



Institut National Polytechnique
Félix Houphouët Boigny Côte d'Ivoire

EDP
Ecole Doctorale Polytechnique

L2IS
Laboratoire Instrumentation
Image et Spectroscopie



Collège Doctoral
Lille Nord de France



SESAM Sciences économiques, sociales,
de l'aménagement et du management
École Doctorale n°73



THESE en Cotutelle

Pour obtenir le grade de

Docteur de l'Institut National Polytechnique

Discipline : Electronique et Electricité appliquées

Et de

Docteur de l'Université de Lille

Discipline : Mathématiques et Applications

présentée et soutenue par

Doumun Mékapeu Solange

le 09 juillet 2021

Contribution au développement des techniques d'analyse d'images multispectrales pour l'aide au diagnostic du paludisme

Directeurs de thèse : **Sophie Dabo-Niang et T. Jérémie Zoueu**

JURY :

Mme Ghislaine GAYRAUD	Professeur, Université de Compiègne (France)	Rapporteur
M. Alain TIEDEU	Professeur, Université Yaoundé I (Cameroun)	Rapporteur
Mme Ndeye NIANG KEITA	Maître de Conférence, CNAM Paris (France)	Examineur
M. Kablan Jerome ADOU	Professeur, Université Felix Houphouët Boigny (Côte d'Ivoire)	Examineur
M. Cristian PREDA	Professeur, Université de Lille (France)	Examineur (Président du jury)
M. Ahmadou ALIOUM	Professeur, Université de Bordeaux (France)	Examineur
Mme Sophie DABO-NIANG	Professeur, Université de Lille (France)	Directrice
M. T. Jérémie ZOUEU	Professeur, Institut National Polytechnique Houphouët Boigny (Côte d'Ivoire)	Directeur

Remerciements

Je tiens avant tout propos à remercier mes encadrants Sophie Dabo-Niang et Jérémie T. Zoueu pour m'avoir guidée tout le long de ce travail, leurs précieux conseils, leurs connaissances scientifiques et pédagogiques m'ont été d'un grand appui tant personnel que scientifique. La patience et la rigueur sont des vertus qu'ils m'ont apprises. Je voudrais leur témoigner toute ma gratitude pour avoir cru en mon travail, au-delà des défis qui se présentaient à nous. Merci de m'avoir encouragée à prendre ce chemin de la thèse et de m'y avoir guidée et laissée l'opportunité de développer mes idées. Je ne les remercierai jamais assez pour leur implication à tous les niveaux pour rendre possible ce travail. Merci Professeur Zoueu ! Vous m'avez appris et montré que « *tout est possible à celui qui croît* ». Merci d'avoir cru en cette parole de « *Esaië 42 :3 : Il ne brisera pas le roseau cassé, et n'éteindra point la mèche qui brûle encore !* ».

Merci Professeur Dabo-Niang ! Votre détermination à aller jusqu'au bout malgré de nombreux obstacles a été pour moi un catalyseur. Ce n'était pas évident, mais vous avez tout donné pour rendre cela possible. Merci encore pour votre disponibilité, votre écoute. Vous m'avez présenté le monde de la géostatistique et de la statistique fonctionnelle, merci pour m'avoir amenée à travailler dans ce domaine.

Je voudrais exprimer ma reconnaissance au directeur de L'INP-HB M. Koffi N'Guessan et au Directeur de l'Ecole Doctorale Polytechniques de l'INP-HB pour l'opportunité qui m'a été accordée de présenter une thèse de doctorat dans cette prestigieuse école. Mes remerciements vont à l'endroit des Professeurs Souleymane Oumtanaga et Michel Babli pour le cadre de travail, les précieux conseils, le soutien qu'ils ont été pour nous dans le laboratoire LARIT, partenaire de L2IS à Abidjan.

Je tiens à exprimer ma gratitude au Président de l'Université de Lille, et au Responsable de la bourse MOBILLEX pour cette opportunité rêvée qu'ils m'ont donnée à travers cette cotutelle et pour le financement de la première phase des échanges. Mes sincères remerciements vont à l'endroit de la Bourse EMS Simons pour leur financement qui a

rendu possible mon deuxième séjour dans la cadre des échanges au cours de cette cotutelle. Je tiens à remercier le Directeur du laboratoire LEM et le CNRS pour leurs précieux financements surtout au cours de cette période difficile du COVID. Merci à Mme Nathalie HAY et Mme Annette Legrand pour leur disponibilité, leur indulgence et d'avoir facilité la mise en place des financements. Un grand merci à l'Agence Suédoise Sida (the Swedish International Development Cooperation Agency) pour son soutien financier à travers le programme ISP (the International Science Programme, Uppsala University). Ma reconnaissance va en l'endroit de Mme Carla Puglia, Professeur à l'Université d'Uppsala en Suède, pour son implication personnelle afin de rendre possible ce financement. Merci au Directeur du Laboratoire Paul Painlevé de l'Université de Lille qui nous a accueillie pendant cette dernière année.

J'adresse tous mes remerciements à Mme Ghislaine Gayraud, Professeur à l'Université de Compiègne, ainsi qu'à M. Alain Tiedeu, Professeur à l'Université Yaoundé I, de l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse. J'exprime ma gratitude à Mme Ndeye Niang Keita, Maître de conférences au CNAM Paris, M. Kablan Jerome Adou, Professeur à l'Université Felix Houphouët Boigny, M. Cristian Preda, Professeur à l'Université de Lille et à M. Ahmadou Alioum, Professeur à l'Université de Bordeaux, qui ont bien voulu être examinateurs. Vous m'honorez de votre présence en tant que juge de cette thèse.

Ces remerciements seraient incomplets si je n'en adressais pas à tous les thésards et les autres membres des laboratoires du LARIT, du L2IS et de l'Axe MéQAME du laboratoire LEM, notamment ceux avec qui j'ai eu l'occasion de travailler et les autres simplement pour les bons moments partagés.

Je tiens à remercier chaleureusement ma famille qui m'a toujours soutenue. Merci d'être à mes côtés même de loin, pour leur compréhension et leur encouragement durant ma thèse. A Fahariat Boukari, Alaa Ali Hassan, Nelly Berille et à tous les membres de l'église de Triolo, merci de m'avoir accueillie comme votre sœur.

Enfin, je remercie mon cher époux et nos enfants pour leur soutien quotidien et leur enthousiasme contagieux à l'égard de mes travaux. Le travail acharné de chaque jour loin de vous me rapprochait de la chaleur familiale. Merci !

DEDICACE

*Je dédie ce travail au DIEU tout PUISSANT qui rend toute chose possible !
A LUI soit la Gloire pour l'éternité ! Amen !*

Table des matières

Remerciements	iii
Table des matières	ix
Liste des figures	xiii
Introduction	ii
1.1 Contexte	19
1.1.1 Le paludisme	19
1.1.2 Enjeux sanitaires	21
1.1.3 Enjeux socio-économiques	22
1.1.4 La lutte antipaludique.....	23
1.2 Objectif de la thèse	26
1.3 Organisation de la thèse	27
Chapitre 2 Diagnostic Automatique du Paludisme	29
2.1 Introduction	31
2.2 Problématique	32
2.3 Moyens.....	33
2.3.1 Mobilisation de l’opinion internationale à réalité du paludisme	33
2.3.2 Mobilisation sur le terrain	33
2.4 Enjeux socio- économiques[17], [18]	34
2.5 Les techniques de diagnostic automatiques du paludisme	35
2.5.1 La technique d’immunochromatographie (ICG): [19] [20] [21] [17] [14]	36
2.5.2 QBC (Quantitative Buffy Coat) :[24].....	37
2.5.3 Diagnostic Moléculaire : [24].....	37
2.5.4 Travaux connexes.....	38

2.5.4.1	Le microscope multispectral multimodal pour diagnostic du paludisme.	38
2.5.4.2	Systèmes basés sur téléphone mobile[25] [26] :.....	39
2.5.4.3	Diagnostic par force magnétophorétique[22], [27] :.....	39
2.6	Conclusion.....	40
Chapitre 3 Acquisition d'image : le Microscope Multimodal et Multispectrale		42
3.1	Introduction.....	44
3.2	Le microscope [31]-[35]:.....	45
3.2.1	Partie optique :	46
3.2.2	Partie électronique :	47
3.2.3	Partie logicielle[1] [32] [35]:.....	48
3.2.4	Les modalités [52][36] :.....	48
3.2.4.1	Transmission / Absorption[31] :.....	49
3.2.4.2	La diffusion de la lumière[37] [52] :.....	50
3.2.4.3	Réflexion.....	53
3.2.5	Les longueurs d'onde et leur intérêt :.....	55
3.3	Conclusion :.....	58
Chapitre 4 A Novel Multispectral Images Normalization Application to Unstained Malaria Infected Blood Smear		60
Chapitre 5 Detection and Segmentation of Erythrocytes in Multispectral Label-Free Blood Smear Images for Automatic Cell Counting.....		93
Chapitre 6 Unsupervised Classification of Erythrocytes in Multispectral Unstained Blood Smear Images for Malaria Automatic Diagnosis.....		120
Chapitre 7 Conclusion et Perspectives.....		142
7.1	Conclusion générale	144
7.2	Discussion.....	146

