucune expérience. Certaines personnes qui n'ont jamais touché un graphique d'histogramme doivent utiliser plus de temps pour comprendre ce que représente l'image.

Avec les notes moyennes de chaque image, nous pouvons dire que celles en relief créés par notre méthode, sont compréhensibles pour les participants. Certaines personnes ont eu des difficultés à lire quelques images parce qu'elles ne connaissaient pas les contenus représentées. Par exemple, une personne ne connait pas le système des '4 saisons' de la terre a eu du mal à comprendre les 4 positions différentes de la terre autour du soleil. Dans certains cas, les participants peuvent distinguer les différents éléments représentés dans l'image, mais il faut donner un peu d'explications pour qu'ils puissent comprendre ces éléments (il faut expliquer la relation entre les 4 saisons et les positions relatives de la terre et le soleil, par exemple, ce qui est normal dans une application pédagogique).

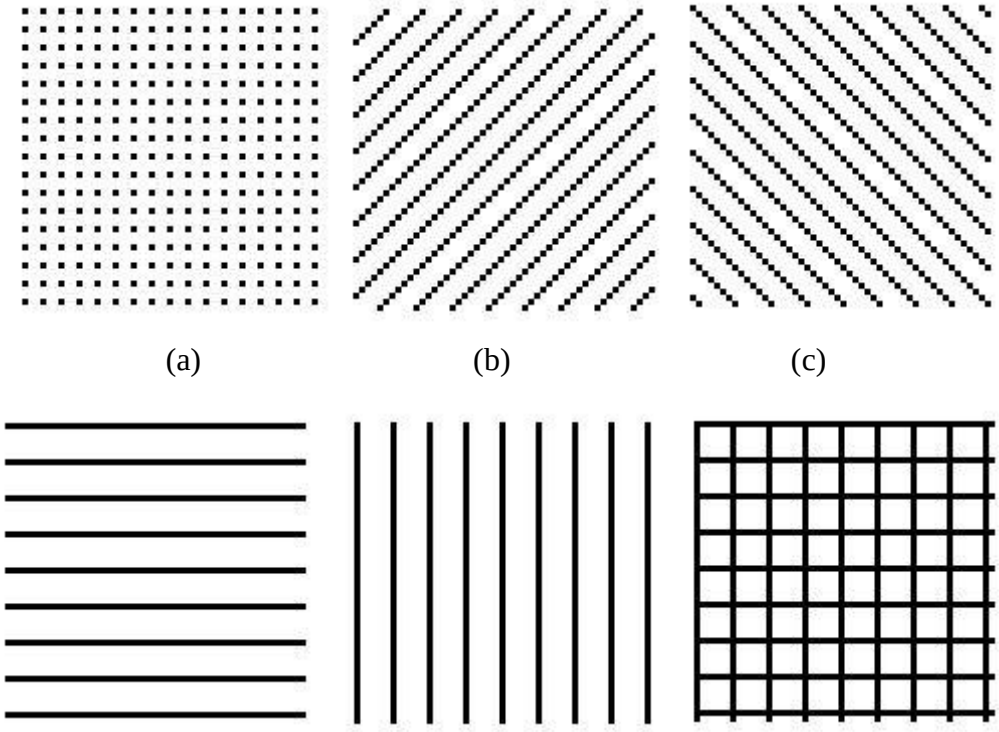
Dans ce test expérimental, nous avons invité seulement un échantillon de huit personnes non-voyantes. Les analyses des résultats ont été effectuées avec l'aide de la *Loi t de Student*, ce qui nous permet une généralisation des résultats [**DeWinter 2013**]. Les tableaux 4.4 et 4.5 ont montré les estimations de l'intervalle général sur le temps d'exploration et sur la note obtenue avec lesquels nous pouvons avoir des perspectives plus globales.

Par exemple, pour l'image a, l'intervalle de confiance estimé du temps d'exploration est entre 161 secondes et 506 secondes ; l'intervalle de note est entre 5.4 et 10. C'est-à-dire que, si nous organisons plusieurs fois des tests avec huit participants différents et aléatoires, pour chaque participant, en général, la probabilité de son temps

d'exploration nécessaire sur la carte a qui se trouve entre 161 secondes et 506 secondes est 95% ; sa note obtenue se trouve entre 5.4 et 10 a la probabilité de 95%.

Parmi toutes les images testées, les camemberts ont eu les intervalles de confiance des notes plus petits que les autres ([9.0, 10] pour la carte d, [9.6, 10] pour la carte h), ensuit ce sont les histogrammes ([8.9, 10] pour la carte c, [8.9, 10] pour la carte g) puis les cartes géographiques ([8.2, 10] pour la carte b, [9.0, 10] pour la carte f), enfin les images astronomiques ([5.4, 10] pour la carte a, [7.2, 10] pour la carte e) ont eu les intervalles de notes plus larges. Nous pouvons considérer que les camemberts, en général, sont plus faciles à comprendre pour les non-voyants, et les images astronomiques (plus complexes et nécessitant plus de connaissances préalables) sont les plus difficiles, ce qui est normal.

À la fin des explorations de chaque participant, nous leur avons demandé de choisir les textures les plus faciles à différencier parmi les 6 textures utilisées dans notre travail (figure 5.21). Enfin, presque tous les participants ont choisi les textures aux motifs verticaux et horizontaux. Ensuite, 2 personnes ont choisi la texture pointillée. Les textures carrelages et obliques sont un peu difficiles à distinguer pour les participants. En revanche, dans ses travaux, Michel Bris [**Bris 2004**] recommande d'éviter d'utiliser les motifs verticaux et horizontaux. En fait, il y a beaucoup de paramètres de motifs qui peuvent influencer la lisibilité d'une texture, par exemple le pas et la largeur du trait. Pour l'instant, nous ne pouvons pas affirmer que les textures préférées sont les motifs verticaux ou horizontaux. Il faudrait faire encore des tests en appliquant tous les différents paramètres pour trouver les textures les plus faciles à distinguer. Par ailleurs le nombre de textures différentes dans l'image ainsi que le voisinage des textures sont aussi des paramètres à prendre en compte.



(a) (b) (c)

(d) (e) (f)

Figure 5.21. Les différents motifs utilisés

(a : motif pointillé, b : motif oblique (45 °), c : motif oblique (-45 °),

d : motif horizontal, e : motif vertical, f : motif carrelage)

CHAPITRE 6

Discussion et conclusion

Sommaire :

6.1 Avantages de la méthode proposée 174
6.2 Du point de vue d'accessibilité du document numérique 176

Les images/cartes en relief (tactiles) sont largement utilisées dans plusieurs domaines comme un outil très important et accessible aux non-voyantes et mal voyantes. La création d'images en relief est un travail assez complexe. Il faut, pour les créateurs, avoir une très bonne connaissance des recommandations générales de la conception d'image en relief, ainsi que d'un ou plusieurs logiciels de retouche d'image. Les créateurs doivent avoir un entraînement pour pouvoir créer une image en relief lisible et compréhensible par les personnes non-voyantes.

Dans notre travail, nous avons présenté une nouvelle méthode de création d'images en relief. Cette méthode utilise les techniques du traitement d'image numérique pour rendre le processus de création d'image en relief plus automatique et plus facile. Les tests que nous avons réalisés, auprès de huit participants, ont montré que les images fabriquées par notre système sont faciles à lire. Les notes que les participants ont obtenues ont montré qu'ils ont bien compris les informations représentées dans les images. Par ailleurs cette méthode est à la portée d'une personne sans connaissances spécifiques préalables avec du matériel aujourd'hui assez banal, ce qui devrait faciliter une autonomie dans le cadre d'une éducation inclusive, par exemple.

6.1 Avantages de la méthode proposée

Pour conclure ce travail de recherche, nous discutons les avantages importants de notre méthode en comparant avec la méthode manuelle.

La raison du lancement de ce projet est de résoudre deux contraintes majeures à la conception manuelle d'une image en relief :

a) La généralisation ;

Il n'y a pas de règles strictes sur la création d'image en relief [**Lobben 2012**], nous pouvons trouver seulement certaines recommandations proposées par les experts. Les créateurs peuvent prendre ces recommandations ou bien utiliser leurs propres règles. Les images créées sont très variables selon les différentes règles appliquées et les différents créateurs.

Avec notre méthode, les images sont créées automatiquement en utilisant les paramètres prédéfinis des techniques de traitement d'image numérique. Notre système propose donc une possibilité de généraliser et formaliser la procédure de la création d'une image en relief. Pour réaliser ce but, les paramètres doivent être définis automatiquement selon les différentes images entrées et respecter au maximum les recommandations proposées. Dans notre travail futur, la tâche de définition automatique des différents paramètres sera mise en œuvre. Une fois ces paramètres fixés, les images créées sont bien généralisées avec les mêmes règles de conception. Pour les personnes non-voyantes, les images en relief devraient être plus faciles à lire si elles ont le même format et respectent les mêmes règles.

b) L'automatisation.

Ce problème est lié à l'efficacité de la création d'image en relief. Nous avons discuté avec certains créateurs, qui nous ont expliqué que pour créer une image en relief, il faut entre quelques heures et plusieurs jours. Pendant les tests expérimentaux, les participants nous ont aussi raconté que pour avoir une image en relief, il faut attendre 2 semaines dès la demande posée au créateur. Le temps de la création est devenu un problème très important.

La méthode proposée peut donner une solution très rapide et facile. Le système prend seulement quelques minutes pour sortir une image en relief. Les tâches des créateurs sont largement simplifiées. Avec la méthode proposée, il est possible pour une personne ordinaire (un enseignant, un membre de la famille d'un non-voyant par exemple) de créer facilement une image en relief.

Après les expérimentations faites avec les utilisateurs, les images en relief créées sont compréhensibles. Le taux de compréhension de l'image peut atteindre 100% si nous donnons un peu d'explications supplémentaires aux participants.

6.2 Du point de vue d'accessibilité du document numérique

Depuis une vingtaine d'années, avec le développement des ordinateurs personnels et des réseaux Internet, l'accessibilité des documents numériques est toujours une demande forte des personnes malvoyantes et non-voyantes. Il est aussi devenu un défi à résoudre. Grace aux avantages des nouvelles techniques, ces personnes sont aujourd'hui capables d'accéder à certains documents numériques. Par exemple, le lecteur d'écran peut lire séquentiellement les textes d'un document et les contenus affichés sur un écran d'ordinateur. Un moteur de ROC peut détecter les textes dans un document et les reconnaitre. Un afficheur Braille est aussi un dispositif pratique pour accéder à des documents contenant du texte. Mais pour les documents avec des contenus plus complexes, l'accessibilité reste très limitée aujourd'hui [**Archambault 2010**], par exemple, pour des documents contenant des images dans une page Web d'Internet (un lecteur d'écran ne peut pas lire une image, mais le titre et la légende peuvent être lus s'ils sont écrits en texte), ou pour les informations graphiques contenues dans une partition musicale [**Encelle 2008**] (dans ce cas, un lecteur d'écran doit avoir une coopération avec des outils spéciaux, par exemple, le logiciel BME [**Jessel 2002**]).

De très nombreux d'applications et logiciels sur le marché peuvent traduire automatiquement les textes visuels en informations auditives (avec une synthèse vocale) ou tactiles (avec un afficheur Braille). Mais jusqu'à présent, nous trouvons très peu de travail concernant la transcription automatique d'une image en image accessible aux non-voyants. Cette procédure est un défi [**Brock 2013**]. Wang [**Wang 2009**] et Kostopoulos [**Kostopoulos 2007**] proposent une possibilité de la création automatique d'une image tactile.

La navigation sur les pages Web est une partie essentielle dans notre vie moderne. L'Internet est une ressource très riche avec lequel nous pouvons trouver une grande quantité de données. Mais les images sur le Web ne sont pas toujours lisibles pour les non-voyants. En effet, les outils informatiques (par exemple, une image sur le Web peut être accessible aux non-voyants avec une souris à retour d'effort et une synthèse vocale [**Jessel 2004**]) et certaines validateurs d'accessibilité de page Web (par exemple, le site W3C

concernant l'accessibilité) peuvent augmenter le taux de lisibilité des pages pour un non-voyant. Les images sont traditionnellement expliquées aux non-voyants par les étiquettes en texte qui sont lus par un lecteur d'écran. Avec cette façon, le taux d'acquisition des informations totale d'une image est très limité.

Notre travail a permis de réaliser une procédure de création automatique d'image en relief, appliquée à des images pédagogiques avec du contenu sémantique simplifié (cartes géographiques, données statistiques, graphiques, mécanique céleste, etc.). Dans notre travail, les types d'images traitées sont plus larges, les images sont aussi plus complexes par rapport aux travaux de Wang [**Wang 2009**] et Kostopoulos [**Kostopoulos 2007**] (limités à des plans de ville). En revanche elles sont plus simples que celles utilisées par Karpathy et Li [**Karpathy 2014a, b**], qui travaillent sur des images réelles, mais les objectifs sont différents.

Notre procédure peut être utilisée comme un outil pratique pour rendre accessible les images numériques (les images scannées ou téléchargées de l'Internet). Bien sûr pour atteindre ce but, il y a encore beaucoup de travail à faire sur l'amélioration des fonctionnalités, l'automatisation de la définition des paramètres et éventuellement la possibilité de la coopération avec les autres modalités de présentation, par exemple avec une carte interactive classique (système image en relief-tablette avec une synthèse vocale) ou un écran tactile avec Feedback- électro-vibration.

Dans le cadre de futurs travaux, le système pourrait encore être amélioré en suivant les recommandations suivantes :

1. Nous proposons de chercher une méthode automatique pour fixer les valeurs de quelques paramètres qui sont fixés manuellement. Cette méthode peut utiliser des algorithmes d'apprentissage, ce qui rendra le système plus intelligent et autonome.

2. Notre travail a été réalisé avec un ROC open-source - Tesseract. Nous proposons de tester d'autres logiciels ROC pour déterminer et utiliser le plus performant.

3. Le logiciel actuel est dépendant d'un autre logiciel. Nous proposons de supprimer cette dépendance, en traduisant le système avec un langage de programmation non propriétaire, afin de fournir un logiciel Standalone.

4. L'étude expérimentale a été effectuée avec des personnes non-voyantes afin d'évaluer la qualité des cartes en relief. Il serait également intéressant d'effectuer une évaluation avec des personnes malvoyantes. Et, dans une seconde phase expérimentale, il serait intéressant d'évaluer le système de transcription d'images avec des transcripteurs.

5. Nous suggérons également de tester la méthode proposée avec d'autres images plus variées, comme des images représentant des schémas de physique, de géométrie, d'analyse, etc.

Bibliographie

Adams 1994 R. Adams & L. Bischof. *Seeded region growing.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 16(6), p. 641-647, 1994.

Ames 1977 W. F. Ames. *Numerical Methods for Partial Differential Equations, Section 1.6.* Academic Press, New York, 1977. ISBN 0-12-056760-1.

Archambault 2010 D. Archambault. *Interaction et usages des modalités non visuelles, accessibilité des contenus complexes.* Les Recherches, Habilitation à Diriger, Université Paris 6, 2010.

Aschwanden 1992 P. Aschwanden & W. Guggenbül. *Experimental Results from a Comparative Study on Correlation type Registration algorithms.* Proc. ISPRS Workshop, p. 268-282, 1992.

Austry 2006 D. Austry. *Le touchant touché- Exploration phénoménologique du toucher thérapeutique.* Actes du colloque: Phénoménologie(s) de l'expérience corporelle. Clermont-Ferrand, 2 et 3 novembre 2006.

Ballard 1981 D. H. Ballard. *Generalizing the Hough Transform to Detect Arbitrary Shapes.* Pattern Recognition, Vol.13, No.2, p.111-122, 1981

Bau 2010 O. Bau, I. Poupyrev, A. Israr & C. Harrison. *TeslaTouch: Electrovibration for Touch Surfaces.* ACM Symposium on User Interface Software and Technology (ACM UIST) 2010

Bay 2006 H. Bay, T. Tuytelaars & L. V. Gool. *SURF: Speeded Up Robust Features.* 9th European Conference on Computer Vision, Graz, Autriche, 2006

Bay 2008 H. bay, A. Ess, T. Tuyelaars & L. V. Gool. *SURF: Speeded Up Robust Features.* Computer Vision and Image Understanding, Vol.110, No3, p. 246-359, 2008.

Berkeley 1944 G. Berkeley. *Essai d'une théorie nouvelle de la vision,* dans : Œuvres choisies. Tome I, Paris, Aubier, 1944

Berveiller 1991 A. Berveiller. *Comment vivre avec une personne aveugle ?* Editions Josette LYON, p. 166, Paris 1991.

Bezdek 1992 J. C. Bezdek & K. Pal. Sankar. *Fuzzy models for pattern recognition*. IEEE Press, 1992.

Bézier 1986 P. Bézier. *Courbes et Surfaces.* Hermès, 1986.

Bouman 2012 C. A. Bouman. *Connected Component Analysis.* Digital Image Processing, 2012

Bris 1999 M. Bris. *Rapport «Tact images & Training».* Paris, France, p. 1–34, 1999

Bris 2004 M. Bris. *Recommandations pour la transcription de document , documentation du service des documents adaptés pour les déficients visuels.* INSHEA, Suresnes, 2004.