7.3 Quelques perspectives :.....	147
Références bibliographiques.....	149
Résumé.....	156

Liste des figures

<i>Figure 1: Cycle de developpement du plasmodium</i>	<i>20</i>
<i>Figure 2: Carte de la mortalite due au paludisme</i>	<i>22</i>
<i>Figure 3: Evolution de la lutte contre le paludisme entre 2000 et 2018[6]</i>	<i>23</i>
<i>Figure 4: Microscope multispectral et multimode [35].....</i>	<i>45</i>
<i>Figure 5: Diagramme des lobes angulaires et aperçu d'une section transversale verticale du microscope multimode multispectrale (merdasa et al., 2013)[35].....</i>	<i>47</i>
<i>Figure 6 : Indicatrices de la diffusion des particules spheriques dans le plan orthogonal a la polarisation (lineaire) de la lumiere incidente.</i>	<i>52</i>
<i>Figure 7 : Representation schematique des phenomenes de reflexion speculaire (a gauche) et de reflexion diffuse (a droite).....</i>	<i>54</i>
<i>Figure 8: Schema de reproduction du parasite dans le foie.....</i>	<i>55</i>
<i>Figure 9: Caracterisation du globule rouge infectee par le plasmodium</i>	<i>56</i>

Glossaire

ADN	Acide DésoxyriboNucléique
AIM	Action and Investment to defeat Malaria 2016-2030 (AIM) - for a malaria-free world (Action et Investissement pour vaincre le paludisme 2016-2030)
ARN	Acide RiboNucléique
DDT	Dichlorodiphényltrichloroéthane
DEL	Diode électroluminescente ou LED (light-emitting diode)
EDP	Ecole Doctorale Polytechnique
GE	Gouttes Epaisses
HRP2	Histidine-Rich Proteine-2
INP-HB	Institut National Polytechnique Félix Houphouët-Boigny
L2IS	Laboratoire d'Instrumentation, Image et Spectroscopie
LEM	Lille Economie Management
MeQuAME	Méthodes Quantitatives Appliquées au Management et à l'Economie
OMS	Organisation Mondiale de la Santé
ONU	Organisation des Nations Unies
P. Falciparum	Plasmodium falciparum
P. Vivax	Plasmodium vivax
PCR	Polymerase Chain Reaction ou Réaction de Polymérisation en Chaîne
PECADOM	Prise En Charge du Paludisme à Domicile
PIB	Produit Intérieur Brut
PNUD	Programme des Nations Unies pour le Développement
QBC	Quantitative Buffy-Coat
SESAM	Sciences Economiques, Sociales, de l'Aménagement et du Management
TDR	Tests de Diagnostic Rapide
UNICEF	United Nations International Children's Emergency Fund (Fonds d'urgence international des Nations unies pour l'enfance) ou Fonds des Nations unies pour l'enfance
UV	Ultra Violet

Introduction

Longtemps considéré comme la méthode de référence de diagnostic du paludisme, l'examen par microscopie du frottis sanguin et de la goutte épaisse présente cependant des limites face à l'urgence. En effet, sa fiabilité et sa sensibilité dépendent principalement du microscopiste dont le manque d'expérience ou sous l'effet de la fatigue, peut conduire à un faux diagnostic ; car cet examen s'est longtemps limité à une simple analyse visuelle de l'échantillon et le comptage qui s'en suit est manuel.

La précision, la rapidité et la puissance de traitement d'une image, maintenant disponible sur un simple ordinateur à travers l'analyse de ses composantes spectrales, donne un véritable essor au traitement de haute résolution de ces images médicales en vue d'un diagnostic automatique.

C'est dans le but de contribuer à l'aide au diagnostic précoce de cette pathologie que le thème suivant a été soumis à notre étude : « CONTRIBUTION AU DÉVELOPPEMENT DES TECHNIQUES D'ANALYSE D'IMAGES MULTISPECTRALES POUR L'AIDE AU DIAGNOSTIC AUTOMATIQUE ET RAPIDE DU PALUDISME ».

Ce sujet rentre dans le cadre d'un projet de recherche qui se poursuit au sein du Laboratoire d'Instrumentation, Image et Spectroscopie (malaria project). Il a commencé par la conception d'un système optique qui est le microscope multimodal multispectral. La contribution de ce mémoire vient en appui pour l'aboutissement de ce projet.

Pour une meilleure compréhension de notre stratégie nous présenterons ici le contexte général de la thèse, suivi de l'objectif et terminerons par l'organisation de notre manuscrit.

1.1 CONTEXTE :

1.1.1 Le paludisme :

Le paludisme est une maladie infectieuse qui touche plus de la moitié de la population mondiale. Il est transmis à l'homme par un parasite de genre plasmodium. Il existe 5 espèces de plasmodium humain dont les formes mortelles sont le *P. Vivax* et le *P. Falciparum*.

Le *P. falciparum* est la forme la plus dangereuse et la plus répandue. Cette forme est responsable de plus 99% des décès dus au paludisme.

Ce parasite est transmis à l'homme par le biais d'un moustique (anophèle femelle) lors de son repas sanguin. Dans le corps humain, le parasite se développe dans les cellules du foie, puis circule dans l'organisme, en entraînant une hémolyse intravasculaire des globules rouges [1]. Il est à nouveau transmis à l'anophèle lors d'une nouvelle pique (Figure 1). Le temps d'incubation est de 7 à 15 jours au bout desquels les premiers symptômes similaires à ceux d'une grippe apparaissent (Courbature, Douleur, Fièvre, Maux de tête, Nausée, Vomissement, etc.).

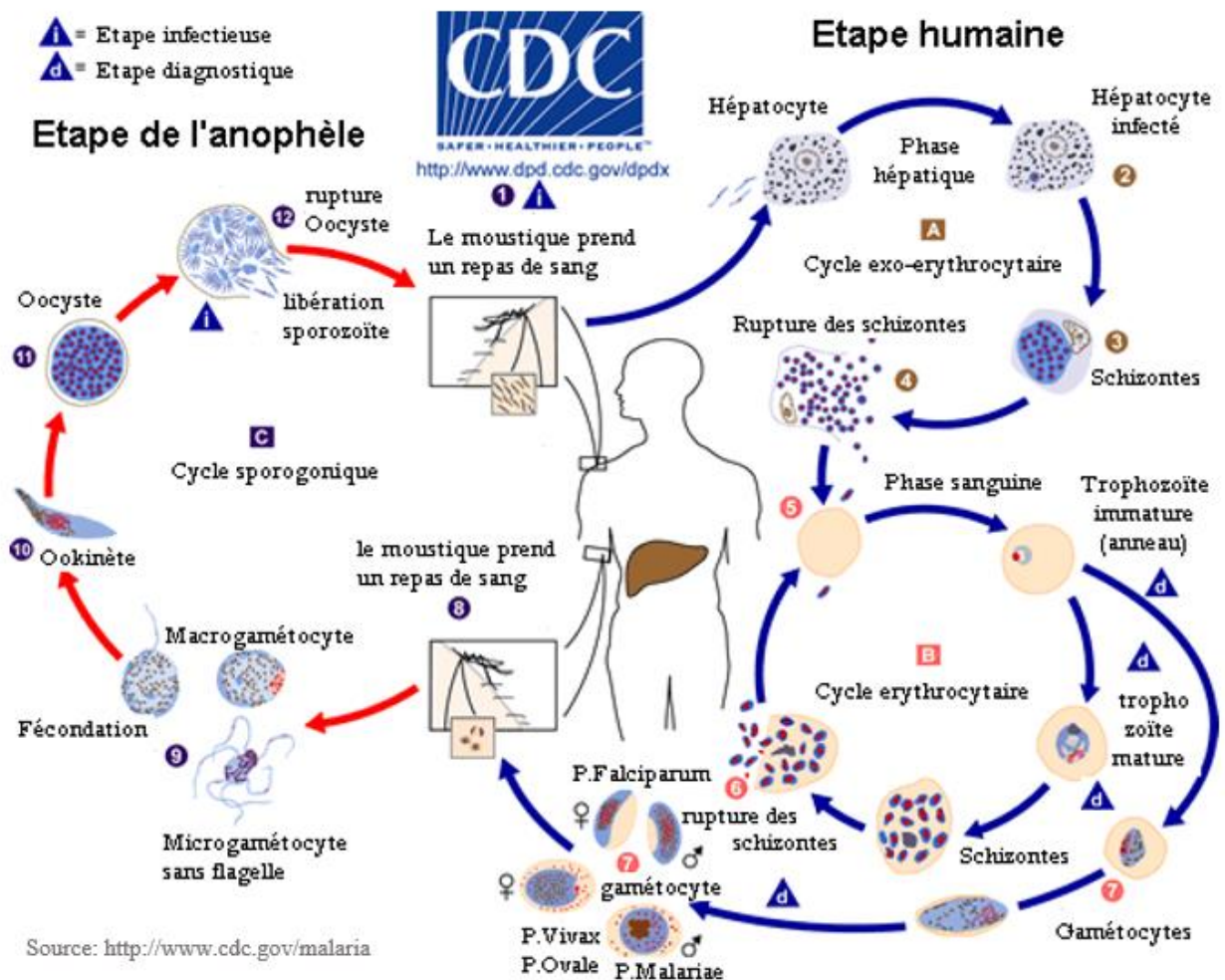


Figure 1: Cycle de développement du plasmodium

Le parasite se reproduisant rapidement, les cas peuvent devenir graves en quelques jours voire quelques heures. Dans le cas d'une contagion au *P. falciparum*, la maladie atteint les formes les plus sévères et souvent mortelle sous 24 heures, sans un traitement. Cette forme de plasmodium peut parfois atteindre les organes vitaux, à la suite d'un retard de prise en charge. Ceci entraîne des complications dont le neuro-paludisme qui est mortel en 2 ou 3 jours ou conduit à des séquelles irréversibles. Les complications se manifestent par une défaillance neurologique, cardiocirculatoire ou respiratoire, des convulsions répétées, une insuffisance rénale, une anémie sévère etc. L'observation de l'un de ces signes est un motif de prise en charge en soins intensif ou en réanimation. Toutefois, avec une prise en charge adéquate, le paludisme se guéri en 3 jours de traitement pour les formes les plus simples. Le pronostic repose essentiellement sur la rapidité du diagnostic. Le retard de celui-ci est l'une des principales causes de décès.

1.1.2 Enjeux sanitaires :

Le paludisme (Malaria en Anglais) est responsable chaque année de près d'un million de décès [2] (OMS, 2020) et l'une des premières causes de consultation dans les centres de santé des pays en voie de développement. En 2019, l'on a enregistré 229 millions de cas et 409 000 de décès imputables au paludisme. Les populations, les plus vulnérables sont les femmes enceintes et enfants de moins de 5 ans ; 67 % de décès concernent les enfants de moins de 5 ans.

L'Afrique subsaharienne est la région qui paye le plus lourd tribut à cette pandémie. Elle est la zone de prédilection du plasmodium falciparum. D'ailleurs 99,7% des cas estimés dans cette zone sont du fait de ce type de plasmodium. De ce fait, elle a enregistré en 2019, 94% des cas et de décès dus à cette maladie majoritairement dans moins de 20 pays [3], [4] (Figures 2 - 3).

Il convient d'indiquer qu'au-delà du problème social et de la santé publique qu'il pose, le paludisme est également un problème économique.

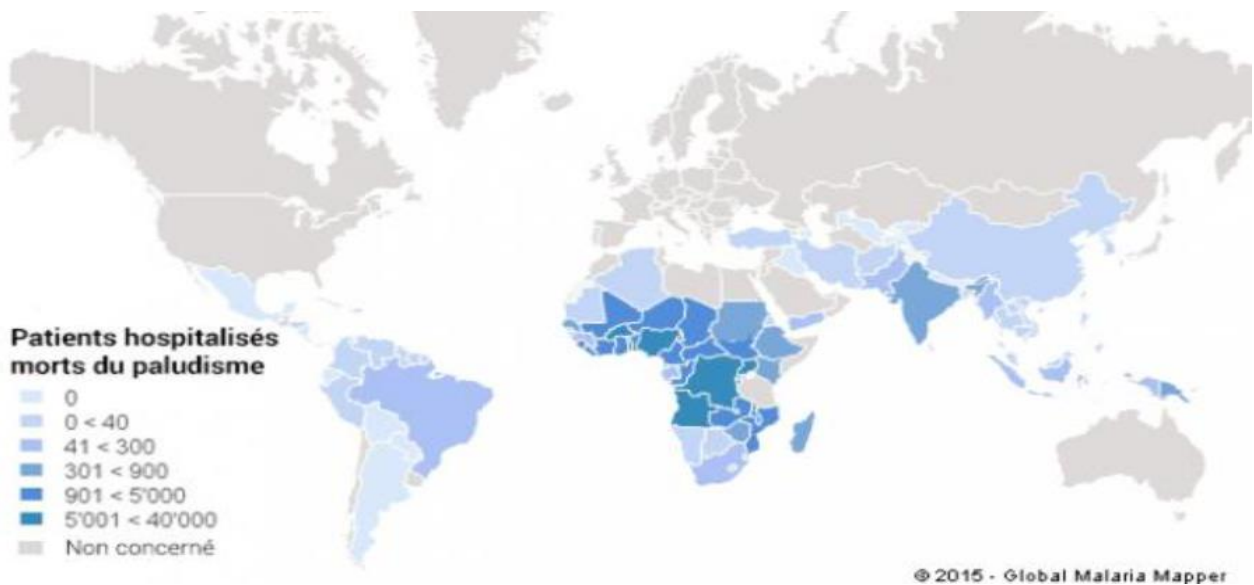


Figure 2: Carte de la mortalité due au paludisme

1.1.3 Enjeux socio-économiques :

La morbidité et l'impact du paludisme en milieux scolaire et professionnel est élevé [5]. Cette forte prévalence constitue une entrave au développement durable des pays endémiques déjà marqués pour la plupart par des crises économiques et politico-sociales qui mettent à mal leur développement. Il est donc impératif de mettre en place une gestion endémique qui dépasse le seul cadre de la santé publique. Selon l'OMS, l'Afrique perd chaque année environ 12 milliards de dollars en PIB du fait du paludisme. En plus de ces coûts indirects, la lutte contre le paludisme engendre un coût direct estimé à plus 5 milliards de dollars par an. En 2018, 2.7 milliards de dollars ont été investis dans la lutte contre le paludisme contre 3.2 milliards de dollars en 2017. Il y a donc une urgence en termes d'investissement pour vaincre cette maladie. Pour garantir une efficacité dans sa lutte, l'OMS s'est dotée de « Stratégie technique mondiale contre le paludisme 2016-2030 ». Cette dernière est complétée par « Action et Investissement pour vaincre le paludisme 2016-2030 (AIM) - pour un monde sans paludisme » de partenariat Roll Back Malaria. Tous deux ayant pour objectif de servir de guide et de sonnette d'alarme dans cette lutte contre le paludisme.

1.1.4 La lutte antipaludique :

Lancé il y a plus de 5 décennies, le programme d'éradication du paludisme, a enregistré bien de succès. Regrouper au sein d'un partenariat nommé **ROLL BACK MALARIA** créé en 1998 [5], avec l'ONU, l'UNICEF, le PNUD, la Banque Mondiale et des partenaires privés, l'OMS est parvenue à éliminer le paludisme en Europe, en Amérique du nord et à réduire la charge de morbidité et de mortalité dans les pays endémiques. Cependant, cet ambitieux objectif « d'un monde sans paludisme » n'est toujours pas atteint. En effet, bien qu'ayant à son actif de réels progrès et des exploits du point de vue de la recherche dans la prise en charge de cette maladie ; le programme s'est heurté à la résistance de l'anophèle, agent vecteur de la maladie, aux insecticides (DDT : dichlorodiphényltrichloroéthane), conçus pour l'éliminer, et celle des parasites aux molécules antipaludiques. La propagation de ces résistances met à mal le succès tant escompté. De plus, un vaccin efficace contre les formes sévères de la maladie et facile d'utilisation reste inaccessible. Ainsi, le paludisme continue d'endeuiller de milliers de famille et d'avoir des conséquences dévastatrices sur les moyens de subsistance et la santé des populations, quoique parfaitement évitable et traitable.

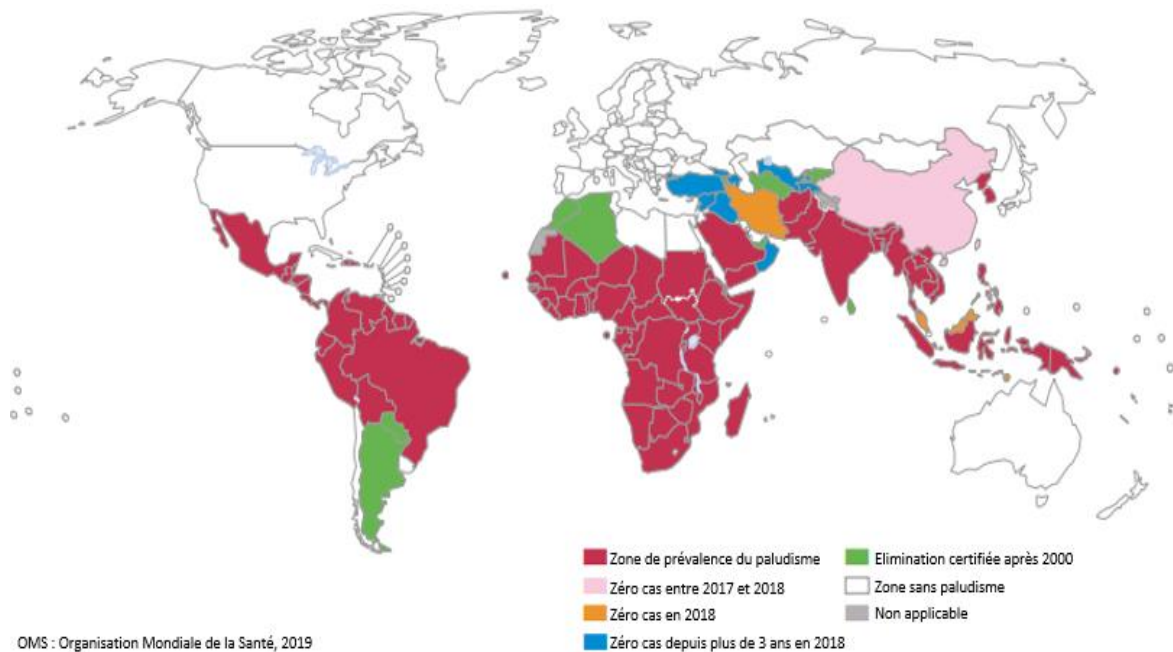


Figure 3: Evolution de la lutte contre le paludisme entre 2000 et 2018 [6]

Une prise en charge rapide du paludisme est salutaire. Il permet de prévenir l'aggravation des cas et empêcher les décès associés à la maladie. De ce fait, une surveillance efficace par un diagnostic exact est cruciale. Cependant, ces décennies de recherche n'ont pu remplacer la microscopie optique qui demeure la méthode de référence en matière de diagnostic du paludisme, par la goutte épaisse et le frottis sanguin.