Brock 2012 A. Brock, P. Truillet, B. Oriola, D. Picard, C. Jouffrais. *Design and User Satisfaction of Interactive Maps for Visually Impaired People.* Computers Helping People with Special Needs, p. 544-551, 2012

Brock 2013 A. Brock. *Thèse : Interactive Maps for Visually Impaired People: Design, Usability and Spatial Cognition.* Université Toulouse 3 Paul Sabatier. 2013

Brock 2015 A. Brock, B. Oriola, P. Truillet, C. Jouffrais & D. Picard. *Interactivity Improves Usability of Geographic Maps for Visually Impaired People.* Human–Computer Interaction 30 (2), p.156-194, 2015

Brunswik 1950 E. Brunswik. *The conceptual frameword of phychology.* International Encyclopedia of United Science, Chicago, The University of Chicago Press, 1950.

Burger 1996 D. Burger. *Les nouvelles technologies dans l'éducation des déficients visuels : actes du colloque tenu à Paris, les 10 et 11 juin 1996.* Colloques INSERM, ISSN 0768-3154, p. 237, 1996.

Buser 1982 P. Buser & M. Imbert. *Psychophysiologie sensorielle*. Hermann. ISBN : 2705659447/9782705659448. 1982

C2RP 2005 *Déficience visuelle* –Etudes et Résultats. Lille, France. 2005

Canny 1986 J. Canny. *A Computational Approach To Edge Detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8 : p. 679-714, 1986.

Cattaneo 2011 Z. Cattaneo & T. Vecchi. *Blind vision: The neuroscience of visual impairment.* Rehabilitation, MIT Press, 2011.

Chauvet 1993 P. Chauvet, J. Lopez-Krahe, E. Taflin & H. Maitre. *System for an intelligent office document analysis, recognition and description.* Signal Processing, Volume 32, p. 161-190, 1993.

Chellali 2010 A. Chellali, I. Milleville-Pennel & C. Dumas. *Influence des systèmes de référence sur les profils de collaboration dans un environnement virtuel.* IHM, 08 Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, p. 63-70, 2010

Chen 2014a Y. Chen, Z. Haddad & J. Lopez-Krahe. *Contribution to the automation or the Tactile Images Transcription Process.* Computer Helping People with Special Needs, 14th International Conference, Part 1, ISSN 0302-9743, p. 642-649, 2014

Chen 2014b Y. Chen, Z. Haddad & J. Lopez-Krahe. *Vers une transformation automatique des images pédagogiques en cartes tactiles.* Actes du Congrès Handicap 2014 - Les technologies d'assistance : de la compensation à l'autonomie. p. 140-145. Paris, 2014,

Chen 2014c Y. Chen. *Une nouvelle mise en correspondance des image -Via la fusion de donnée visuelles et inertielles.* Edition Universitaires Européennes, ISBN-978-3-8417-3670-3, 2014.

Chen 2015 Y. Chen, Z. Haddad. *Apport dans la reconnaissance des symboles pour l'accès haptique aux images.* Handicap, Vieillissement, Indépendance, Insertion, Technologie- Actes du colloque Jeunes Chercheurs, jeunes chercheuses, l'Université Paris 8 Vincennes, 2015.

Choi 2013 S. Choi & K. J. Kuchenbecker. *Vibrotactile Display: Perception, Technology, and Applications.* Proceedings of the IEEE, 101, p. 2093–2104, 2013.

Danielsson 1990 P. E. Danielsson & O. Seger. *Generalized and Separable Sobel Operators.* Machine vision for three-dimensional scenes, Herbert Freeman, Academic Press, 1990.

Deriche 1988 R. Deriche, J. P. Cocquerez & G. Almouzi. *An efficient method to*

build early image description. 9ème ICPR, Rome (Italy), 1988.

DeWinter 2013 J.C.F. de Winter. *Using the Student's t-test with extremely small sample sizes.* Practical Assessment, Research & Evaluation, 2013

Duda 1972 R. O Duda & PE. Hart. *Use of the Hough transformation to detect lines and curves in pictures.* Communication of ACM, Vol 15, No.1, p.11-15, 1972.

Dudani 1997 S. A. Dudani, K. J. Breeding, & R. B. McGhee. *Aircraft identification byMoment Invariants.* IEEE Transactions on Computers, Vol C-26, No. 1. 1977

Edman 1992 P. Edman. *Tactile graphics*. New York, USA: AFB press, p. 550, 1992

Encelle 2008 B. Encelle, N. Jessel, J.Mothe. *Aide à l'accessibilité des non-voyants aux partitions musicales.* HANDICAP, Paris, 10/06/2008-12/06/2008, Cépaduès Edition, p.118-122, 2008.

Epshtein 2010 B. Epshtein, E. Ofek & Y. Wexler. *Detecting Text in Natural Scenes with Stroke Width Transform.* Computer Vision and Pattern Recognition (CVPR), IEEE Conference, 2010.

Flusser 2000 J. Flusser. *On the Independence of Rotation Moment Invariants.* Pattern Recognition, vol. 33, p. 1405–1410, 2000.

Flusser 2006 J. Flusser & T. Suk. *Rotation Moment Invariants for Recognition of Symmetric Objects.* IEEE Trans. Image Proc., vol. 15, p. 3784–3790, 2006.

Fonseca 2000 M. J. Fonseca & J. A. Jorge. *Using Fuzzy Logic to Recognize Geometric Shapes Interactively.* Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on 7-10 May 2000.

Fortier 1999 M-F. A. Fortier, D. Ziou, C. Armenakis & S. Wang. *Nouvelles perspectives en détection de contours : Textures et images multispectrales.* Vision Interface 1999, Trois Rivières, Canada, 1999.

Foulke 1992 E. Foulke & P. H. Hatlen. *A collaboration of two technologies. Part 1 : Perceptual and cognitive processes : their implications for visually impaired persons.* British Journal of Visual Impairment,

10, p. 43-46, 1992.

Gaunet 2005 F. Gaunet & X. Briffault. *Exploring the Functional Specifications of a Localized Wayfinding Verbal Aid for Blind Pedestrians: Simple and Structured Urban Areas.* Human-Computer Interaction, 20, p. 267–314, 2005.

Gllavata 2004 J. Gllavata, R. Ewerth & B. Freisleben. *Text Detection in Images Based on Unsupervised Classification of High-Frequency Wavelet Coefficients.* Proceeding or the 17th International Conference on Pattern Recognition, IEEE, Vol.1, p.425-428, 2004.

Harris 1988 C. Harris & M. Stephens. *A Combined Corner and Edge Detector.* Proceedings of The Fourth Alvey Vision Conference (Manchester, UK), p. 147-151, 1988.

Hart 2009 P. E. *Hart. How the Hough Transform was Invented.* IEEE Signal Processing Magazine, Vol 26, Issue 6, p. 18 – 22, 2009

Hatwell 1986 Y. Hatwell. *Toucher l'espace, la main et la perception tactile de l'espace.* Presses universitaires de Lille, p. 21, 14, 40, 1986.

Hatwell 1992 Y. Hatwell. *Le développent des perceptions tactiles et des coordinations visuo-tactiles : implication pour l'éducation des déficients visuel.* Le Courrier de Suresnes, n°56, CNEFEI, p.59, 1992.

Hatwell 2000a Y. Hatwell, A. Streri & E. Gentaz. *Toucher pour Connaître-Psychologie cognitive de la perception tactile manuelle.* Psychologie et science de la pensée, Presse Universitaire de France, p. 193, 2000.

Hatwell 2000b Y. Hatwell & F. Martinez-Sarocchi. *La lecture tactile des cartes et dessins bidimensionnels, et l'accès des aveugles aux œuvres d'art.* Dans Y. Hatwell, A. Streri & E. Gentaz. Toucher pour Connaître-Psychologie cognitive de la perception tactile manuelle. Psychologie et science de la pensée, Presse Universitaire de France, p.267-286, 2000.

Hatwell 2003a Y. Hatwell. *Le développement perceptivo-moteur de l'enfant aveugle. Enfance* 1 Vol. 55, p. 88-94, 2003.

Hatwell 2003b Y. Hatwell. *Psychologie cognitive de cécité précoce. Sciences*

humaines, Dunod, p. 121 , 2003.

HID 2003 *Handicaps-Incapacités-Dépendance.* Revue française des Affaires sociales n°1-2, 57ème année, 2003

Hinton 1993 R. A. L. Hinton. *Tactile and audio-tactile image as vehicle for learning.* In Non-visual Human-computer Interactions, p. 169-179, 1993

Horowitz 1974 S. L. Horowitz & T. Pavlidis. *Picture Segmentation by a directed split-and-merge procedure.* Proceedings of the 2nd International Joint Conference on Pattern Recognition, Copenhagen, Denmark , p. 424-433, 1974

Hough 1962 P. V. C Hough. *Methods and Means for Recognizing Complex Pattern.* No.3, December 1962.

Hu 1962 M. K. Hu. *Visual Pattern Recognition by Moment Invariants.* IEEE Transactions on Information Theory, Vol IT-8, 1962.

Jacobson 1996 R. D. Jacobson. *Talking tactile maps and environmental audio beacons : An orientation and mobility development tool for visually impaired people.* In ICA Comission on maps and diagrams for blind and visually impaired people: Needs, solutions and developments, p. 1–22, Ljubjiana, Slovenia.1996

Jessel 2002 N. Jessel, B. Tornil, B. Encelle. *BME Braille Music Editor : une solution pour l'accessibilité de l'information musicale par les non-voyants.* Journée thématique BrailleNet 2002 : Livres / Books et Cartables électroniques pour l'intégration scolaire des jeunes handicapés visuels, Paris (FRANCE), 03/05/2002.

Jessel 2004 N. Jessel, B. Tornil, B. Encelle. *Using SVG and a force feedback mouse to enable blind people to access 'Graphical' Web based documents.* Computers Helping People with Special Needs, p. 228-235. Springer Berlin Heidelberg, 2004.

Jung 2008 C. Jung, Q. Liu & J. Kim. *A new approach for text segmentation using a stroke filter.* Signal Processing 88, p.1907-1916, 2008.

Jung 2009 C. Jung, Q. Liu & J.Kim. *A stroke filter and its application to text localization.* Pattern Recognition Letters 30, p. 114-122, 2009.

Karpathy 2014a A. Karpathy, A. Joulin et F. F. Li. *Deep fragment embeddings for*

bidirectional image sentence mapping. In Advances in neural information processing systems. p. 1889-1897, 2014.

Karpathy 2014b A. Karpathy et F. F. Li. *Deep visual-semantic alignments for generating image descriptions.* arXiv preprint arXiv:1412.2306. 2014.

Kimura 1993 T. D. Kimura, A. Apte & V. Vo. *Recognizing multi-stroke geometric shapes: An experimental evaluation.* In Proceedings of the ACM conference on User Interface and Software Technology (UIST'93), p. 121–128, Atlanta, GA, ACM Press, 1993.

Kostopoulos 2007 K. Kostopoulos, K. Moustakas, D. Tzovaras, G. Nikolakis, C. Thillou & B. Gosselin. *Haptic Access To Conventional 2d Maps For The Visually Impaired.* Journal on Multimodal User Interfaces, Vol. 1, No. 2, 2007.

Lahav 2003 O. Lahav & D. Mioduser. *A blind person's cognitive mapping of new spaces using a haptic virtual environment.* Journal of Research in Special Educational Needs, Volume 3, Number 3, p. 172–177, 2003.

Lederman 1979 S. J. Lederman & D. Kinch. *Texture in tactual maps and graphics for the visually handicapped.* Journal of Visual Impairment & Blindness, 73, p. 217–227, 1979.

Lederman 1987 S. J. Lederman & R. L. Klatzky. *Hand movement: A window into haptic object recognition.* Cognitive Psychology, 19, p. 342-368, 1987.

Li 2000 H. Li, D. Doermann & Omid Kia. *Automatic Text Detection and Tracking in Digital Video.* IEEE Trans. Image Process., vol. 9, p. 147–156, Jan. 2000.

Li 2005 Z. Li, X. Huang, Y. Li, M. Wang. *Ren yin gong cheng Ergonomics.* HUA TAi Publishing, Taiwan, Chine. ISBN-10: 957609562X/ISBN-13:9789576095627. 2005.

Lia 1998 J. Lia & R. M. Gray. *Text and Pictures Segmentation by the Distribution Analysis of Wavelet Coefficients.* IEEE Int'l. Conference on Image Processing, p. 790-794, 1998

Lloyd 2000 R. Lloyd. *Understanding and learning maps.* In R. Kitchin & S.

Freundschuh (Eds.), Cognitive Mapping: Past Present and Future, p. 84–107, NewYork: Taylor & Francis, 2000

Lobben 2012 A. Lobben & M. Lawrence. *The Use of Environmental Features on Tactile Maps by Navigators Who Are Blind.* The Professional Geographer, 64, 95–108, 2012.

Lopez-Krahe 1988 J. Lopez-Krahe & P. Pousset. *Transformée de Hough discrète et bornée. Application à la détection de droites parallèles et du réseau routier.* Traitement du Signal, Vol. 5, No. 4, p.281-290, 1988.

Lowe 2004 D. G. Lowe. *Distinctive image features from scale-invariant keypoints.* Int. J. of Computer Vision, 60, 2, p. 91-110, 2004.

MacQueen 1967 J. B. MacQueen. *Some Methods for classification and Analysis of Multivariate Observations.* Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297, 1967

Maitre 1985 H. Maitre. *Un panorama de la transformation de Hough*. Traitement du Signal. Volume 2-N °4 p.305-317, 1985.

Milar 1994 S. Millar. *Understanding and Representing Space: Theory and Evidence from Studies with Blind and Sighted children.* Oxford, Oxford University Press, 1994.

Montagu 1979 A. Montagu. *La peau et le toucher- Un premier langage.* Seuil, France, 1979.

Moravec 1980 H. Moravec. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover.* PhD, Carnegie-Mellon University (CMU), Robotics Institute, 1980.

Morrongiello 1995 BA. Morrigiello, B. Timney, GK. Humphrey, S. Anderson & C. Skory. *Spatial knowledge in blind and sighted children.* J Exp Chil Psychol. 59(2) : 211-33, 1995..

Narayan 2012 E. Narayan, Y. Birla & G. K. Tak. *Enhancement of Fuzzy C-Means Clustering using EM Algorithm.* International Journal of Computer Applications (0975 – 8887) Volume 43– No.13, April 2012.

Nino 1960 F. Nino. *Qu'est –ce qu'un aveugle*. Le Courier –Le monde en relief des aveugles, p.4, Juin 1960.

Nomura 1990 J. Nomura. *The secrets of colors-latest science of color introduction.* BunkoNesco Books, Japonais. ISBN-10 : 4890360700/ISBN-13 : 9784890360703. 1990

Pissaloux 2010 E. Pissaloux, Y. Chen & R. Velazquez. *Image Matching Optimization via Vision and Inertial Data Fusion: Application to Navigation of the Visually Impaired.* International Journal of Image and Graphics, Vol. 10, No3, p. 1-14, 2010

Prewitt 1970 J. M. S. Prewitt. *Object enhancement and extraction.* Picture processing and Psychopictoris, Academic Press, 1970.

Priya 2011 R. K. Priya, C. Thangaraj et C. Kesavadas. *Fuzzy C-Means method for Colour Image Segmentation with L*U*V* Colour transformation.* IJCSI International Journal of ComputerScience Issues, Special Issue, ICVCI-2011, Vol. 1, Issue 1, November 2011

Rege 2012 P. P. Rege & C. A. Chandrakar. *Text-Image Separation in Document Images Using Boundary/Perimeter Detection.* ACEEE Int. J. on Signal & Image Processing, Vol. 03, No. 01, Jan 2012.

Rice 2005 M. T. Rice, R. D. Jacobson, R. G. Golledge & D. Jones. *Design considerations for haptic and auditory map interfaces.* Cartography and Geographic Information Science, 32, p. 381–391, 2005.

Roberts 1963 L. G. Roberts. *Machine perception of three-dimensional soups*. Doctoral dissertation, Massachusetts Institute of Technology, 1963.

Rosenfield 1969 A. Rosenfield. *Picture Processing by Computer*. Academic, New York, 1969.

Sahbani 2005 M. H. Sahbani & K. Hamrouni. *Segmentation d'images texturées par transformée en ondelettes et classification C-moyenne floue*. SETIT 2005.

Shambharkar 2011 S. Shambharkar & S. Tirpude. *Fuzzy C-Means Clustering For Content Based Image Retrieval System.* ICBMG 2011, IPCSIT vol.20, IACSIT Press, Singapore, 2011.

Sherwood 2006 L. Sherwood & A. Lockhart. *Physiologie humaine.* De Boeck

Université, Bruxelles. p. 148. 2006.

Smith 1997 S. M. Smith, J. M. Brady. *SUSAN—A New Approach to Low Level Image Processing.* Int. J. of Computer Vision 23(1), p. 45-78, 1997.