La recherche scientifique, en ce qui concerne le diagnostic du paludisme, est un domaine très actif et plusieurs solutions ont vu le jour notamment le TDR, le PCR, le QBC. Mais des limites non négligeables n'ont jusque-là pas permis leur déploiement effectif dans les centres de santé des pays endémiques.

Le TDR, bien qu'étant accessible et approuvé par l'OMS, doit être confirmé par la microscopie. D'autant plus qu'il ne permet pas d'apprécier la parasitémie et d'identifier l'espèce de parasite en cause. Sans compter qu'il est négatif pour les parasitémies faibles et positif en présence d'auto-anticorps jusqu'à 2 à 3 semaines après un traitement efficace [7].

Le QBC (quantitative Buffy-Coat) est une technique de microscopie par fluorescence très sensible. Cependant, il nécessite du matériel spécifique non accessible en zone endémique outre le fait qu'il ne permet pas d'identifier les espèces.

Le PCR (Polymerase Chain Reaction ou Réaction de Polymérisation en Chaîne) [8], qui est une technique de biologie moléculaire, permet de détecter et d'identifier les espèces du parasite du paludisme. Il nécessite cependant une technique de pointe non accessible en zone endémique.

En définitive, la microscopie reste la méthode indéniable en ce qui concerne le diagnostic du paludisme quoiqu'elle soit limitée. En effet, le processus, entièrement manuel, peut être source d'erreur tant au niveau de la confection et du marquage de l'échantillon que de l'analyse microscopique qui s'en suit. Le résultat est subjectif, dépend du niveau de parasitémie (négatif pour les parasitémies faibles) et surtout sujet à une lenteur qui rend difficile la prise en charge des cas urgents. Par conséquent, la mise en place d'une méthode de diagnostic automatique et rapide est d'une urgence capitale dans la lutte contre cette tueuse infantile des pays pauvres.

C'est dans un tel contexte que, que le Laboratoire d'Instrumentation, Image et Spectroscopie (L2IS), de l'Institut National Polytechnique Félix Houphouët-Boigny (INP-HB) de Yamoussoukro en Côte d'Ivoire et des partenaires suédois ont mis sur pied ces dernières années un microscope multispectrale et multimodale, ayant entre autres applications, le diagnostic du paludisme. Basée sur l'imagerie multispectrale, cet instrument est capable d'effectuer un marquage spectral qui permet de détecter les globules rouges parasités et d'établir ainsi le diagnostic automatique du paludisme [6-7]. De plus, il a pour avantage d'automatiser le processus de comptage des érythrocytes et de mesurer la parasitémie qui détermine la conduite thérapeutique à tenir. Cet outil, bien qu'étant une prouesse du point de vue de la recherche scientifique nécessite des ajustements pour atteindre son objectif de diagnostic automatique et rapide du paludisme, aussi bien au niveau du temps d'acquisition que des traitements qui s'en suivent. Il s'avère donc nécessaire d'optimiser les techniques de traitement et d'analyse des images multispectrales pour parvenir au diagnostic automatique et rapide du paludisme.

L'imagerie étant un domaine évolutif avec des composantes interdisciplinaires, elle offre l'avantage d'associer les spécialités telles que l'optique, la spectroscopie et l'analyse statistique. Cette thèse est le fruit d'une opportune collaboration entre L2IS spécialisé dans l'instrumentation optique, électronique et dans l'imagerie, et le LEM en particulier l'axe MeQuAME (Méthodes Quantitatives Appliquées au Management et à l'Economie). MeQuAME a développé une longue expertise dans le domaine de l'analyse statistique spatiale, du Big Data et Machine Learning. Cette collaboration s'est effectuée dans le cadre d'une convention de cotutelle signée entre l'INP-HB l'institut hôte du laboratoire L2IS et l'université de Lille qui héberge le Laboratoire LEM sous la coordination des écoles doctorales EDP Yamoussoukro (Côte d'Ivoire) et SESAM Lille (France) et le Co-encadrement du Professeur Dabo-Niang et du Professeur Zoueu.

1.2 OBJECTIF DE LA THESE :

Vue le contexte supra défini, notamment le besoin d'optimisation des techniques d'analyses et de traitement d'images multispectrales pour parvenir à un diagnostic automatique, rapide et exact, les objectifs de la thèse sont les suivants :

- Développer une méthode de reconstruction d'image de fond basée sur l'image de l'échantillon qui puisse donner lieu à une nouvelle équation de normalisation.
- Proposer une méthode de détection et d'imputation des valeurs aberrantes afin de traiter les pixels de hautes intensités représentant les bruits impulsionnels dans l'image acquise par le microscope multispectral et multimodal.
- Développer une méthode de segmentation d'image multispectrale, rapide et efficace dans le contexte d'image de faible contraste.
- Proposer une méthode de classification (différenciation) des érythrocytes selon qu'ils soient infectés ou non par le parasite du paludisme.
- Mettre en place une méthode de comptage qui permette d'établir la parasitémie élément primordial pour la conduite thérapeutique.

La littérature regorge de différents types d'approches quant au traitement et à l'analyse d'images. Pour ce qui nous concerne, nous nous intéressons aux possibilités qu'offrent les techniques mathématiques et statistiques pour l'analyse et le traitement des données d'images multispectrales. Plus précisément, il s'agit de mettre en œuvre les techniques géostatistiques d'analyse variographique pour la reconstruction d'un champ aléatoire gaussien, le rééchantillonnage par **Bootstrap**, la détection des valeurs aberrantes par la méthode **Hampel**, la standardisation statistique en application de la loi de **Beer Lambert** et le déploiement de techniques d'apprentissage automatique pour la classification des cellules sanguines.

1.3 ORGANISATION DE LA THESE :

Ce manuscrit est développé en six chapitres, précédés de ce chapitre introductif et suivis par la conclusion et les perspectives.

Le chapitre 1 de ce manuscrit de thèse présente les objectifs de nos présents travaux et en précise le contexte.

Le chapitre 2 énonce la problématique du diagnostic automatique du paludisme. Il en donne les forces et les faiblesses ainsi que les enjeux économique et sanitaire de ce type de diagnostic.

Le chapitre 3 se focalise sur la description du dispositif expérimental constitué essentiellement du Microscope Multispectral et Multimodal. Il présente également l'état de l'art en matière d'acquisition des images et des traitements statistiques associés.

Le chapitre 4 est dédié à notre première contribution qui concerne la reconstruction de l'image de fond d'un frottis sanguin. Elle présente les deux méthodologies utilisées à savoir le rééchantillonnage par Bootstrap selon le principe de Monté Carlo et la simulation géostatistique d'un champs aléatoire Gaussien basée sur l'analyse variographique. Pour cette contribution, un article a été soumis.

Le chapitre 5 s'intéresse au processus de détection et de segmentation des érythrocytes dans les frottis non marqués sur la base de la loi de Beer Lambert. Il donne une description explicite des méthodes mises en place dans ce processus. Cette deuxième contribution a fait l'objet d'un article publié.

- ✓ S. Doumun, S. Dabo and J. Zoueu, "Detection and segmentation of erythrocytes in multispectral label-free blood smear images for automatic cell counting", *J. Spectr. Imaging* 9, (2020). doi: <http://dx.doi.org/10.1255/jsi.2020.a10>.

Le chapitre 6 présente les différentes approches de classification et les différents algorithmes mis en œuvre pour chaque approche et en présente les résultats. Cette dernière contribution sera soumise prochainement.

Le chapitre 7 Présente le bilan du travail effectué et en donne les perspectives.

Diagnostic Automatique du Paludisme

2.1 INTRODUCTION

Qualifiée de tueuse silencieuse, la maladie à plasmodium continue de faire des ravages socio-économiques et d'endeuiller de nombreuses familles à travers le monde. Loin devant le COVID19 (131 millions de cas en 2 ans)[11] et l'Ebola (64 000 cas en 45 ans)[12], le paludisme enregistre 229 millions de cas et 409 000 décès[13], chaque année.

L'Afrique subsaharienne est la région du monde qui paye le plus lourd tribut, depuis plus cinq décennies tant sur le plan sanitaire que socio-économique. Elle a enregistré, à elle seule, 94% de taux de morbidité, de mortalité et une perte annuelle de 12 milliards de dollars de PIB[14]. Bien qu'ayant à son actif de réels progrès et des exploits du point de vue de la recherche, un demi-siècle de lutte n'a pu parvenir à une solution définitive dans la prise en charge de cette pandémie.

Toutefois, une prise en charge rapide est salutaire dans le traitement de cette maladie. Une surveillance efficace par un diagnostic précoce et exact est cruciale dans ce contexte. En effet, elle permet non seulement, d'éviter que la maladie ne devienne mortelle mais réduit son intensité et contribue à freiner sa transmission. Cependant, force est de constater que ces décennies de recherche n'ont pu remplacer la microscopie qui demeure la méthode de référence en matière de diagnostic du paludisme, par la goutte épaisse et le frottis sanguin quoiqu'elle soit limitée. Pour cause, le processus, entièrement manuel, peut être source d'erreur tant au niveau de la confection et du marquage de l'échantillon que de l'analyse microscopique qui s'en suit. Il est subjectif, dépend du niveau de parasitémie et sujet à une lenteur qui rend difficile la prise en charge des cas urgents. Face à ces insuffisances, la mise en place de techniques de diagnostic automatique est un domaine très actif et des nouveaux instruments ont vu le jour.