Sobel 1968 I. Sobel & G. Feldman. *A 3*3 Isotropic Gradient Operator for Image Processing.* Presented at the Stanford Artificial intelligence Project in 1968.

Tatham 1991 A. F. Tatham. *The design of tactile maps : theoretical and practical considerations.* In M. Rybaczak & K. Blakemore (Eds.), Proceedings of international cartographic association : mapping the nations, London, UK : ICA, p. 157–166, 1991

Thinus-Blanc 1997 C. Thinus-Blanc & F. Gaunet. *Representation of space in blind persons : vision as a spatial sense.* Psychological Bulletin, 121, 20-42, 1997.

Tobin 2003 M. J. Tobin, J. Greaney & E. Hill. *Le Braille : problèmes de structure, d'enseignement der d'évaluation.* In In Y. Hatwell, A. Streri & E. Gentaz. Toucher pour Connaître-Psychologie cognitive de la perception tactile manuelle. Psychologie et science de la pensée, Presse Universitaire de France, p. 237-254, 2000.

Tremeau 1997 A. Tremeau & N. Borel. *A region growing and merging algorithm to color segmentation.* Pattern Recognition Volume 30, Issue 7, p. 1191–1203, July 1997.

Truchetet 1998 F. Truchetet. *Ondelette pour le signal numérique*. Editions Hernes, p. 11, 1998.

Ungar 2000 S. Ungar. *Cognitive mapping without visual experience.* In R. Kitchin & S. Freundschuh (eds) Cognitive Mapping: Past Present and Future, London: Routledge, 2000.

Vanderheiden 2012 G. C. Vanderheiden & J. B. Jordan. *Design for people with functional limitation.* In G. Salvendy (Ed.), Handbook of human factors and Ergonomics. Hoboken, NJ, USA: John Wiley & Sons, Inc. 2012

Villasenor 1995 J. D. Villasenor, B. Belzer & J. Liao. *Wavelet Filter Evaluation for image Compression.* IEEE Transactions on Image Processing,

Vol.4, No.8, 1995

Wacker 1969 A. G. Wacker. *A cluster approach to finding spatial boundaries in multi-spectral imagery.* LARS Information Note 122969, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, 1969.

Wahl 1981 F. M. Wahl, K. Y. Wong, & R. G. Casey. *Block Segmentation and Text Extraction in Mixed Text/Image Documents.* Research Report RJ 3356, IBM Research Laboratory, SanJose, CA, 1981.

Wahl 1982 F. M. Wahl. *A New Distance Mapping and its Use for Shape Measurement on Binary Patterns.* Research Report RJ 3361, IBM Research Laboratory, San Jose, CA, 1982.

Wang 2009 Z. Wang, B. Li, T. Hedgpeth & T. Haven. *Instant Tactile-Audio Map: Enabling Access to Digital Maps for People with Visual Impairment.* In Assets'09 Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility. Pittsburgh, Pennsylvania, USA: ACM Press. p. 43–50. 2009

WHO 2012 WHO. *Visual Impairment and blindness Fact Sheet N °282*. World Health Organization, 2012.

Xie 2002 Y. Xie & Q. Ji. *A New Efficient Ellipse Detection Method.* In Proc. ICPR (2), p. 957-960, 2002.

Yatani 2012 K. Yatani, N. Banovic & K. Truong. *Space Sense: representing geographical information to visually impaired people using spatial tactile feedback.* In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems – CHI'12 New York, New York, USA: ACM Press, p. 415 – 424, 2012

Zadeh 1965 L. A. Zadeh. *Fuzzy sets.* Information and Control 8 (3): p. 338–353, 1965.

Zucker 1976 S. W. Zucker. *Region growing: Childhood and adolescence.* Computer Graphics and Image Processing Volume 5, Issue 3, Pages 382–399, September 1976.

Liste des publications

Edwige PISSALOUX, **Yong CHEN**, Ramiro VELAZQUEZ. I*mage Matching Optimization via Vision and Inertial Data Fusion: Application to Navigation of the Visually Impaired.* International Journal of Image and Graphics, Vol. 10, No3, p. 1-14, 2010.

Yong CHEN, Jaime LOPEZ KRAHE. *La Reconnaissance des Formes dans l'Enrichissement des Images. Application pour l'Accessibilité Tactile des Personnes Aveugles.* Handicap, Vieillissement, Indépendance, Technologie- Actes du colloque Jeunes Chercheurs, jeunes chercheuses, IFRATH, ISBN-9782953689938,, Saint Denis, du 13 au 14 juin 2013, p. 7-16.

Yong CHEN, Jaime LOPEZ KRAHE. *Shape Recognition, Image Analysis and Artificial Vision for Blind People Using Mainstream Devices.* V International Congress on Design, Research Networks and Technology for all. Madrid, SPAIN, Proc. DRT4ALL 2013, p. 228-234.

Yong CHEN, Zehira HADDAD, Jaime LOPEZ KRAHE. *Vers une transformation automatique des images pédagogiques en cartes tactiles* . Actes du Congrès Handicap 2014- Les technologies d'assistance : de la compensation à l'autonomie, IFRATH, ISBN-9782953689945, Paris 11-13 juin 2014, p. 140-145.

Yong CHEN, Zehira HADDAD, Jaime LOPEZ KRAHE. *Contribution to the automation or the Tactile Images Transcription Process.* Computer Helping People with Special Needs, 14th International Conference, ICCHP 2014, Part 1, ISSN 0302-9743, page 642-649.

Gérard UZAN, **Yong CHEN**, Laurie JACOB. *Règles à appliquer aux ouvrages numérisés pour leur accessibilité par des personnes déficientes visuelles.* Laboratoire CHArt THIM, Université Paris 8, Saint-Denis, 2014 [Rapport interne]

Yong CHEN. *Une nouvelle mise en correspondance des images. Via la fusion de données visuelles et inertielles.* Edition Universitaires Européennes, ISBN-9783841736703, 2014.

Yong CHEN, Zehira HADDAD, Jaime LOPEZ KRAHE. *A New Approach for Automatic Transformation of Pedagogic Images into Tactile Images.* AMSE JOURNALS-AMSE

IFRATH Publication –2014- Modelling C; Vol. 75; N °2; pp 43-54

Yong CHEN, Zehira HADDAD. *Apport dans la reconnaissance des symboles pour l'accès haptique aux images.* Handicap, Vieillissement, Indépendance, Insertion, Technologie- Actes du colloque Jeunes Chercheurs, jeunes chercheuses, IFRATH, ISBN 9782953689952, p.31-39, 2015.

Laurie JACOB, **Yong CHEN**, Gérard UZAN. *Correction vocalisée et mise en accessibilité de documents numérisés : l'exemple du projet « Correct ».* Handicap, Vieillissement, Indépendance, Insertion, Technologie- Actes du colloque Jeunes Chercheurs, jeunes chercheuses, IFRATH, ISBN 9782953689952, p.73-80, 2015.

Article accepté:

Zehira HADDAD, **Yong CHEN**, Jaime LOPEZ KRAHE. *A pattern recognition to make accessible the geographic images for blind and visually impaired.* IEEE International Conference on Image Processing, ICIP 2015, September 27-30, Québec City, Canada, 2015.

Annexe A

Codage Braille

1. La numérotation des six points d'une cellule braille :

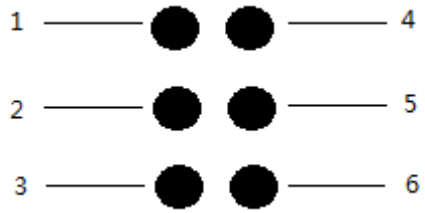


Figure A1. Les six points d'une cellule Braille

2. Dans les tableaux suivants, on présente les symboles de base codés dans notre travail :

Tableau A1. Les chiffres

Caractère braille	***Points embossés***	***Symbole***	***Signe***
⠁	1-6	un	1
⠃	1-2-6	deux	2
⠉	1-4-6	trois	3
⠙	1-4-5-6	quatre	4
⠑	1-5-6	cinq	5
⠋	1-2-4-6	six	6
⠛	1-2-4-5-6	sept	7
⠓	1-2-5-6	huit	8
⠊	2-4-6	neuf	9
⠚	3-4-5-6	zéro	0

Tableau A2. Symboles braille codés

Caractère braille	*Points embossés*	*Symbole de base*	*Signe*	*Caractère braille*	*Points embossés*	*Symbole de base*	*Signe*
⠁	1	lettre a	a	⠺	2-4-5-6	lettre w	w
⠃	1-2	lettre b	b	⠭	1-3-4-6	lettre x	x
⠉	1-4	lettre c	c	⠽	1-3-4-5-6	lettre y	y
⠙	1-4-5	lettre d	d	⠵	1-3-5-6	lettre z	z
⠑	1-5	lettre e	e	⠯	1-2-3-4-6	c cédille	ç
⠋	1-2-4	lettre f	f	⠿	1-2-3-4-5-6	e accent aigu	é
⠛	1-2-4-5	lettre g	g	⠷	1-2-3-5-6	a accent grave	à
⠓	1-2-5	lettre h	h	⠮	2-3-4-6	e accent grave	è
⠊	2-4	lettre i	i	⠾	2-3-4-5-6	u accent grave	ù
⠚	2-4-5	lettre j	j	⠢	1-6	a accent circonflexe	â
⠅	1-3	lettre k	k	⠣	1-2-6	e accent circonflexe	ê
⠇	1-2-3	lettre l	l	⠬	1-4-6	i accent circonflexe	î
⠍	1-3-4	lettre m	m	⠻	1-4-5-6	o accent circonflexe	ô
⠝	1-3-4-5	lettre n	n	⠱	1-5-6	u accent circonflexe	û
⠕	1-3-5	lettre o	o	⠫	1-2-4-6	e tréma	ë
⠏	1-2-3-4	lettre p	p	⠯	1-2-4-5-6	i tréma	ï
⠟	1-2-3-4-5	lettre q	q	⠳	1-2-5-6	u tréma	ü
⠗	1-2-3-5	lettre r	r	⠪	2-4-6	oe ligaturé	oe
⠎	2-3-4	lettre s	s	⠆	2-3	point-virgule	;
⠞	2-3-4-5	lettre t	t	⠒	2-5	deux-points	:
⠥	1-3-6	lettre u	u	⠖	2-3-5	point d’exclamation	!
⠧	1-2-3-6	lettre v	v	⠲	2-5-6	point	.
⠂	2	virgule	,	⠦	2-6	point d’interrogation	?

Annexe B

Images testées en relief

Carte a) :

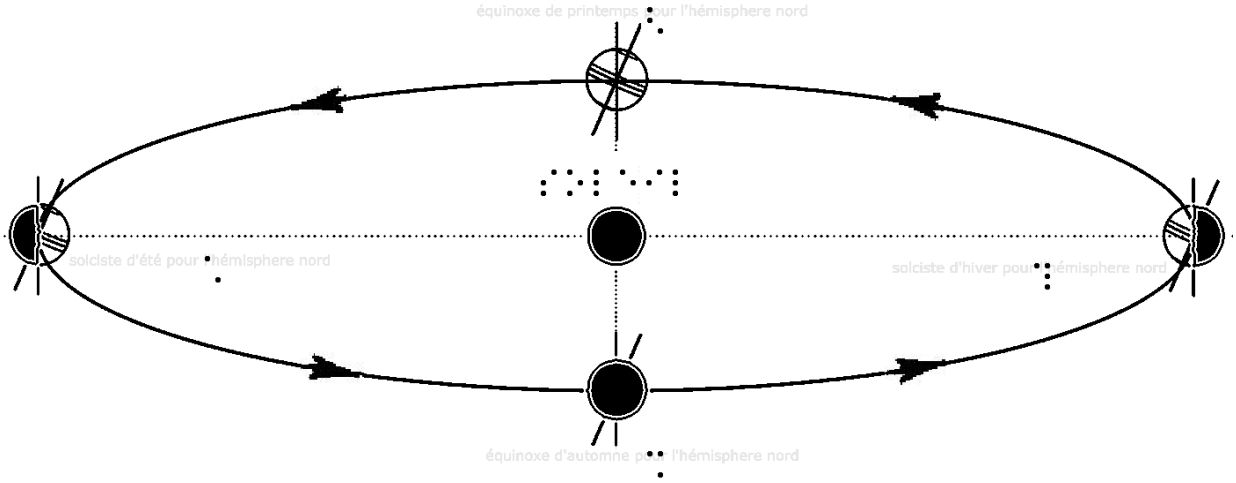
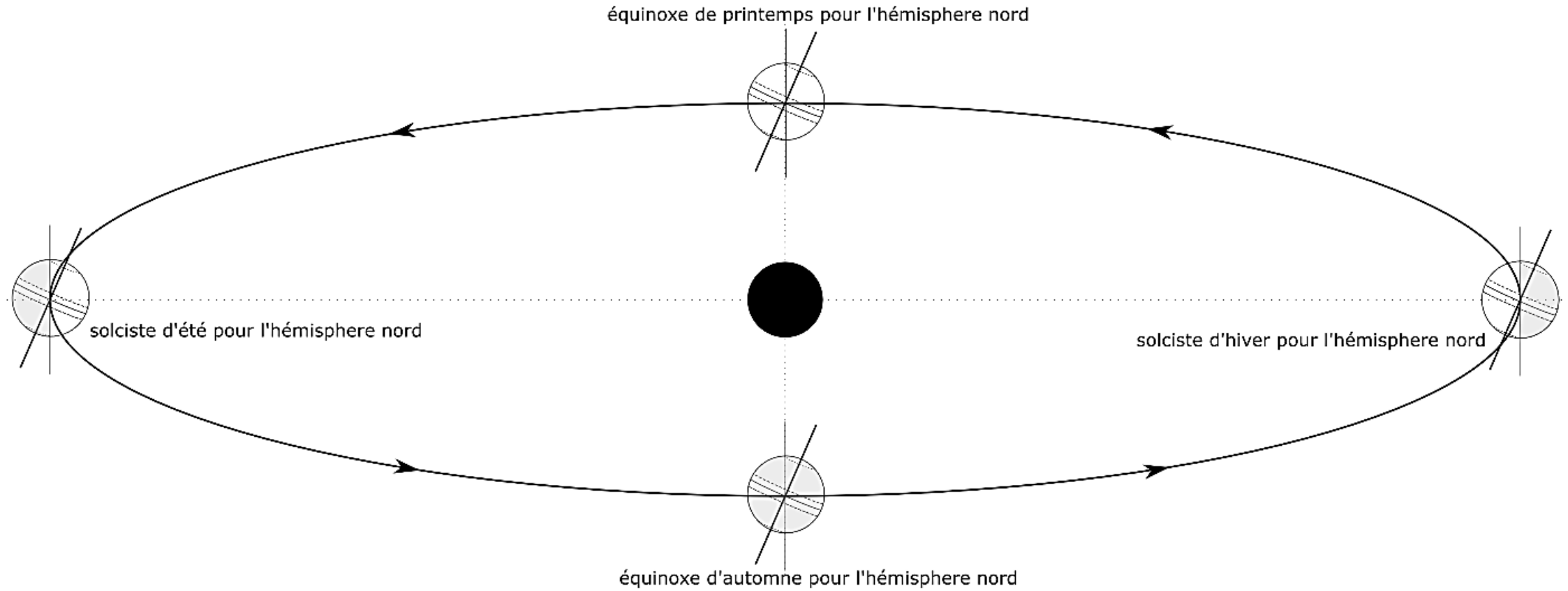
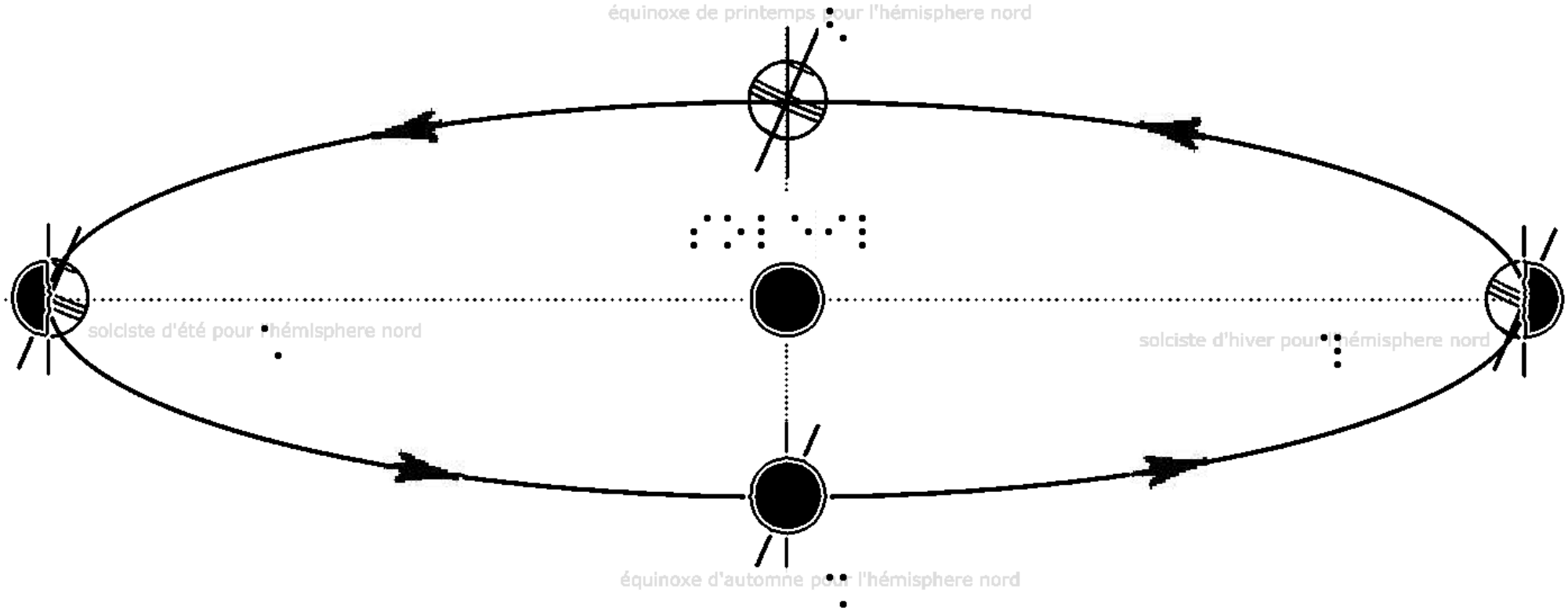


Figure B1. Image astronomique (Carte a)

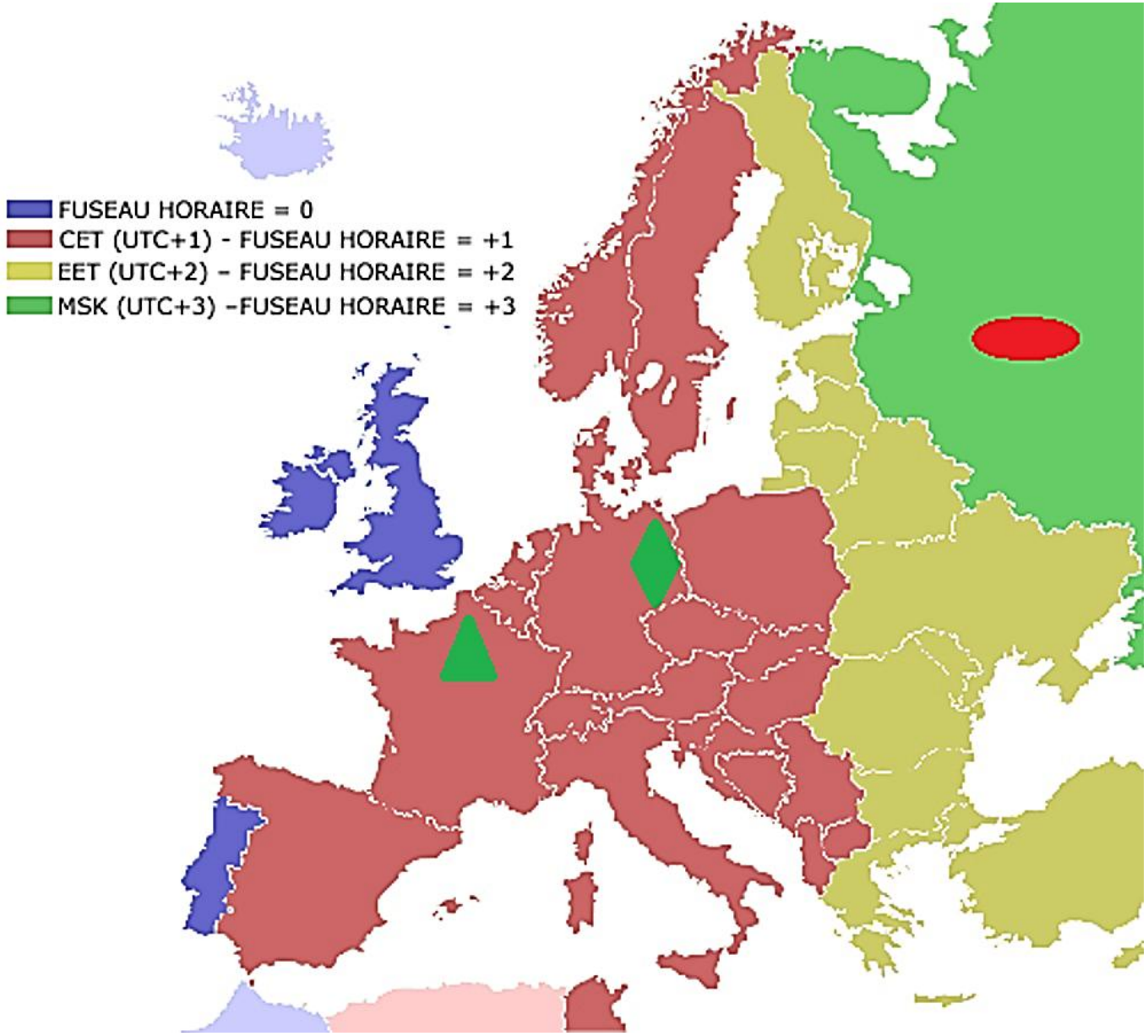

équinoxe de printemps pour l'hémisphere nord
solciste d'été pour l'hémisphere nord
solciste d'hiver pour l'hémisphere nord
équinoxe d'automne pour l'hémisphere nord

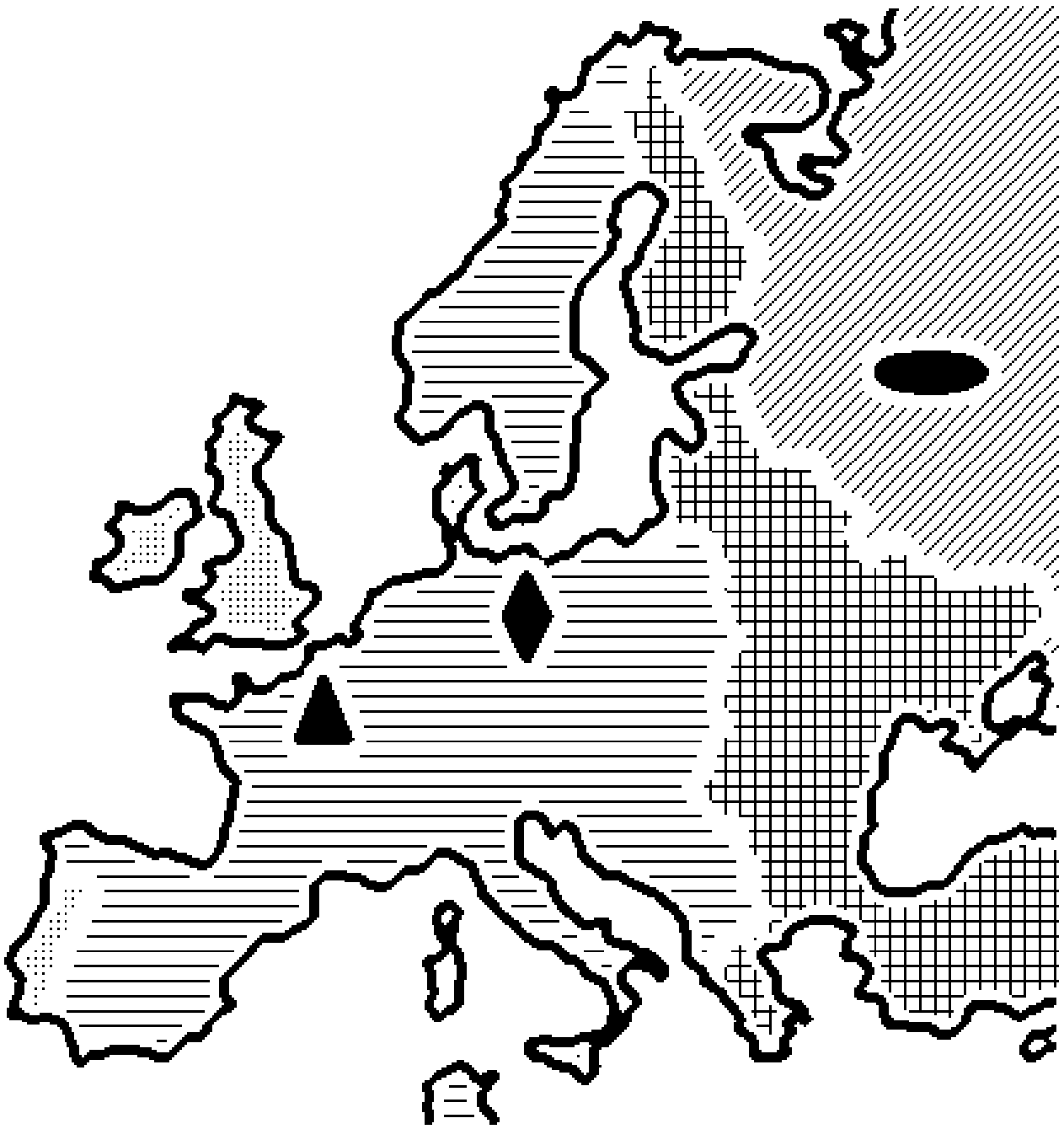

équinoxe de printemps pour l'hémisphere nord
solciste d'été pour hémisphere nord
solciste d'hiver pour hémisphere nord
équinoxe d'automne pour l'hémisphere nord

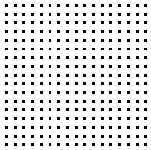
Catre b) :

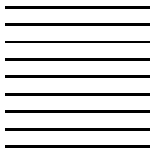


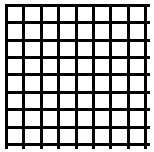
Figure B2. Carte géographique (Carte b)

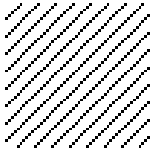

FUSEAU HORAIRE = 0
CET (UTC+1) - FUSEAU HORAIRE = +1
EET (UTC+2) – FUSEAU HORAIRE = +2
MSK (UTC+3) –FUSEAU HORAIRE = +3

















Carte c) :

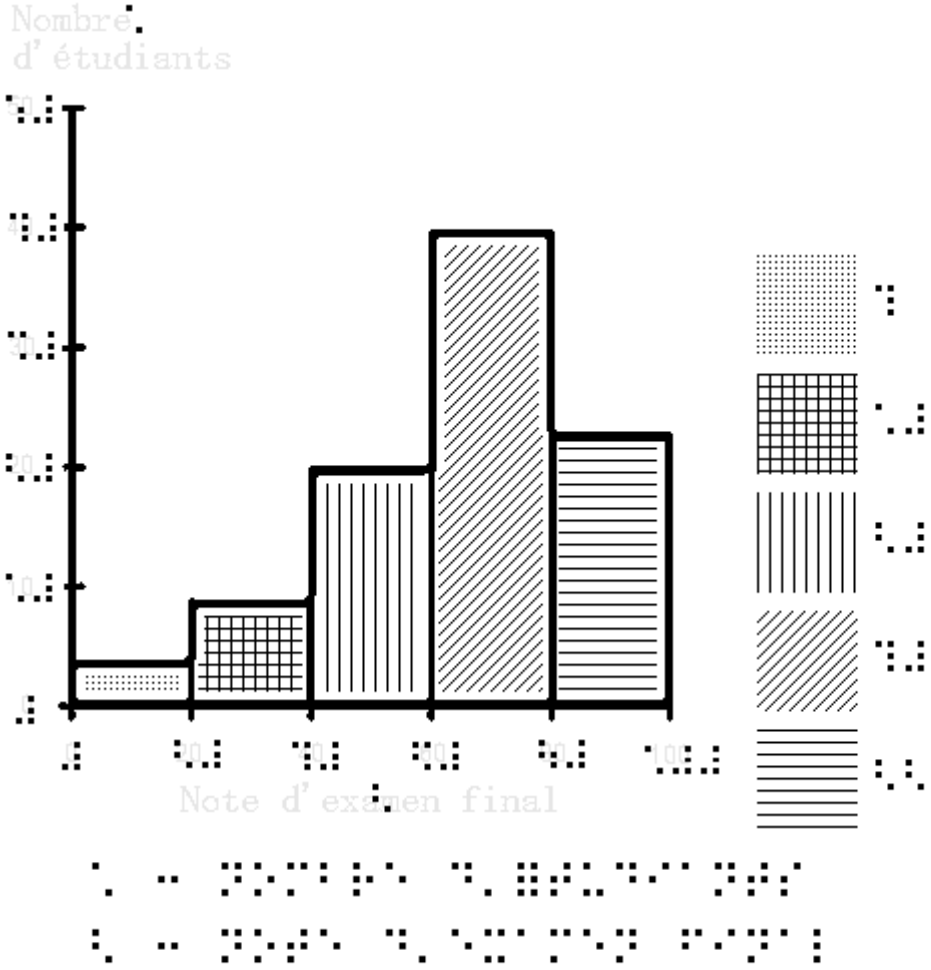
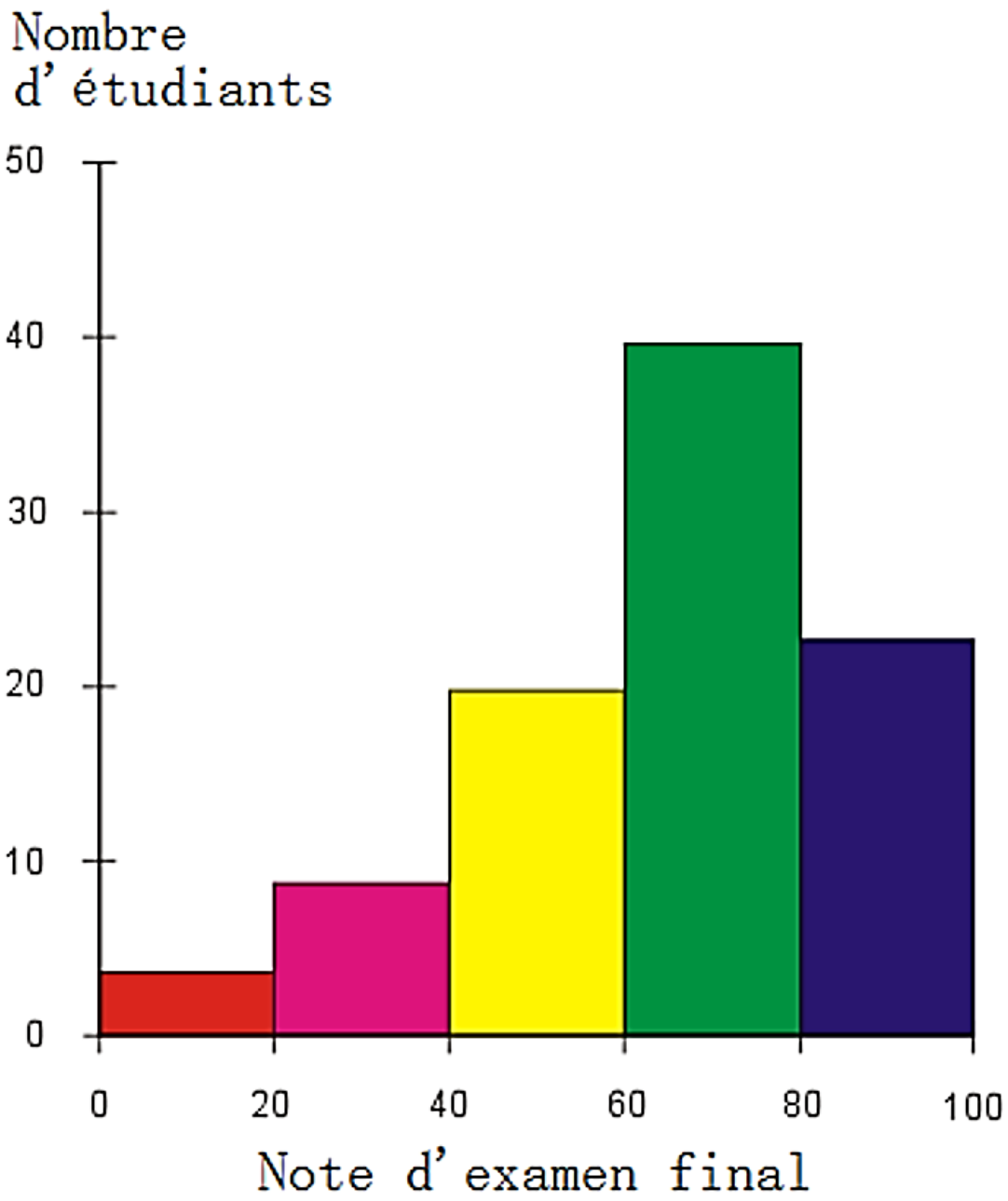
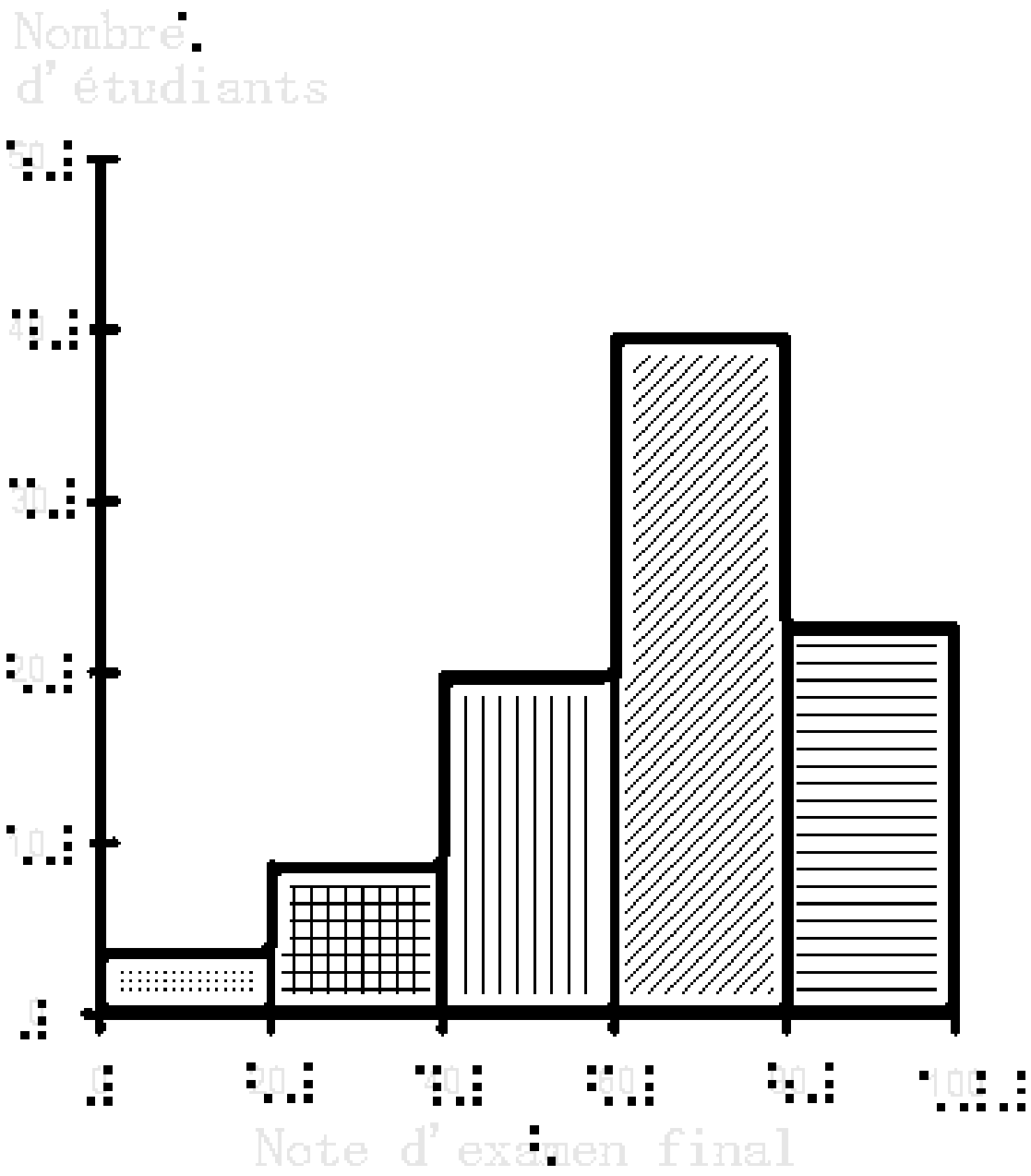
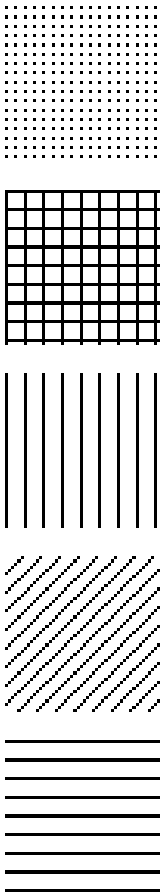