Dans ce chapitre, nous vous présenterons les différentes techniques de diagnostic automatique développées ces dernières décennies. Les forces et les faiblesses de celles-ci ainsi que quelques travaux connexes.

2.2 PROBLEMATIQUE :

Malgré des progrès remarquables et les avancées technologiques, la lutte contre le paludisme connaît une stagnation ces dernières années [15]; voire perd du terrain, en particulier dans les pays OMS de l'Afrique où la charge est élevée. Le nombre de cas en baisse entre 2010 (247 millions) et 2014 (217 millions) a amorcé une remontée dès 2015 pour atteindre 229 millions de cas en 2019.

Le diagnostic rapide et un traitement précoce représentent l'un des piliers incontournables de la lutte contre cette pandémie. Selon l'OMS [16] les perturbations dans le diagnostic et le traitement du paludisme pourraient entraîner des dégâts et des mortalités considérables. En effet, toujours selon l'organisme, 1% de perturbation entraînerait 19 000 décès supplémentaires et 25 à 50 % de perturbations 46 000 à 100 000 décès supplémentaires en Afrique subsaharienne. De plus, face à la menace de résistance constante aux médicaments antipaludiques, l'OMS recommande que dans tous les cas présumés, le paludisme soit confirmé par un diagnostic basé sur la recherche des plasmodiums.

Demeurant la référence, les techniques de la microscopie conventionnelle de frottis sanguin et de goutte épaisse permettent une étude morphologique des hématozoaires et la différenciation des espèces pour le frottis et une très bonne sensibilité pour la goutte épaisse. Elles utilisent une méthodologie simple mais exigeante en précision et rigoureuse qui requiert un long apprentissage et dont la sensibilité est corrélée avec le temps d'observation. La réalisation de ce type de diagnostic reste un défi même pour un biologiste averti. Non seulement, sa pratique n'est pas exempte d'erreur humaine tant au niveau du marquage que pour l'observation ; mais encore, le résultat dépend du niveau de parasitémie du patient. Il peut être négatif pour les parasitémies faibles. Aussi la lenteur de cet examen rend-elle difficile la prise en compte de l'urgence palustre d'où la nécessité d'un diagnostic automatique.

Les nouvelles méthodes de diagnostic à promouvoir dans le cadre des recherches devront atteindre un double objectif ; être accessible à tous, surtout aux populations défavorisées ou en situation d'urgence, être autonome et fournir un diagnostic invariablement exact.

2.3 MOYENS :

2.3.1 Mobilisation de l'opinion internationale à réalité du paludisme :

Faire régresser le paludisme d'ici 2030 étant l'un des objectifs du millénaire, le plan international de lutte contre le paludisme, mène une campagne afin de faire connaître le fléau que représente cette maladie au grand public. Se faisant, il incite l'opinion internationale à se mobiliser autour de la question. Le but pour l'OMS est d'inciter les politiques et les organismes privés à investir davantage dans la lutte contre cette pandémie à travers le financement des actions de prévention, la recherche de nouvelles molécules pour le traitement rapide du paludisme et la mise en place de nouveaux outils de diagnostic. Etant donné que, sur le terrain, la prise en charge précoce des malades du paludisme par un traitement efficace est salutaire. Il est donc impératif de développer de nouveaux outils de diagnostic rapide et automatique. L'OMS dans son communiqué de presse du 20 Novembre 2020 [16] lance cet appel à l'innovation suivant : « *L'élimination du paludisme dans tous les pays, en particulier ceux où la charge de morbidité est élevée, nécessitera probablement des outils qui ne sont pas disponibles aujourd'hui. En septembre 2019, le Directeur général de l'OMS a lancé un "défi de l'éradication du paludisme", appelant la communauté mondiale de la santé à investir davantage dans la recherche-développement de nouveaux outils et de nouvelles méthodes de lutte contre le paludisme. Ce message a été réaffirmé dans le rapport d'avril 2020 du Groupe* ». Invitant ainsi la communauté internationale à davantage investir dans la prise en charge de cette pandémie.

2.3.2 Mobilisation sur le terrain :

Le travail sur le terrain concerne essentiellement la sensibilisation, la prévention et les traitements. La sensibilisation a pour but de faire comprendre l'origine et les modes de

transmissions du paludisme. Les populations sont ainsi encouragées à utiliser des moustiquaires afin de prévenir la transmission de la maladie.

Cette prévention passe également par l'assainissement du cadre de vie des populations et la distribution de moustiquaires imprégnées longue durée. L'utilisation de ce dernier reste le moyen le plus efficace de protection individuelle contre cette maladie.

Pour ce qui concerne le traitement, la prise en charge précoce et efficace par un médicament antipaludéen des malades passe par un diagnostic rapide et exact.

2.4 ENJEUX SOCIO- ECONOMIQUES [17], [18] :

La réalité de l'impact négatif du paludisme sur les performances économiques des entreprises ne fait aucun doute. L'impact socio-économique de cette maladie sur les populations et en particulier, en milieu professionnel continue d'être élevé malgré les efforts de lutte pour l'éradication de cette parasitose.

Cette maladie est l'une des principales causes de morbidité et mortalité dans le monde. Elle est responsable de la majorité de l'absentéisme des travailleurs et de la baisse de la productivité des entreprises. Elle est perçue en cela comme une sérieuse menace au développement économique des entreprises par les responsables de celles-ci. Considéré comme une maladie de la pauvreté qui favorise le renforcement de cette dernière, le paludisme est un frein et très coûteux pour le développement économique des pays endémiques.

Le préjudice causé par le paludisme touche toutes les couches sociales :

- ✓ Les employeurs par l'impact direct sur la main d'œuvre et indirect sur l'environnement économique. En effet, un épisode palustre fait perdre à l'entreprise environ 10 jours de travail en plus de la baisse de productivité qui s'en suit à la reprise du travail par l'employeur. De plus, la prise en charge des frais médicaux est un poids non négligeable pour les entreprises.
- ✓ Au niveau de l'employé et du citoyen lambda, le paludisme affecte aussi bien physiquement, moralement qu'au niveau de son revenu. Selon les propos du

dirigeant de Roll Back Malaria dans son rapport 2017 [5], « *Il a été démontré que chaque cas de paludisme avait un coût pour les ménages estimé à 2,67 dollars US (entre 0,34 dollar et 7,66 dollars US), correspondant aux frais directs déboursés. Chez les adultes, la maladie provoque une perte de productivité moyenne de 3,4 jours (entre 2 et 6 jours), soit un coût indirect supplémentaire de 10,85 dollars US au minimum. Les mères et les autres adultes qui s'occupent des malades sacrifient 2 à 4 jours additionnels chaque fois qu'un enfant ou un autre membre de la famille contracte le paludisme, ce qui génère à nouveau des coûts indirects supplémentaires pour les ménages* »

Sachant, qu'une personne n'est pas immunisée après une crise de paludisme, elle peut faire plusieurs crises au cours d'une période donnée et qu'elle peut le transmettre, il en ressort sans aucun doute que le paludisme est très nuisible à l'environnement socio-économique des pays endémiques. Il est donc indéniable de mettre en place une politique de prise en charge rapide et précise afin de réduire l'intensité de la maladie et les risques de transmission par un diagnostic et un traitement adéquat.

2.5 LES TECHNIQUES DE DIAGNOSTIC AUTOMATIQUES DU PALUDISME :

Pour atteindre l'objectif de la prise en charge précoce du paludisme par un diagnostic rapide et efficace, des efforts ont été consentis pour mettre en place des outils de diagnostic automatique. Dans la littérature, de nombreuses techniques ont été rapportées. Elles permettent de détecter les parasites du paludisme chez un patient ou de mettre en évidence sa présence chez ce dernier. Dans ce qui suit nous rappelons quelques méthodes de diagnostic automatiques du paludisme, les forces et les faiblesses de celles-ci.

2.5.1 La technique d'immunochromatographie (ICG) [14] [17] [19] [20] [21] :

Le Test de Diagnostic Rapide en abrégé (TDR) pour le paludisme, aussi connus sous l'appellation "Dipsticks", est l'outil diagnostique recommandé par l'OMS dans toutes les structures de soins privées comme publiques dépourvues de service de microscopie; et également à l'échelle communautaire dans le cadre de la Prise En Charge du Paludisme à Domicile (PECADOM). Ces tests reposent sur la détection des protéines spécifiques de l'hématozoaire à l'aide d'une bandelette réactive par immunochromatographie. Plusieurs types de protéines plasmodiales peuvent être détectés selon la particularité du test. L'histidine-rich protéine-2 (HRP2) est spécifique du *P. falciparum*, l'espèce la plus dangereuse. La lactate déshydrogénase plasmodiale (pan-pLDH) comme l'aldolase est commune à toutes les espèces.

Cette méthode a le gros avantage d'être rapide, avec une réponse dans le quart d'heure. Il est facile d'interprétation et ne requiert aucune compétence particulière pour la lecture. Les spécificités et sensibilités de ces tests varient selon le niveau de parasitémie. De l'ordre de 80-90% de sensibilité à un seuil de 200 parasites/ μ l, elle augmente à 95-100% à 2000 parasites/ μ l (21) en comparaison à la microscopie (100%). Les TDR sont ainsi pertinents pour le diagnostic de personnes symptomatiques et n'exigent pas de matériel coûteux ou compliqué.

Cependant, il présente plusieurs inconvénients dont :

- ✓ Un nombre important de faux positifs ; dans la mesure où, ces tests peuvent être positifs plusieurs jours après guérison de la pathologie, dû à la présence d'auto-anticorps ou de facteur rhumatoïde à des taux élevés. De plus, la persistance de la circulation de l'HRP2 après disparition des parasites du sang est trouvée jusqu'à 15 jours après la négativité des tests microscopiques. Une lecture tardive après le dépôt des réactifs peut être également à la base de cet état des faits.
- ✓ Les faux négatifs sont dus à une faible parasitémie de l'ordre de 100 parasites par μ l, soit 0,002 % d'hématies infectées.

2.5.2 QBC (Quantitative Buffy Coat) [24] :

Le principe de cette technique microscopique de fluorescence repose sur l'utilisation d'un fluorochrome (l'acridine orange) capable de se fixer sur le noyau du parasite. La recherche du Plasmodium se fait dans 50µl de sang recueillis dans un tube à hématocrite, après concentration par centrifugation et lecture au microscope à fluorescence. La sensibilité de cette technique serait comparable à celle de la Goutte Epaisse pour des infections supérieures à 100 parasites/µl. Elle varie de 41% à 93% pour des parasitémies inférieures à 100 parasites/µl. La spécificité pour *P. falciparum* est élevée (93-98%) mais chute à environ 50% pour les infections causées par les autres espèces.

De très bonne sensibilité et spécifique, elle a l'avantage d'être rapide. Elle est d'une excellente valeur prédictive négative, elle peut détecter la présence du parasite même en cas de très faible parasitémie. Le QBC est facile d'apprentissage.

Il présente cependant un inconvénient majeur de toxicité de l'acridine et donc nécessite un personnel expérimenté mais aussi le coût élevé du fait de de l'appareillage et des réactifs, donc inapproprié sur le terrain. De plus, il ne permet pas le diagnostic d'espèce.

2.5.3 Diagnostic Moléculaire [24] :

L'amplification génique par PCR est la technique la plus sensible qui permet de détecter de très faibles parasitémies de l'ordre de 0.3 parasite/µl de sang, avec une possibilité de quantification de l'ADN plasmodial en utilisant la PCR quantitative, et ce chez des patients parfois asymptomatiques. La PCR a également une excellente valeur prédictive négative avec une spécificité absolue si elle est réalisée dans de bonnes conditions. L'amplification du gène codant pour la petite sous unité 18S de l'ARN ribosomal permet aussi l'identification des espèces plasmodiales en cause en utilisant une "Nested" PCR, ainsi que la différenciation entre une rechute et une réinfection.

Plusieurs méthodes basées la PCR existent [19] :

- ✓ la PCR nichée (nested-PCR) permet la détection sans quantification de parasites à un seuil de 2 à 6 parasites/µl ;

- ✓ la PCR multiplex a une sensibilité supérieure, détectant jusqu'à 0,2 parasite/ μl et la QT-NASBA (nucleic acid sequence based amplification) jusqu'à 0,01 parasite/ μl .
- ✓ la PCR quantitative permet de mesurer la parasitémie et la RT-PCR différencie les formes adultes des formes gamétocytaires. De sensibilité similaire à la PCR nichée, la LAMP (loop-mediated isothermal amplification) est un outil de biologie moléculaire utilisable en milieu isolé. La sensibilité des méthodes PCR dépend principalement du volume analysé.

La PCR a l'inconvénient d'avoir un coût élevé. Réservé aux laboratoires spécialisés, elle nécessite une bonne technicité car il présente un risque de contamination et a un temps d'analyse relativement long (24 h environ) [22]. Cette expertise est insuffisamment présente dans les régions isolées où le paludisme est prédominant.

2.5.4 Travaux connexes :

En marge de ces techniques de diagnostic automatique du paludisme, plusieurs travaux de recherche ont vu le jour et des instruments de technologie de pointe ont été développés, ayant des applications qui sont des prouesses du point de vue de la recherche scientifique.

Nous pouvons citer en cela :

2.5.4.1 Le microscope multispectral multimodal pour diagnostic du paludisme :

Construit dans le cadre de la lutte contre le paludisme, par un diagnostic automatique, il permet à travers l'acquisition et le traitement d'une image de dimension 4 (2 dimensions spatiales, 1 géométrique et 1 suivant les longueurs d'onde) de pouvoir établir le diagnostic différentiel du paludisme par un marquage spectral [23] [24]. Cette image est acquise suivant 13 longueurs d'onde en 3 géométries. Toutefois cette technique se heurte à la

qualité de l'image qui comporte des bruits optiques et électroniques dont l'extraction conduit à un très long temps d'acquisition.

2.5.4.2 Systèmes basés sur téléphone mobile [25] [26] :

L'objectif est de développer un système de diagnostic du paludisme automatisé et portable. Un tel système, développé pour les appareils mobiles (téléphones mobiles et tablettes), résout le problème d'accessibilité des pays à faibles revenus, par des outils de diagnostic peu coûteux. Il permet l'accès aux centres de santé dans les communautés éloignées et, surtout, ne dépendant pas d'une expertise approfondie ou d'un équipement de détection diagnostic coûteux. Ce système utilise des échantillons de sang colorés au Giemsa combinés à une microscopie optique pour identifier l'espèce *Plasmodium falciparum* qui est en cause des cas mortels du paludisme. Dans ce type de système des techniques de traitement d'image et d'intelligence artificielle sont couplés aux algorithmes de reconnaissance basés sur des classificateurs faibles avec un apprentissage adaptatif pour identifier les parasites *Plasmodium*. Pour faciliter l'apprentissage aux algorithmes de reconnaissance un prétraitement est apporté en supprimant l'arrière-plan autour des cellules sanguines.

2.5.4.3 Diagnostic par force magnétophorétique [22] [27] :

Cette technique tire profit des propriétés magnétiques significativement différentes entre les globules rouges sains et les globules rouges parasités. En effet, l'hémozoïne, pigment caractéristique du paludisme, a une structure cristalline qui modifie les propriétés magnétiques et optiques du globule rouge [28], [29]. Le développement de la technique de diagnostic par force magnétophorétique permet une détection de haute sensibilité de l'hémozoïne. Toutefois, l'utilisation d'un champ magnétique externe est une limite non négligeable à la performance de détection et de séparation des cellules.

2.6 CONCLUSION

L'automatisation du diagnostic du paludisme est une solution prometteuse dans la lutte contre cette pandémie mais elle reste un défi et un domaine de recherche très actif. Les outils actuels sont souvent coûteux et impliquent des composants matériels sophistiqués, ce qui rend difficile leur déploiement dans les zones à ressources limitées qui sont pourtant les plus affectées. En l'absence de vulgarisation de ces méthodes du fait des limites non négligeables, la microscopie optique reste la méthode de diagnostic de référence et la plus accessible aux centres de santé des pays endémiques. Cependant, pour un diagnostic, l'OMS recommande, préalablement au traitement, une confirmation parasitologique rapide par le TDR [30] comme solution pour les zones difficiles d'accès et / ou il n'existe pas de personnel expérimenté. Ces tests devraient être confirmés par une goutte épaisse afin de caractériser l'espèce plasmodiale en cause, et de déterminer la parasitémie qui définit la conduite thérapeutique.

Acquisition d'images : le
Microscope Multimodal et
Multispectral

3.1 INTRODUCTION

La méthode standard et traditionnelle de diagnostic du paludisme dans les centres de santé des pays en développement est la goutte épaisse et /ou le frottis sanguin. Cette technique utilise un microscope classique à source blanche et le marquage chimique de l'étalement de sang pour colorer les parasites, les identifier et les compter manuellement. Le résultat de ce type de diagnostic dépend entièrement de l'expérience du microscopiste. Aussi, cet examen est très couteux en temps, ce qui rend difficile la prise en charge rapide des cas urgents.

C'est donc, dans le souci d'aider à faire un diagnostic rapide et invariablement exact, que le laboratoire L2IS a initié et développé avec ses partenaires un microscope multimodal et multispectral pour le diagnostic du paludisme. Ce diagnostic est possible à travers l'acquisition, l'analyse et le traitement d'images multispectrales de frottis sanguin non marqué. Ainsi le marquage préalablement chimique est remplacé par un marquage spectral ; aussi le comptage qui s'en suit est entièrement automatique. Il tire ainsi profit de la technologie d'imagerie multispectrale.

Relativement nouvelle, cette technologie combine les avantages de la spectroscopie optique et la puissance de la visualisation d'objet obtenue par imagerie optique. La spectroscopie, méthode versatile, étant la science de l'acquisition et de l'exploration des caractéristiques spectrales de la matière. Elle est l'étude de l'interaction lumière-matière qui permet d'analyser d'une manière rapide et non-invasive sa pureté, ses propriétés électroniques et certains paramètres physico-chimiques. Quant à l'imagerie, elle est la science et la technologie d'acquisition et d'exploration de données spatiales et temporelles à partir d'objets dans le but d'obtenir des informations. Cette combinaison donne à l'imagerie multispectrale, la capacité d'établir des empreintes spectrales uniques des objets captés dans les régions électromagnétiques d'intérêt et procure des performances positives là où la vision couleur ne permet pas de conclure.