Figure B3. Image histogramme (Carte c)


Nombre
d'étudiants
50
40
30
20
10
0
0
20
40
60
80
100
Note d'examen final


Nombre d'étudiants
0
10
20
30
40
0
20
40
60
80
100
Note d'examen final



Carte d) :

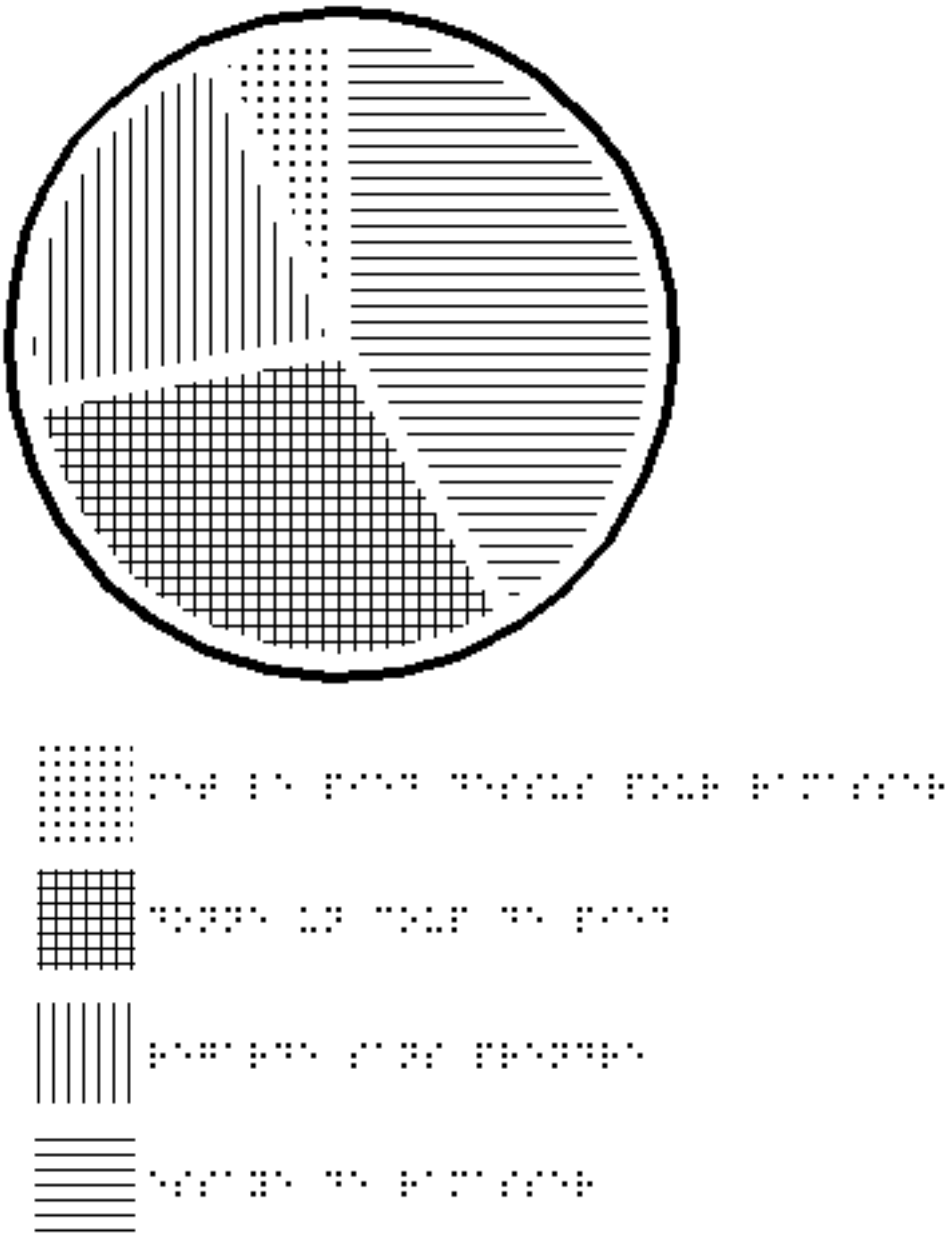
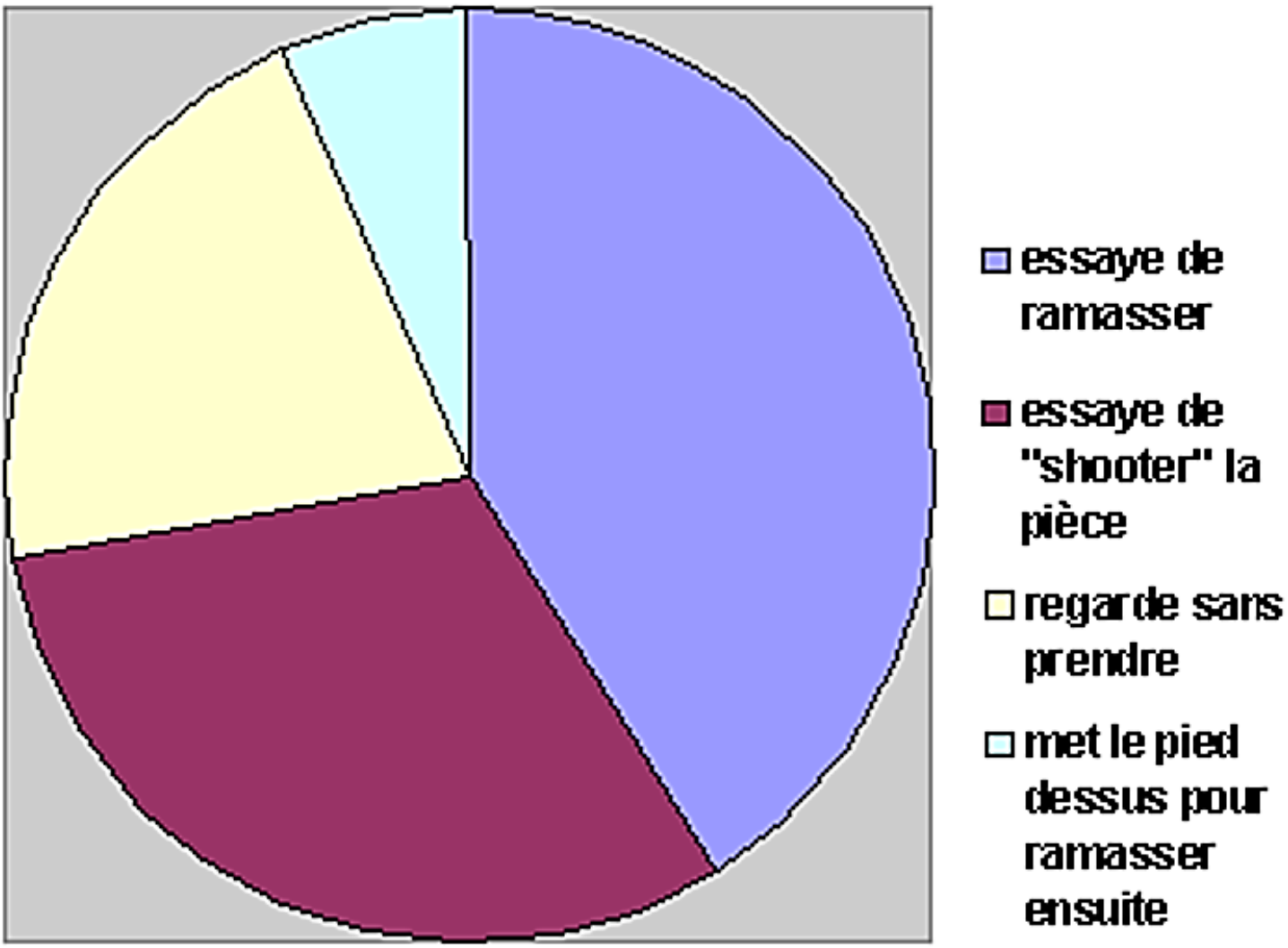
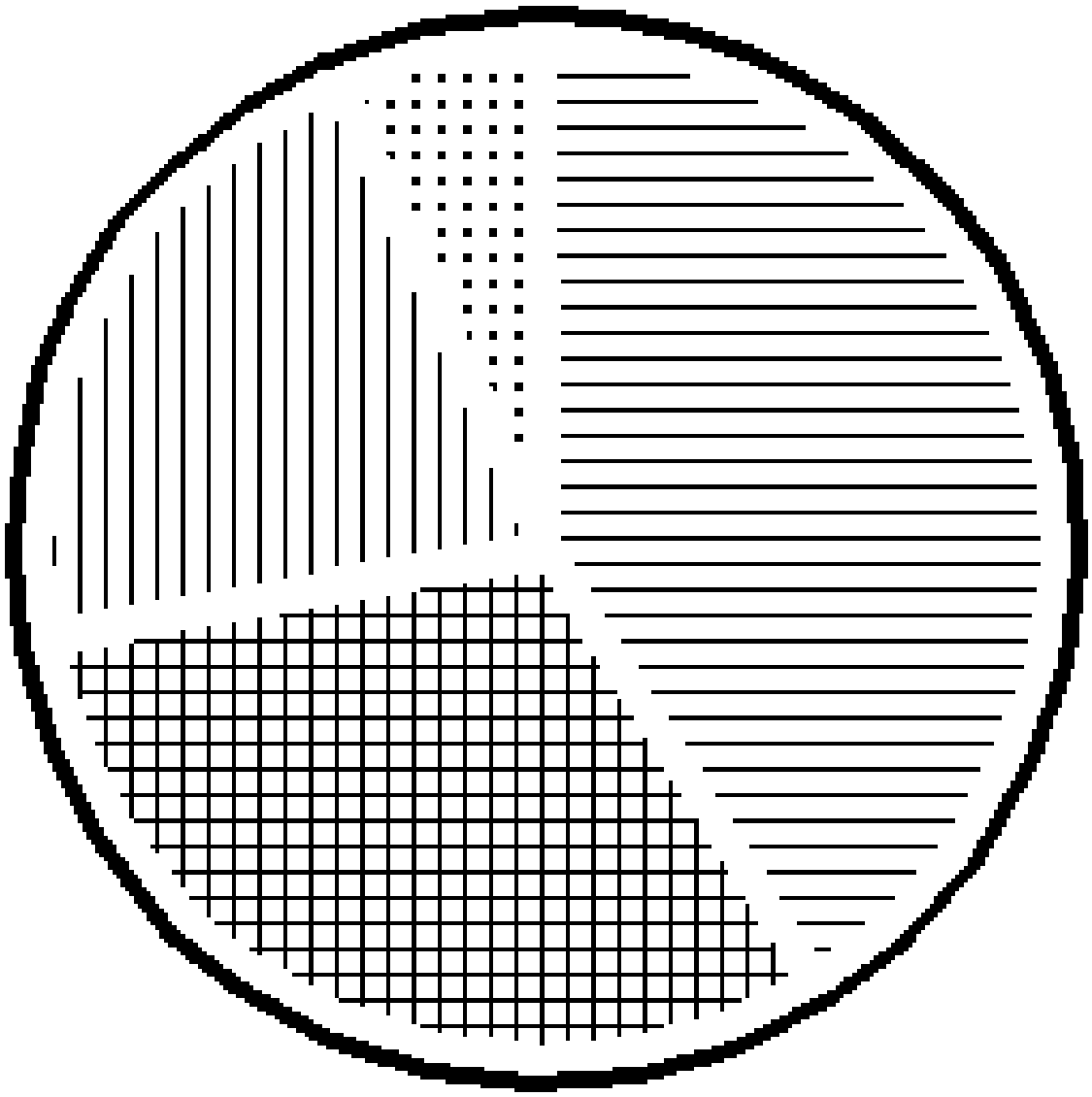


Figure B4. Image camembert (Carte d)


essaye de ramasser
essaye de "shooter" la pièce
regarde sans prendre
met le pied dessus pour ramasser ensuite



Carte e) :

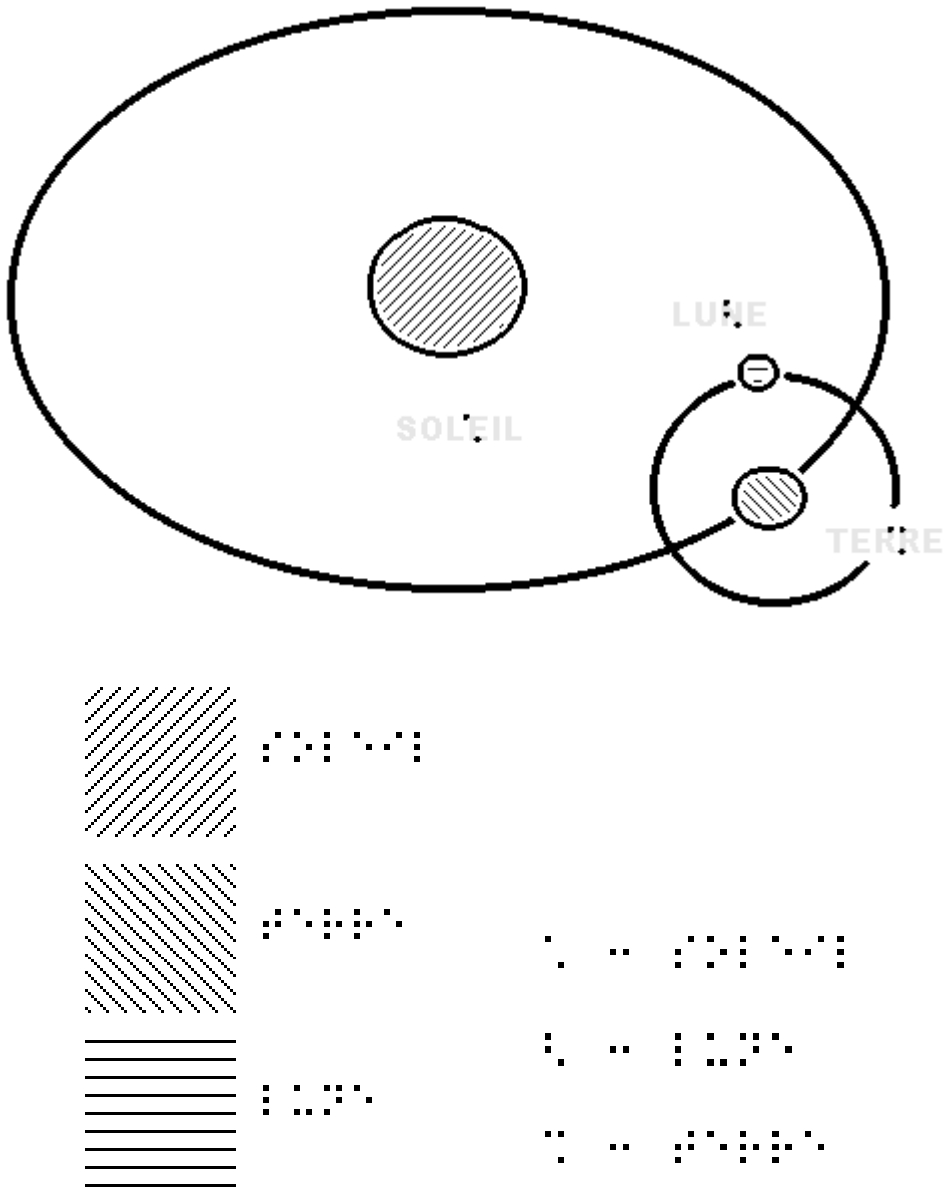
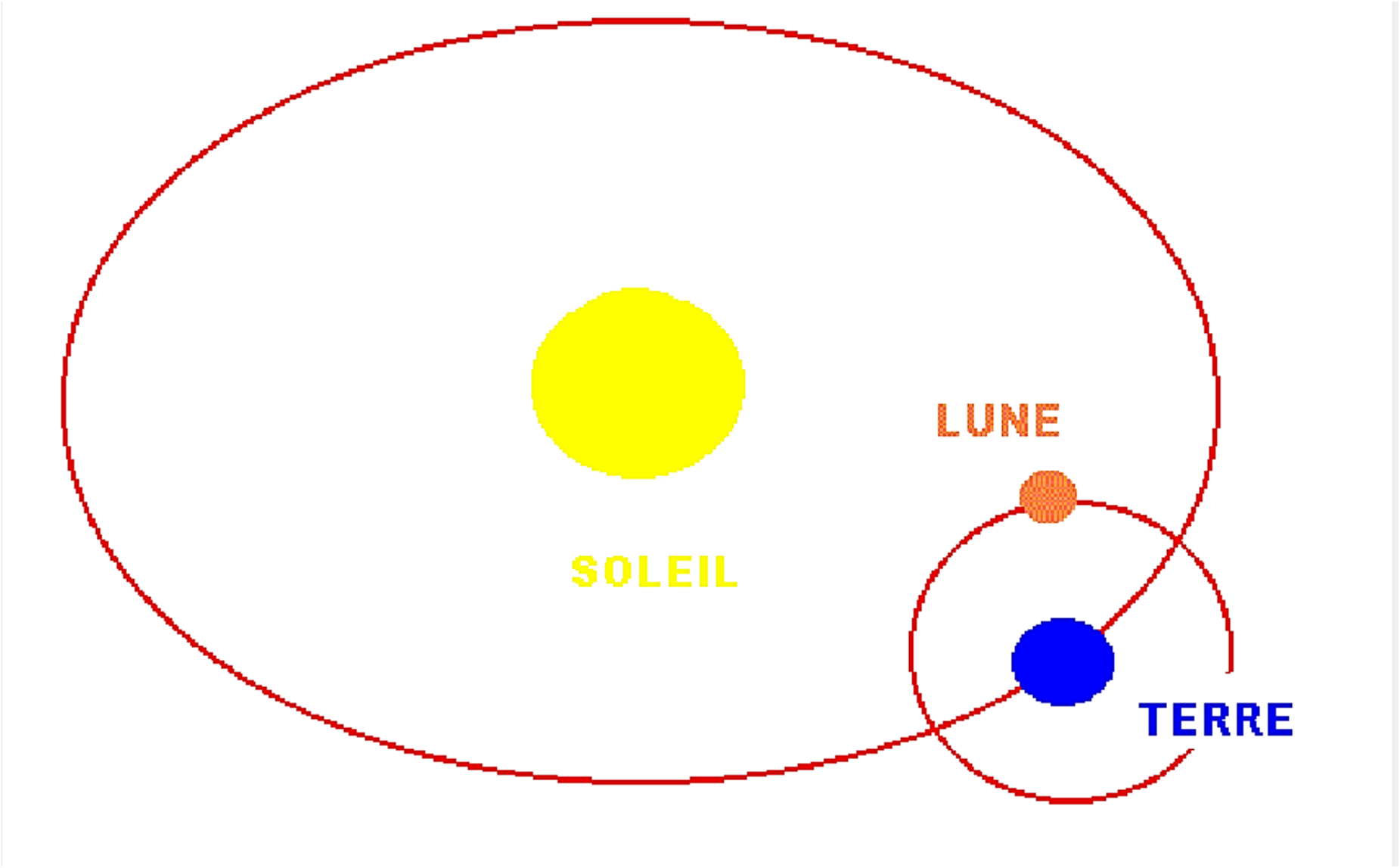
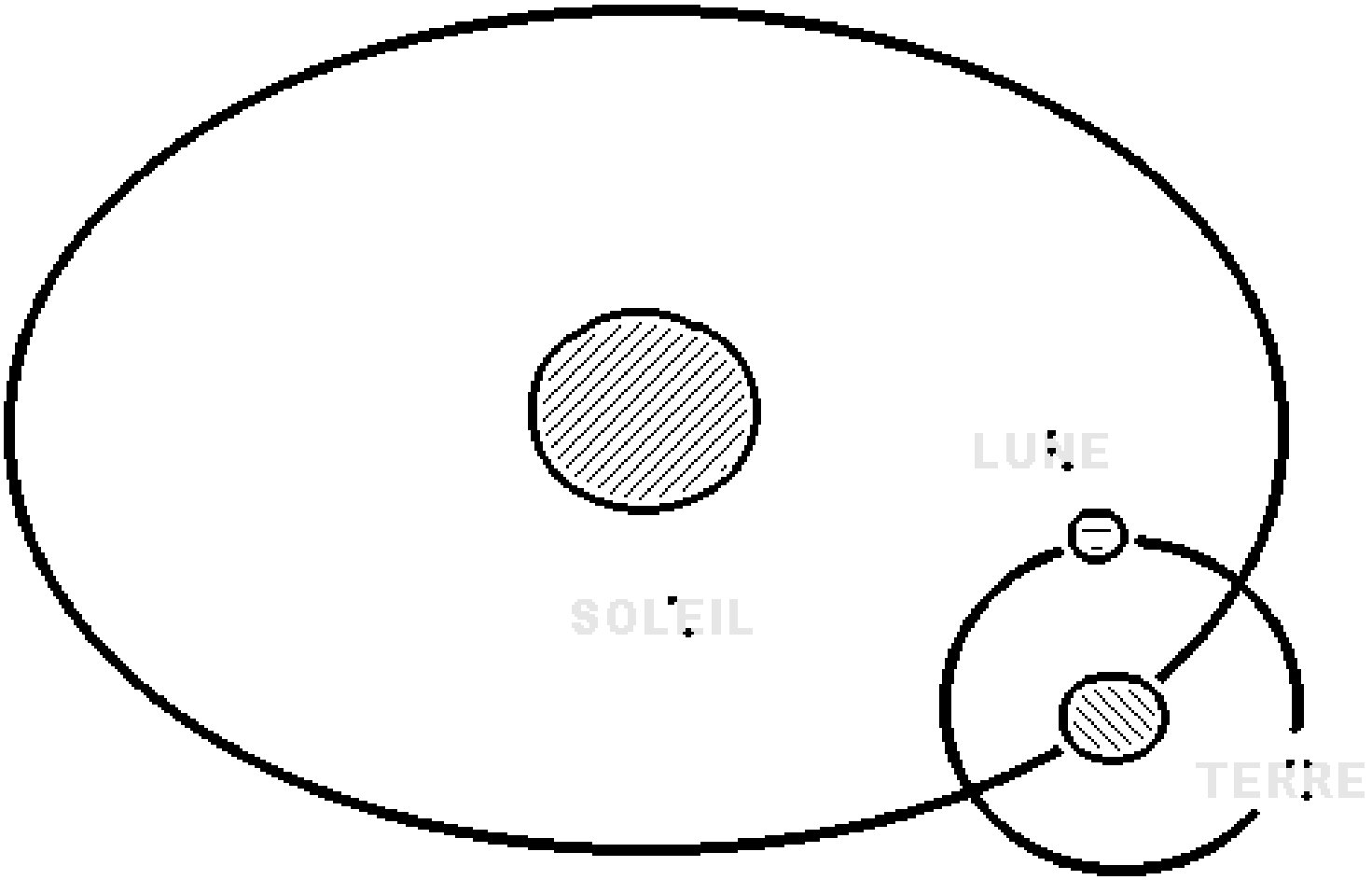
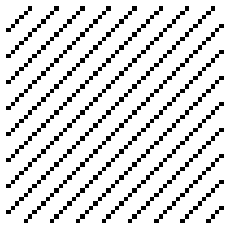
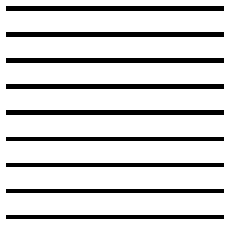


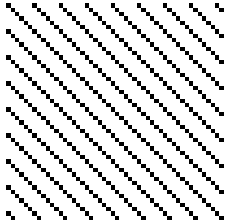
Figure B5. Image astronomique (Carte e)


LUNE
SOLEIL
TERRE


LUNE
SOLEIL
TERRE





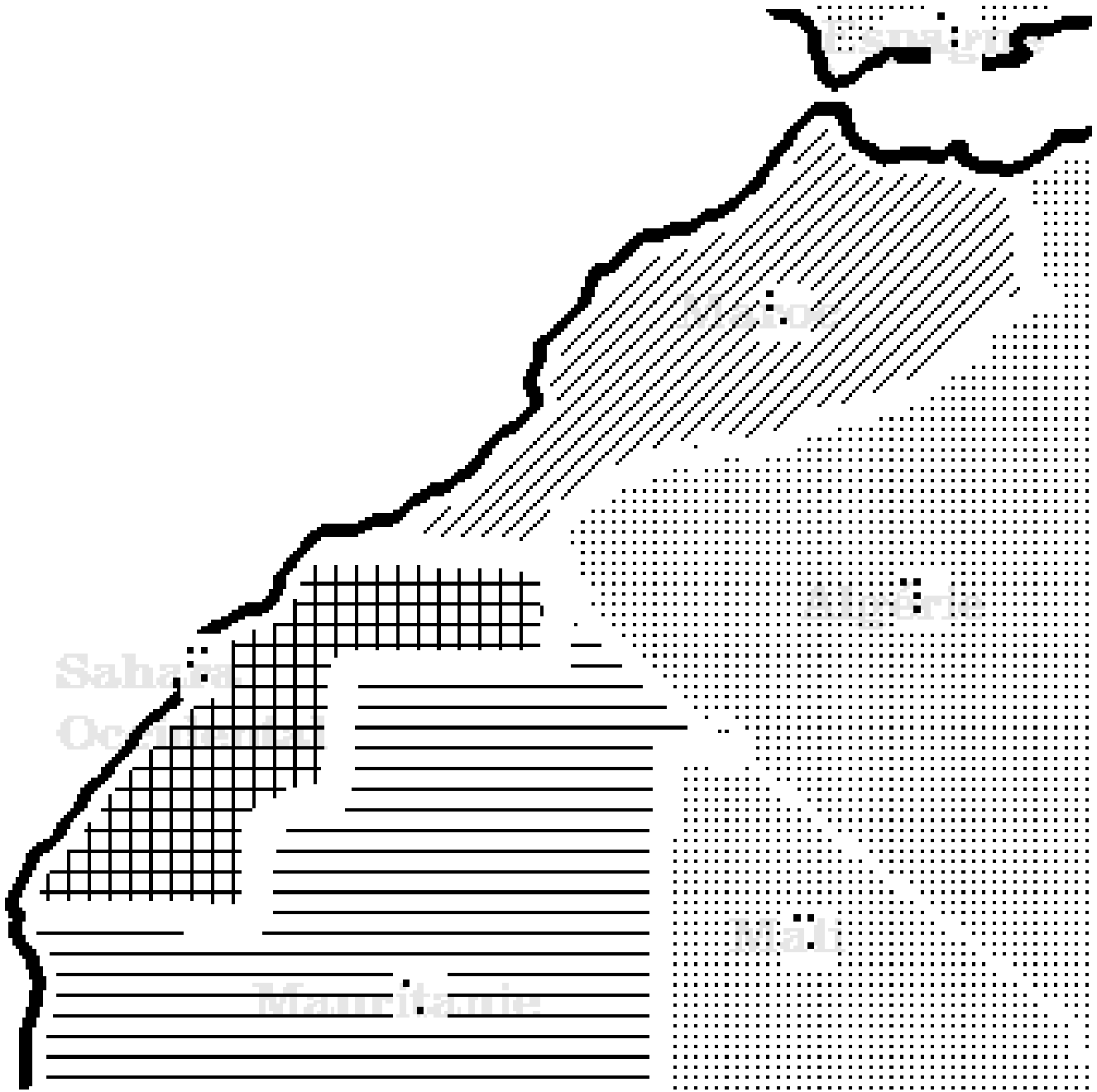


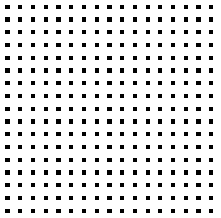
Carte f) :

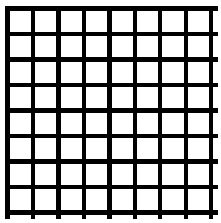


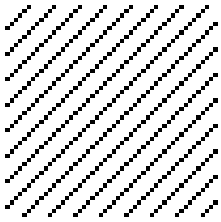
Figure B6. Carte géographique (Carte f)

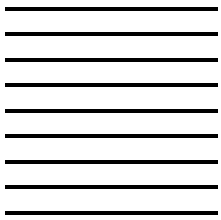

Espagne
Maroc
Algérie
Sahara
Occidebtal
Mauritanie
Mali


Maroc
Algérie
Sahara
Occidental
Mali
Mauritanie









Carte g) :