Dans une image multispectrale, chaque pixel (échantillon spatial) véhicule un spectre sous forme d'un vecteur de valeurs scalaires mesurant le rayonnement électromagnétique

incident. Ceci crée une image multivariée à trois dimensions contenant l'information spatiale (deux dimensions) et l'information spectrale (une dimension). Ces images en niveau de gris sont constituées de données capturées à des fréquences spécifiques du spectre électromagnétique.

C'est donc, cette technologie qui a été mise en œuvre à travers le dispositif expérimental décrit dans ce chapitre.

3.2 LE MICROSCOPE [31] [35]:

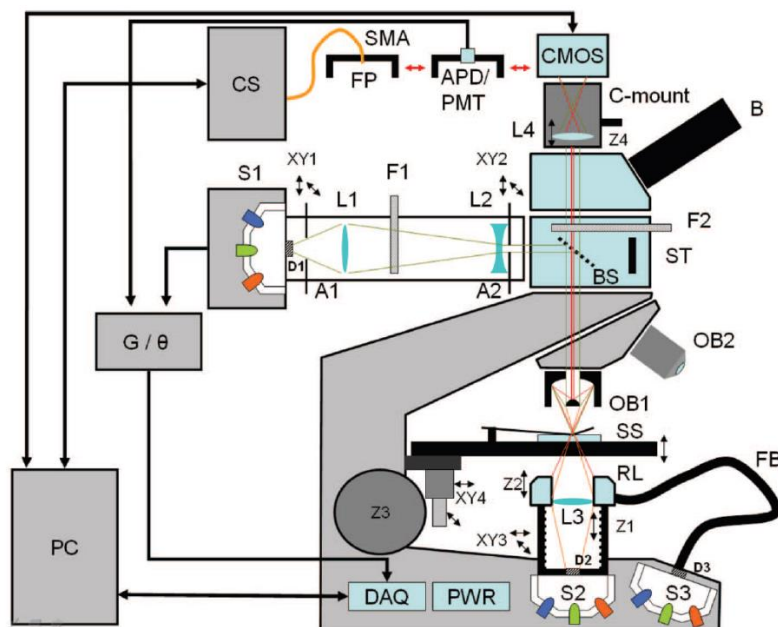


Figure 4: Microscope multispectral et multimode [35]

L'instrument est un microscope métallurgique modifié ; microscope de Brunel compound model SP80. Ce dispositif expérimental est constitué de trois parties : un module optique, un module électronique et un module logiciel, dont l'ensemble permet l'acquisition des images multispectrales à analyser pour établir le diagnostic automatique du paludisme.

3.2.1 Partie optique :

Dans cette partie, la source de lumière blanche à filament d'origine servant pour l'imagerie réfléchissante a été remplacée par trois nouveaux modules de multiplexage DEL S1, S2 et S3. Ces modules sont composés chacun de 13 DEL réparties sur un support diffuseur de forme parabolique dont le foyer correspond à la sortie du faisceau. Ces sources lumineuses représentent trois géométries d'acquisition : S1 pour le mode réflexion, S2 pour le mode transmission et S3 pour la diffusion. Pour chaque multiplexage donné, les 13 diodes émettent respectivement les longueurs d'ondes suivantes : 375 nm, 400 nm, 435 nm, 470 nm, 525 nm, 590 nm, 625 nm, 660 nm, 700 nm, 750 nm, 810 nm, 850 nm and 940 nm. Ces 3 modules sont associés à leurs sorties à des diffuseurs (respectivement D1, D2 et D3). Ces diffuseurs ont pour rôle d'obtenir, à la sortie de chaque source, des faisceaux avec une répartition spatiale uniforme, homogène et identique, de type Lambertien dont la radiance spatiale ne varie pas avec l'angle et avec la géométrie de la DEL utilisée [37] (Figures 4 - 5).

S1, S2 et S3 fournissent un éclairage d'excitation monochromatique de la gamme UV au proche infrarouge à partir de 3 lobes angulaires différents par rapport au lobe d'observation.

La source S3 (mode de diffusion) est conduite vers l'échantillon à l'aide d'une fibre optique (FB) couplée à un anneau (RL) qui permet de rayonner un faisceau circulaire d'épaisseur 3mm sur un diamètre de 60mm. Cet anneau peut être réglé suivant l'axe vertical (Z2) et permet ainsi de faire varier l'angle de diffusion.

La source S2 du mode transmission est conduite à l'échantillon (SS), à l'aide d'une lentille convergente (L3) fixée à un canal (Z1) disposant des mouvements de translations XY3. Une manette (Z3) permet de déplacer le porte-échantillon dans le sens vertical avec une possibilité de mouvement dans le plan de l'échantillon (XY4). Le microscope dispose d'objectifs de type plan (OB2) et de type casse grain (OB1). L'objectif de type casse grain a été choisi, pour faire face aux problèmes d'aberration chromatique. En effet, les objectifs traditionnels de type plan, entraînent un changement du plan focal, lorsque l'on change de longueur d'onde d'illumination. Pour conduire les faisceaux de la source S1 vers le

porte-échantillon, nous disposons d'une lentille convergente (L1), d'une lentille divergente (L2) de deux diaphragmes (A1 et A2), de mouvements de réglage XY1 et XY2 et d'une fente réglable F1. Cet ensemble permet d'aider à l'alignement du faisceau et fixer la taille du faisceau.

L'oculaire mécanique initial du microscope de Brunel a été substitué par une caméra monochrome CMOS présentant une profondeur de codage de 12bits/pixel (5 mégapixels Guppy - 503B, Vision Allied Technology, avec un capteur de MT9P031 de Micron / Aptina), avec une taille de pixel de $2,2 \mu\text{m} \times 2,2 \mu\text{m}$. Il est précédé d'un système de focalisation (C-mount) composé d'un ensemble de lentilles, désigné par L4 et d'un système de réglage Z4.

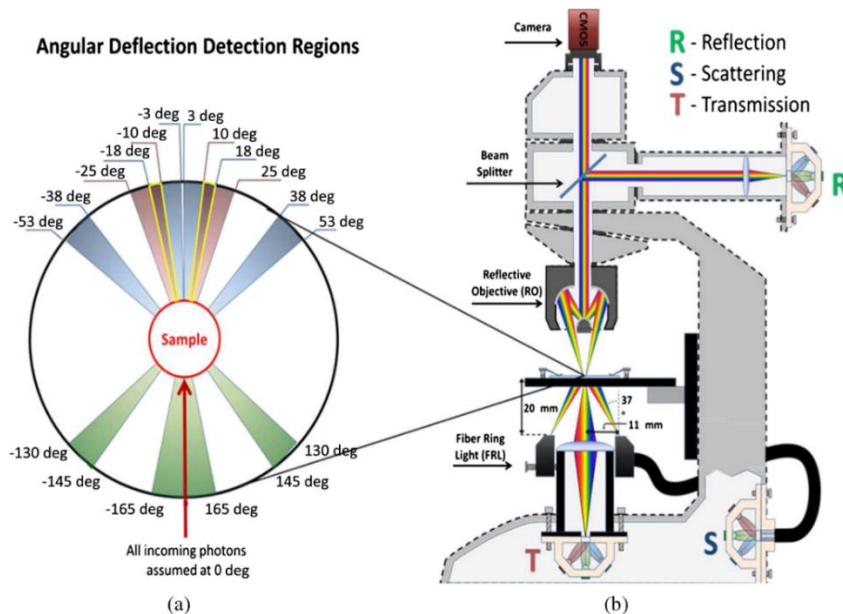


Figure 5: Diagramme des lobes angulaires et aperçu d'une section transversale verticale du microscope multimode multispectral (MERDASA et al., 2013)[35]

3.2.2 Partie électronique :

Le circuit électronique est contrôlé par une carte d'acquisition de données USB (DAQ, National Instruments, NI USB-6009) et permet d'effectuer le multiplexage entre différents modes angulaires et différentes bandes spectrales. Les DEL sont alimentées par une tension continue de 9 Volts. Outre le DAQ contrôlant le courant et la source, les

entrées analogiques permettent de caractériser les sources en termes de courbes de tension (U-I). La caméra est reliée à un ordinateur par un système de transmission.

3.2.3 Partie logicielle [31] [32] [35] :

L'équipement est contrôlé par le logiciel LabVIEW (National Instruments, NI). Le programme contrôle l'exposition et le gain de la caméra, le multiplexage des DEL et le courant à travers la carte d'acquisition de données NI, DAQ. L'interface graphique est construite autour d'une image d'aperçu en direct de la caméra pour une localisation et une mise au point faciles. Les curseurs sont utilisés pour ajuster le gain, l'exposition et les courants des DEL, tandis que les menus déroulants sont utilisés pour sélectionner la bande de longueur d'onde et le mode angulaire, c'est-à-dire la réflexion, la transmission ou la diffusion. Un histogramme en direct permet à l'opérateur d'ajuster les paramètres, pour éviter la saturation de la plage dynamique. Lorsque les paramètres d'un mode et d'une bande ont été choisis, ils sont ajoutés à une ligne dans un protocole de mesure. Les colonnes du tableau sont : mode angulaire, longueur d'onde, courant DEL, gain de la caméra, exposition, pause entre l'exécution de chaque ligne. Chaque ligne correspond à l'acquisition d'une image monochromatique. Lorsque l'ensemble du protocole est exécuté, le programme enregistre les images sur le disque dur avec une copie du protocole de mesure.