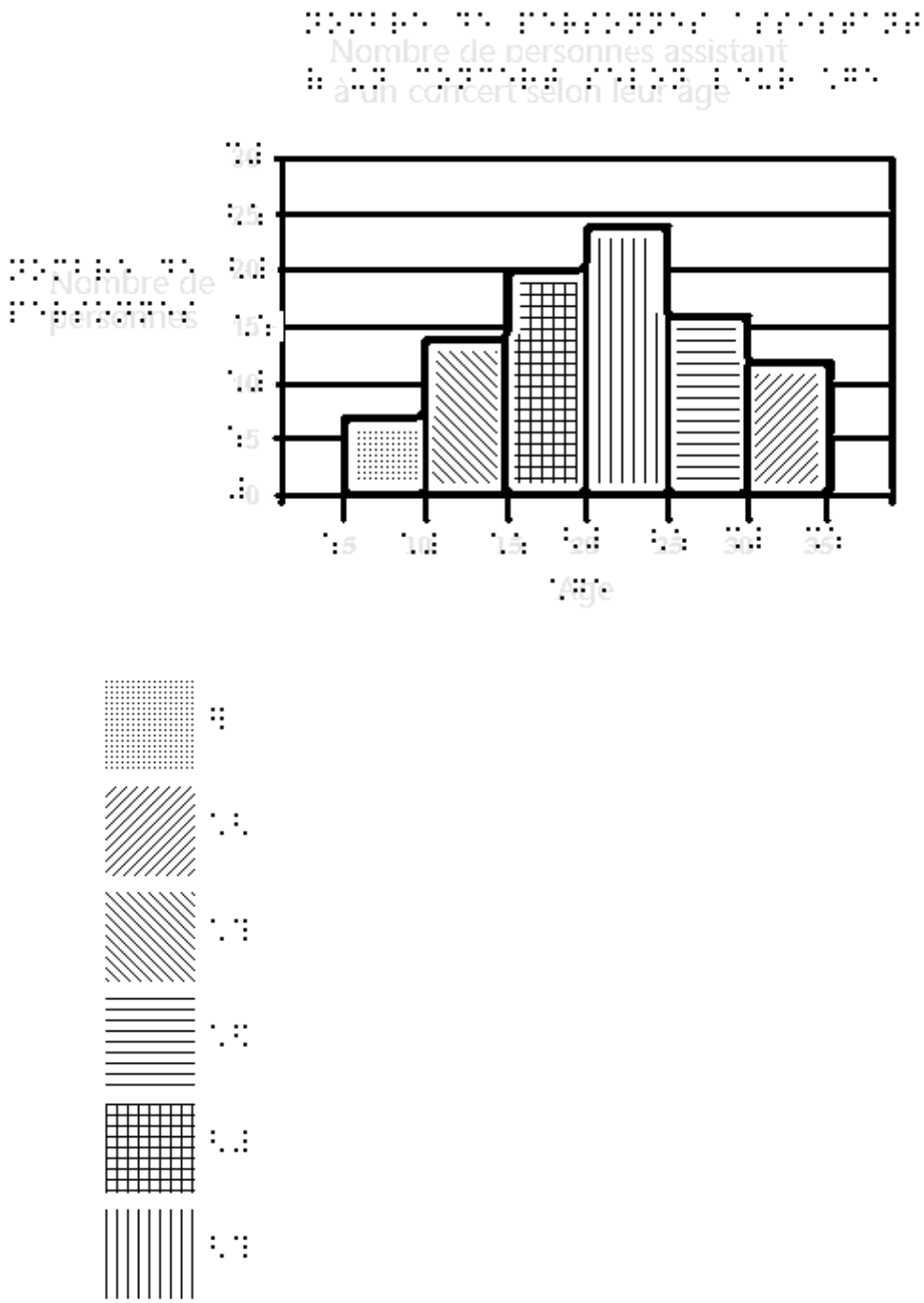
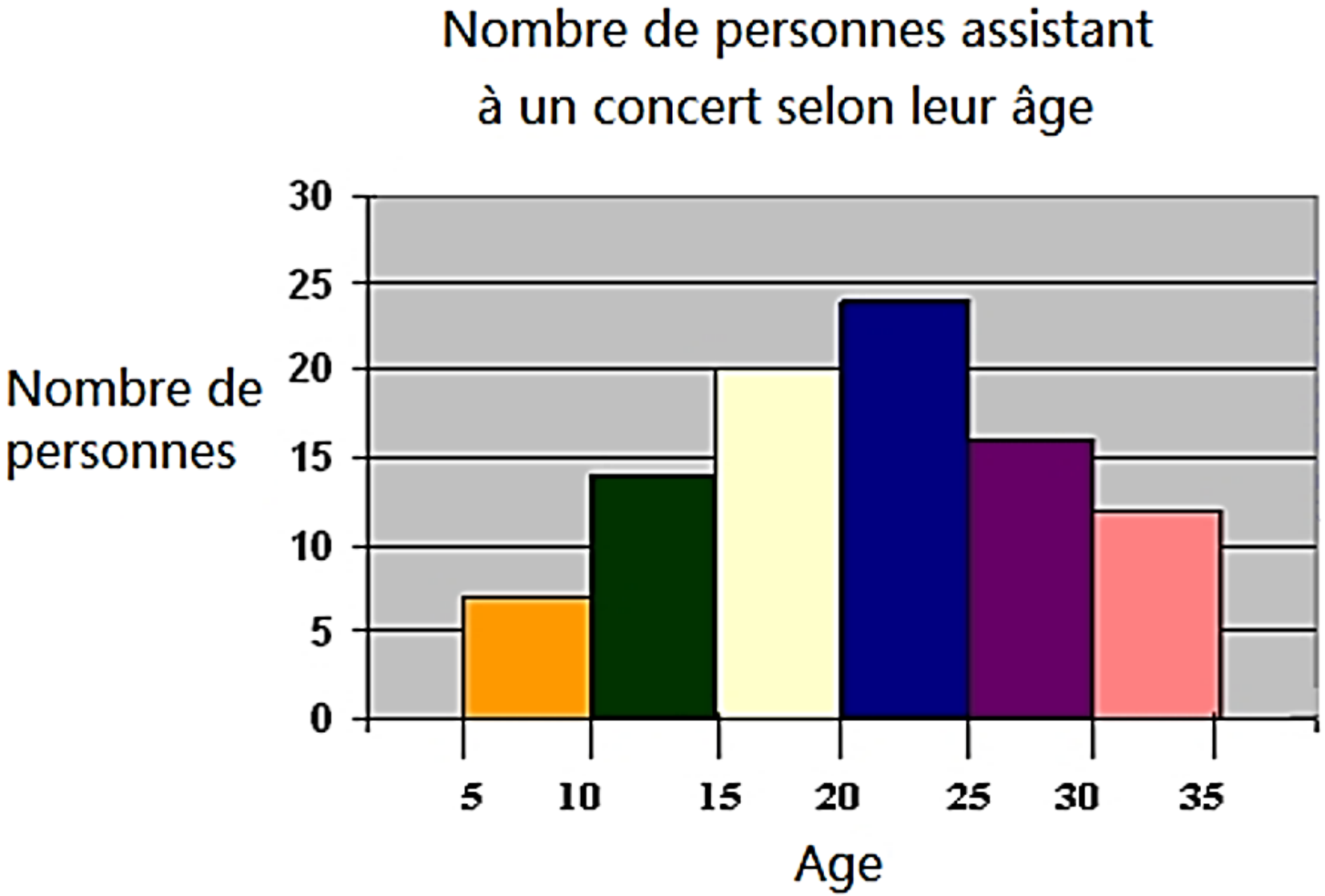
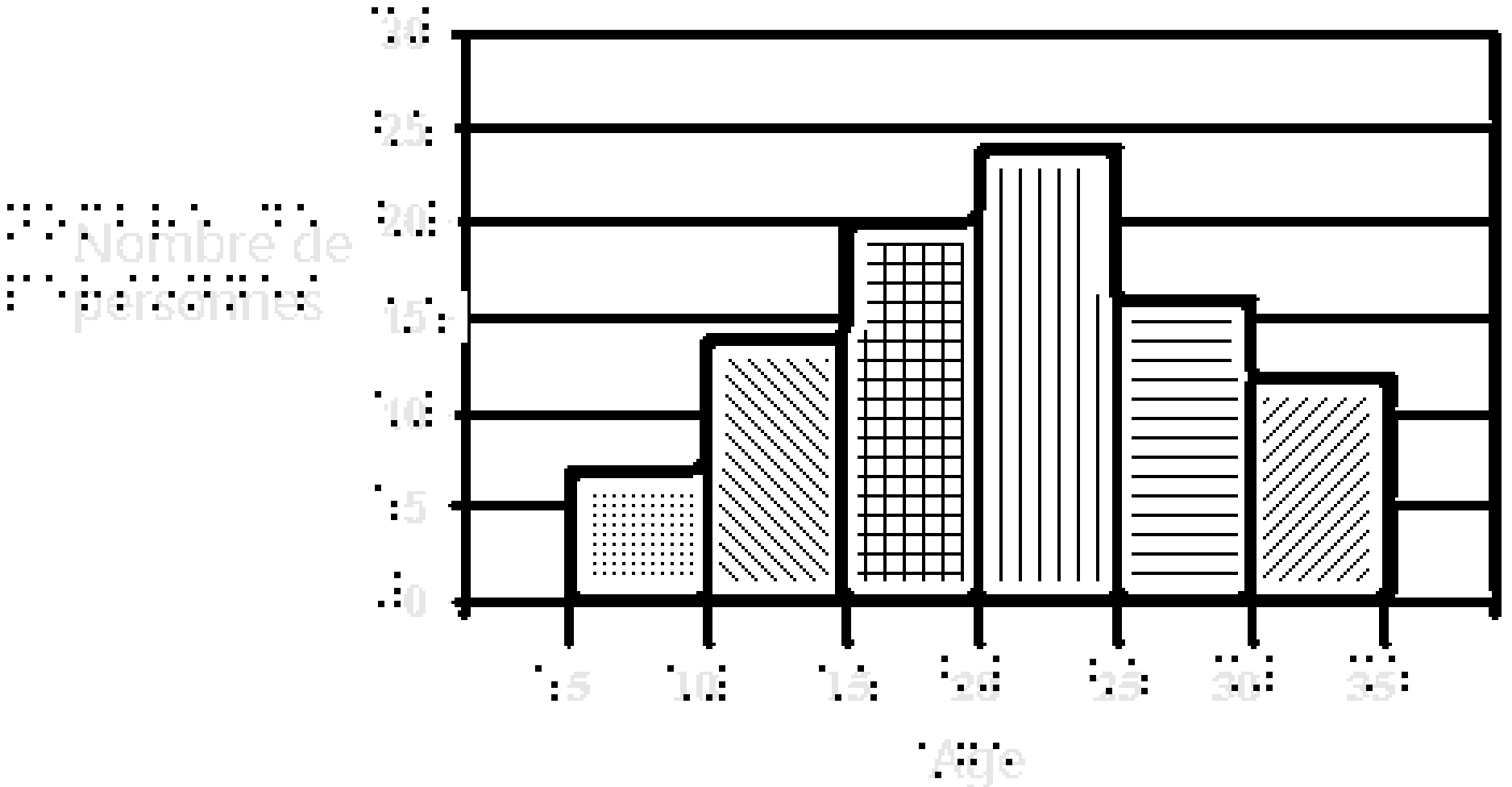
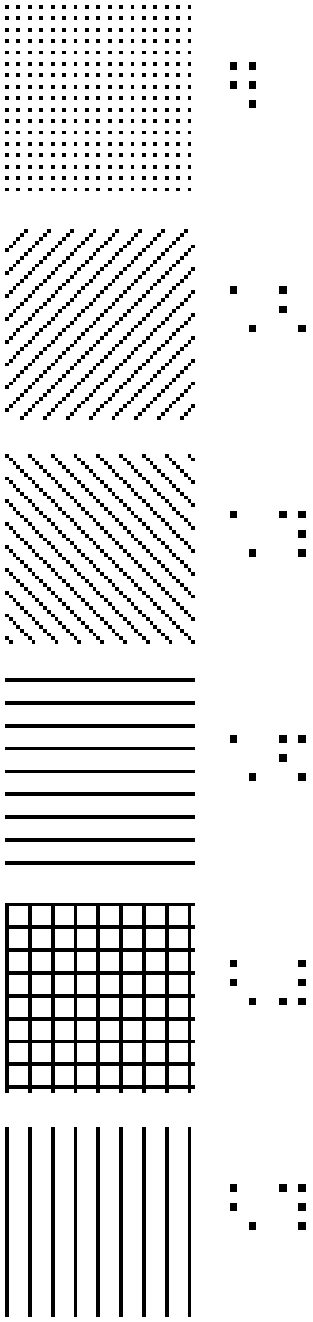


Figure B7. Image histogramme (Carte g)


Nombre de personnes assistant
à un concert selon leur âge
Nombre de personnes
30
25
20
15
10
5
0
5
10
15
20
25
30
35
Age

Nombre de personnes assistant à un concert selon leur âge





Carte h) :

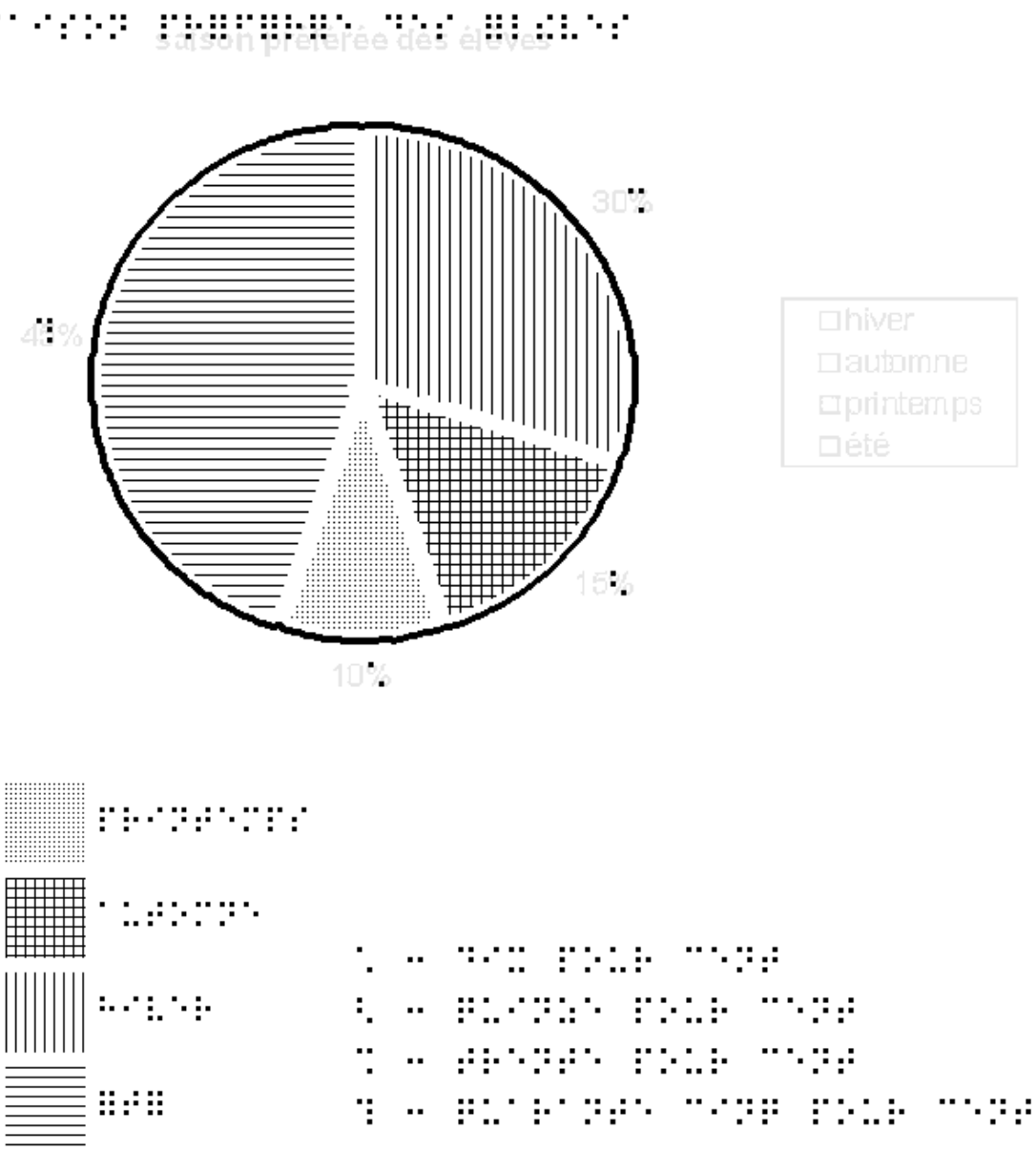
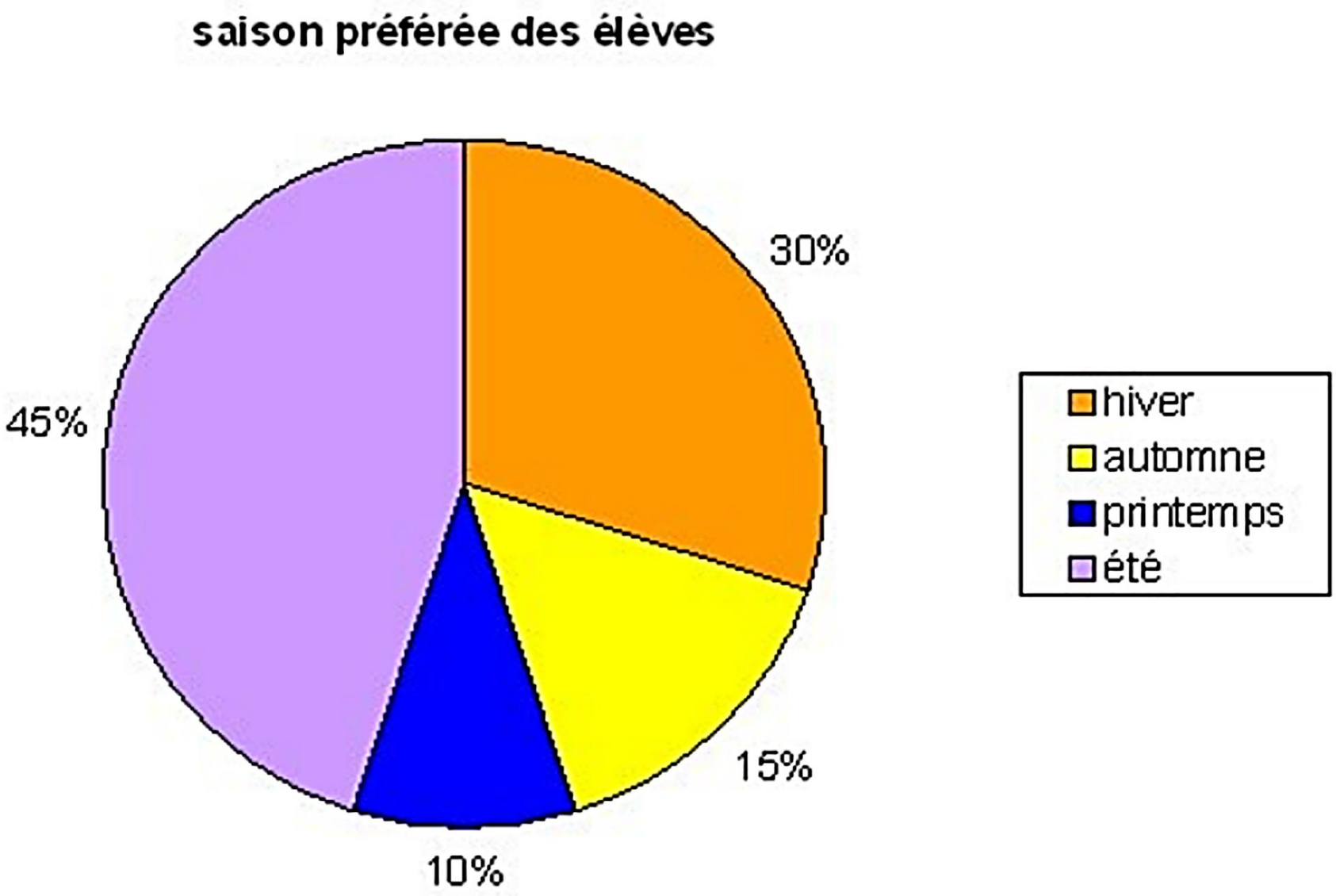
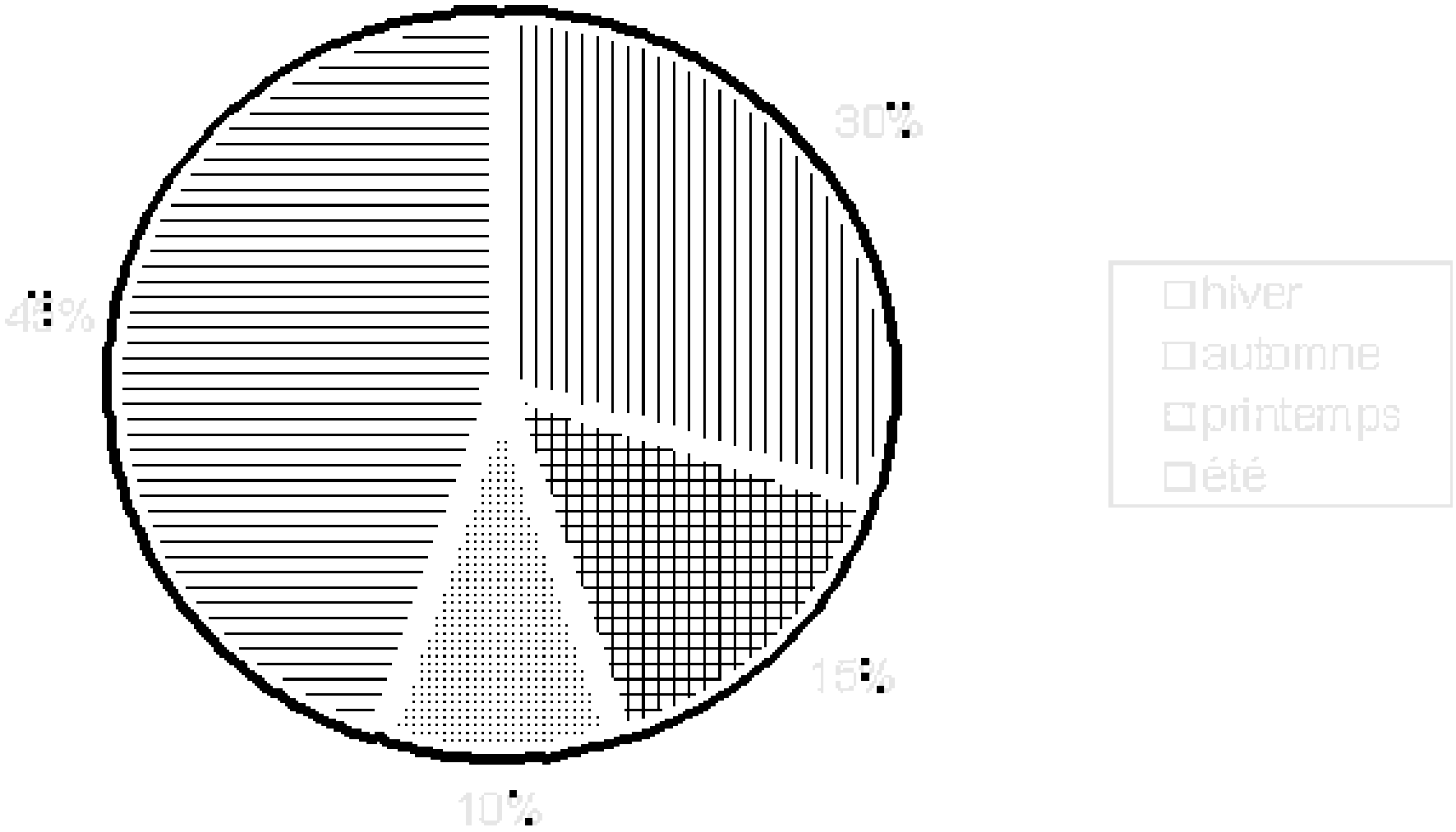
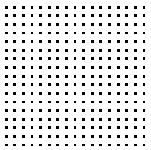


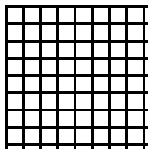
Figure B8. Image camembert (Carte h)

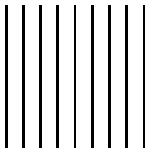

saison préférée des élèves
30%
15%
10%
45%
hiver
automne
printemps
été

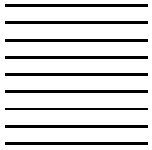
Saison préférée des élèves











Annexe C

Image astronomique réalisée en relief

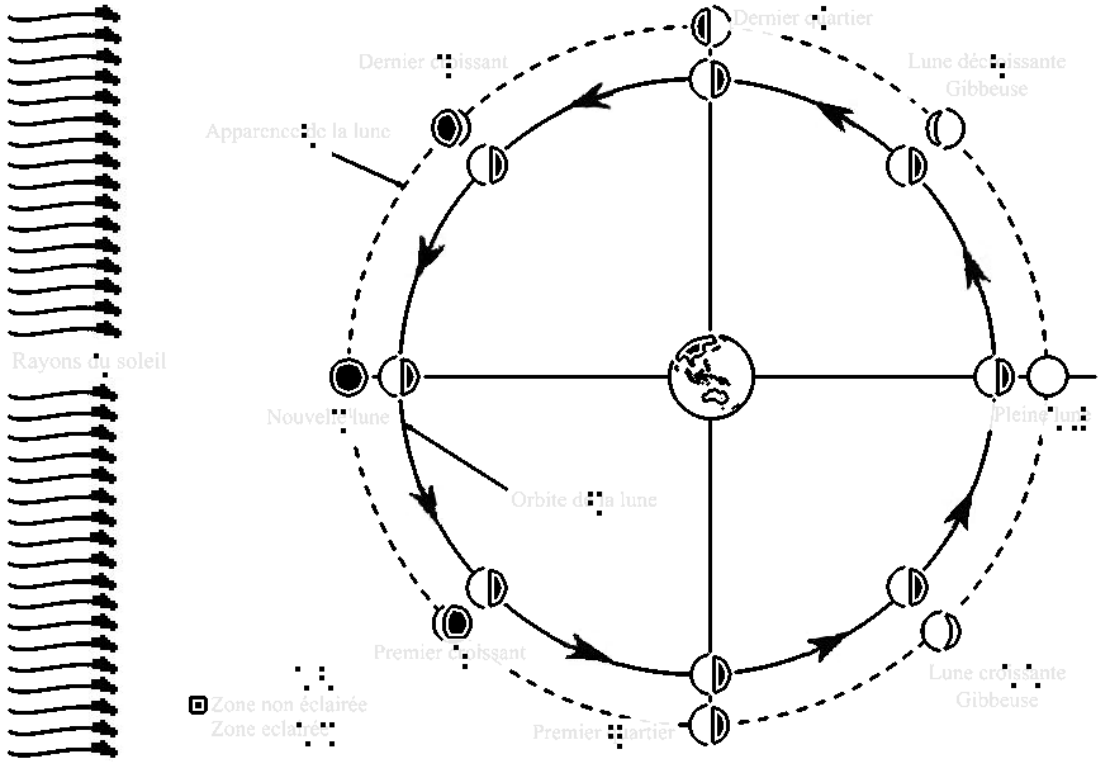
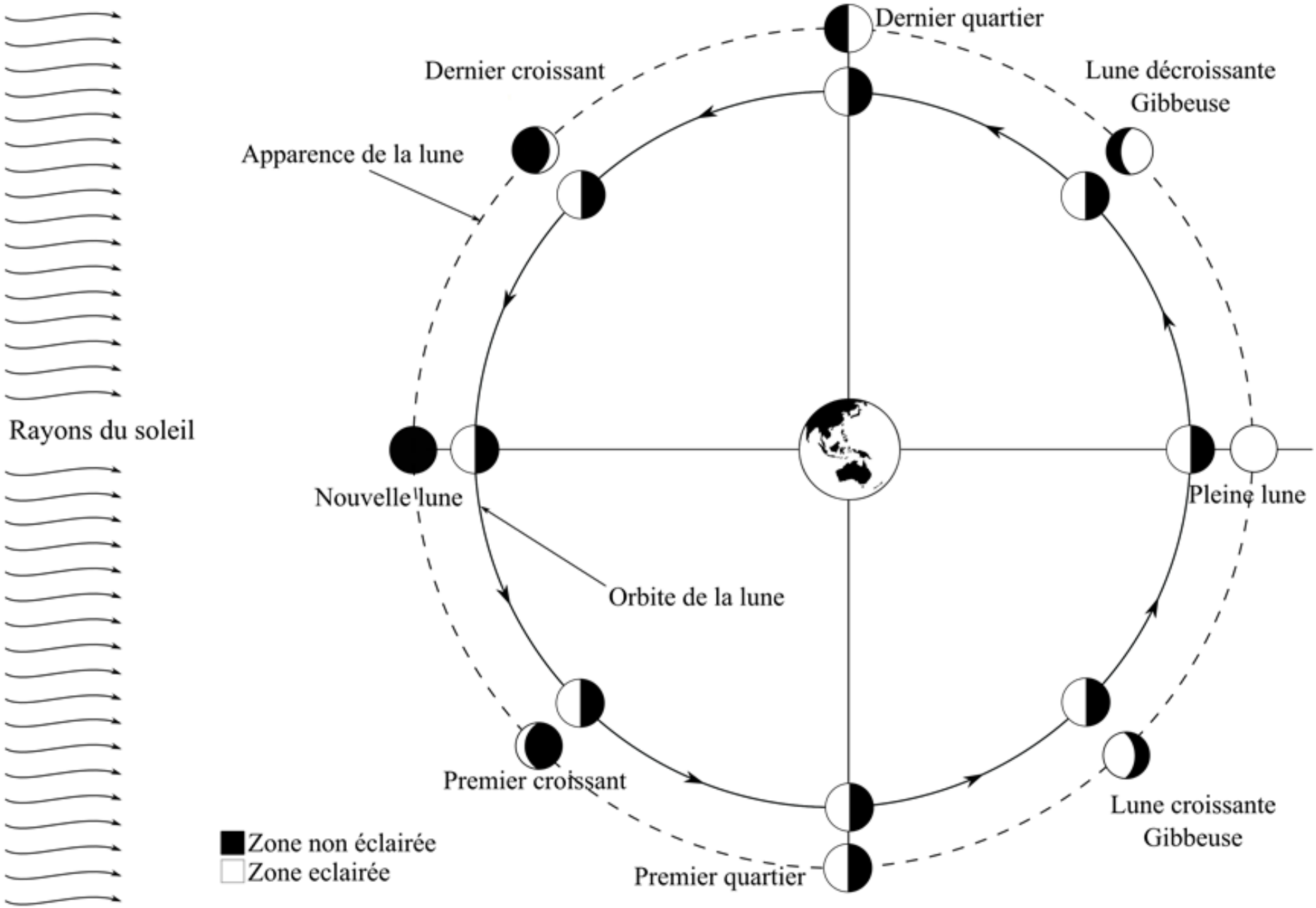
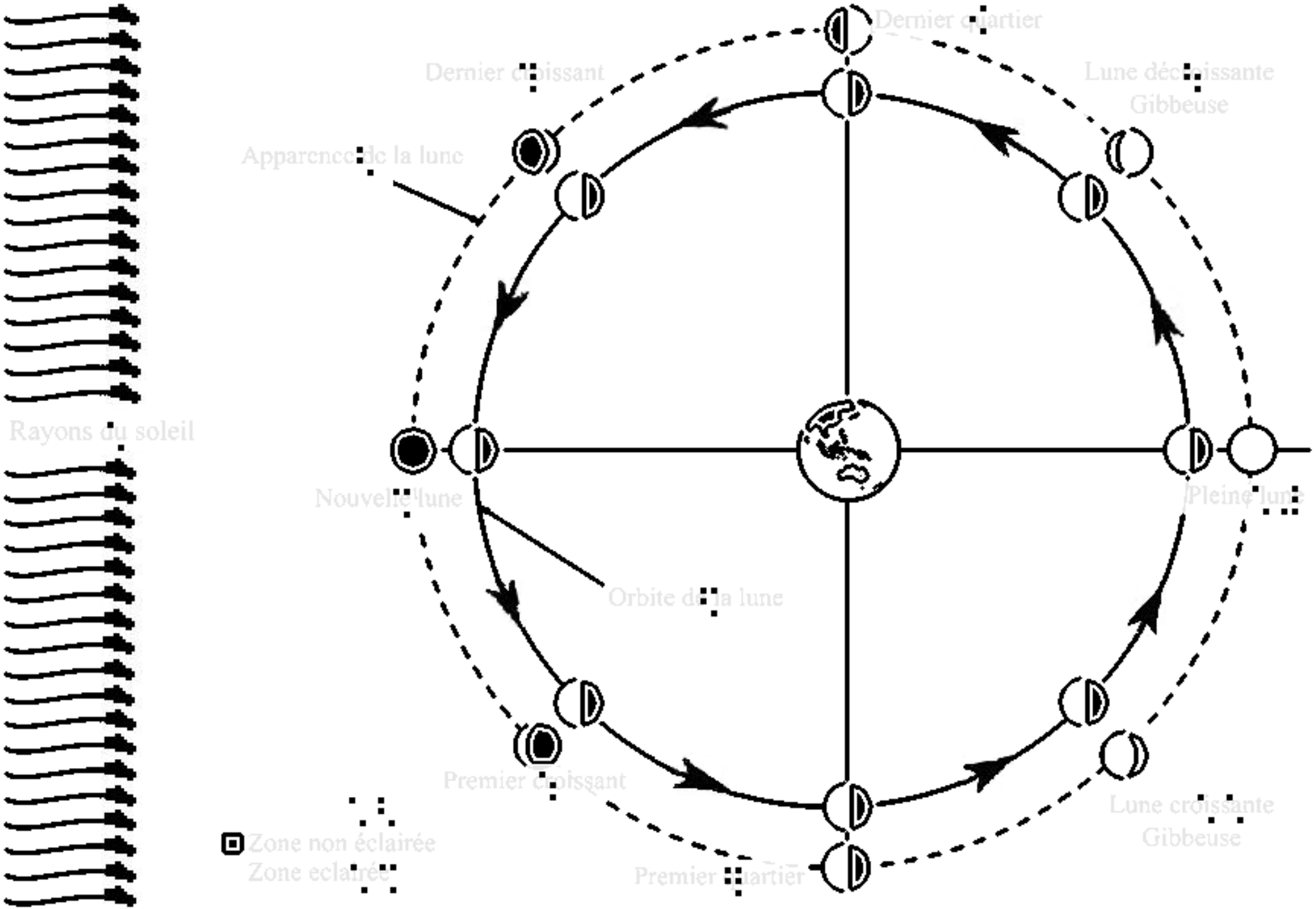


Figure C1. Image astronomique réalisée en relief (phase de la lune)


Rayons du soleil
Dernier quartier
Dernier croissant
Lune décroissante
Gibbeuse
Apparence de la lune
Nouvelle lune
Pleine lune
Orbite de la lune
Premier croissant
Lune croissante
Gibbeuse
Premier quartier
Zone non éclairée
Zone eclairée


Dernier quartier
Dernier croissant
Lune décroissante
Gibbeuse
Apparence de la lune
Rayons du soleil
Nouvelle lune
Pleine lune
Orbite de la lune
Premier croissant
Lune croissante
Gibbeuse
Zone non éclairée
Zone eclairée
Premier quartier