3.2.4 Les modalités [52] [36] :

Le microscope multimodal et multispectral utilise trois géométries d'acquisition qui correspondent à trois différentes techniques spectroscopiques. Sachant que les méthodes spectrométriques se composent en catégories à savoir la spectrométrie des rayonnements et la spectrométrie de masse, les techniques utilisées dans cette étude concerne la spectrométrie de rayonnement. Cette dernière est fondée sur l'étude des spectres de l'interaction entre la matière et les rayonnements électromagnétiques qu'elle absorbe, transmet, émet ou diffuse.

Dans cette section, nous donnerons une description succincte de ces différentes modalités d'acquisition développées dans le dispositif expérimental : la Transmission, la réflexion et la diffusion.

3.2.4.1 Transmission / Absorption [31] :

Dans la spectroscopie de transmission, la source et le détecteur sont disposés de part et d'autre de l'échantillon. Les rayonnements émis par la source interagissent avec l'échantillon, en étant plus ou moins absorbés par l'échantillon, pénètrent et se propagent dans le second milieu (celui du détecteur), en produisant un rayonnement transcrit en image visualisable. L'image obtenue est en niveau de gris et représente différents contrastes dépendant de la quantité d'électrons ayant traversé l'échantillon. Le but de cette technique étant de donner accès à la structure microscopique interne des objets observés, chaque point (pixel) de l'image correspond à une information le long d'une trajectoire linéaire à travers l'échantillon.

Quant à la spectroscopie d'absorption, elle consiste à mesurer l'atténuation de la radiation incidente traversant un milieu, pour en identifier les transitions électroniques et éventuellement en tirer les concentrations des substances absorbantes. Elle est liée à la transmission par une proportionnalité inverse. Celle-ci est mise en évidence par la loi de Beer-Lambert.

Loi de Beer-Lambert

La loi de Beer-Lambert permet de comprendre les phénomènes d'absorption des rayonnements électromagnétiques par les molécules d'un corps. C'est une loi fondamentale basée sur la comparaison de l'intensité du faisceau incident I_0 et du faisceau transmis I . En fait, cette loi traduit la proportionnalité entre l'intensité lumineuse absorbée à une longueur d'onde donnée et la quantité de matière absorbante. Dans ce cas, l'expression de cette loi prend la forme suivante :

$$A_\lambda = -\log(T_\lambda) = \varepsilon_\lambda c l \quad \text{avec} \quad T_\lambda = \frac{I}{I_0}, \quad I(l) = I_0 e^{-u_\alpha}$$

- A_λ Absorbance du milieu traversé par le rayonnement,

- T_λ Transmittance de ce même milieu,
- ε_λ Coefficient d'extinction molaire, caractéristique de la substance étudiée à une longueur d'onde donnée en $L mol^{-1}cm^{-1}$. Ce coefficient est une grandeur intrinsèque de la matière. Il révèle la capacité qu'a le milieu à absorber le rayonnement incident.
- l Épaisseur traversée en cm,
- c Concentration en mol/l ,
- u_α Coefficient d'absorption.

3.2.4.2 La diffusion de la lumière [37] [52] :

La diffusion d'une onde électromagnétique est le phénomène par lequel, un rayonnement électromagnétique tel que la lumière ou un faisceau de particules est dévié dans diverses directions par une interaction avec d'autres objets selon une loi angulaire qui lui est propre. Pour qu'il ait diffusion, il faut que l'indice de réfraction du corps soit différent de celui du milieu environnant. Il existe plusieurs types de diffusions selon les couples de particules impliqués. Les caractéristiques de la diffusion vont donc être dépendantes à la fois de la longueur d'onde du rayonnement incident et de la taille des particules. L'origine du phénomène peut être associée à plusieurs phénomènes élastiques lorsque le faisceau diffusé se fait sans variation de la longueur d'onde :

Diffusion Rayleigh à partir de dipôles tels que les molécules ; Diffusion de mie à partir de sphères réfractives ou réfractions aléatoires par interface interne dans des matrices troubles. Il existe également des processus de diffusion inélastiques tels que la diffusion de Raman et la diffusion Compton. Le phénomène de diffusion est dit isotrope lorsqu'il est reparti uniformément dans toutes les directions ou sinon il est dit anisotrope.

L'intensité du faisceau diffusé dépend d'un certain nombre de paramètres tels que : l'angle d'incidence, l'indice de réfraction, la longueur d'onde et le coefficient de diffusion. Lorsqu'il est éclairé, un corps diffusant devient une source quasi-ponctuelle de lumière et l'intensité du rayonnement diffusé décrit autour de chaque particule diffusante une nappe

qui est caractéristique de la particule. Si la géométrie de cette dernière est isotrope, cette nappe est symétrique par rapport à la direction de propagation de l'onde incidente. La forme de cette nappe dépend fortement du rapport entre la dimension de la particule et la longueur d'onde de la radiation incidente. Pour des particules de petites tailles devant la longueur d'onde, on parle de diffusion Rayleigh. En revanche, pour des particules de tailles supérieures ou voisines de la longueur d'onde, on parle de diffusion de Mie. Dans ce cas, les particules à étudier sont sphériques, mais la théorie de Mie peut être appliquée sur des particules de formes quelconques en faisant des approximations.

3.2.4.2.1 Diffusion Rayleigh :

La diffusion Rayleigh a lieu lorsque la taille des particules est très faible devant la longueur d'onde (inférieures à $\lambda/10$) de la radiation si bien que les variations spatiales du champ électromagnétique peuvent être négligées à l'échelle du nuage électronique. La particule est alors soumise à la force magnétique et à la force électrique. La première est négligeable devant la seconde. Elle est un cas limite de la diffusion de Mie.

3.2.4.2.2 Diffusion de Mie [38] :

La théorie de Mie, est une théorie de la diffraction de la lumière par des particules sphériques. La diffusion de Mie désigne la diffusion par des particules dont le rayon oscille entre 0.1 et 10 fois la longueur d'onde. Les indicateurs d'intensité deviennent plus complexes, et la diffusion a lieu préférentiellement vers l'avant au fur et à mesure que la taille de la particule augmente (Figure 6) [37].

Dans le premier cas, le rayon de la particule est très inférieur à la longueur d'onde optique : cela correspond au domaine de diffusion de Rayleigh, pour lequel la diffusion se fait autant vers l'avant que vers l'arrière.

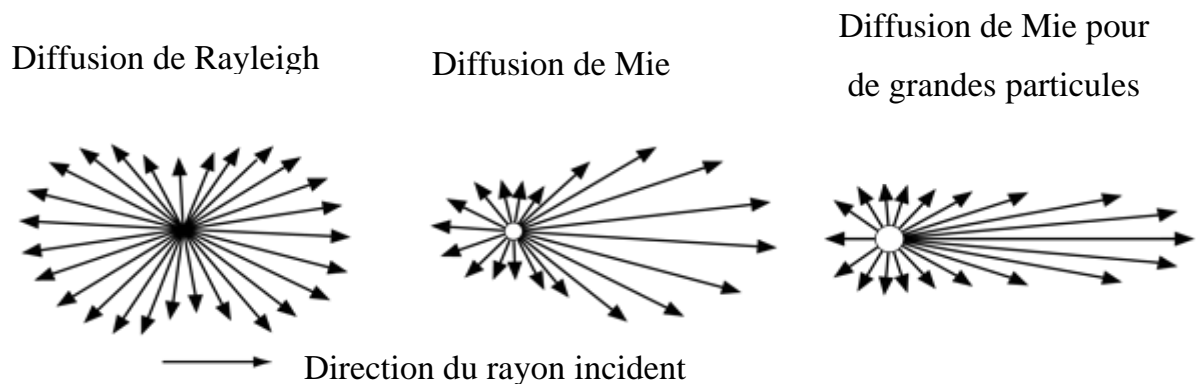


Figure 6: Indicatrices de la diffusion des particules sphériques dans le plan orthogonal à la polarisation (linéaire) de la lumière incidente.

3.2.4.2.3 Diffusion géométrique :

Lorsque les particules sont grandes devant la longueur d'onde (supérieures à $10 \cdot \lambda$), les lois de l'optique géométrique suffisent à traiter le problème. Dans les tissus biologiques, compte tenu de l'étendue de la distribution de taille des diffuseurs (Figure 6) ; tous les régimes de diffusion peuvent avoir lieu simultanément. Toutefois, les caractéristiques de diffusion observées dépendent beaucoup des conditions d'observation. En effet, sachant que la diffusion de Rayleigh est isotrope et celle de Mie surtout dirigée vers l'avant, la lumière transmise à travers le milieu sera en majorité affectée par de grosses particules (diffusion de Mie) alors que la contribution des petites particules (diffusion de Rayleigh) sera principalement observée dans la lumière rétrodiffusée. Au regard des structures de tailles diverses, les théories de Rayleigh et Mie ne sont pas directement applicables aux tissus biologiques. On raisonne alors à l'échelle macroscopique avec un coefficient de diffusion moyen caractérisant le milieu. Ce coefficient de diffusion μ_s exprimé en cm^{-1} , peut être défini de manière similaire à l'absorption. L'intensité balistique I (quantité de lumière qui traverse un milieu diffusant non absorbant d'épaisseur L sans être diffusée) s'exprime sous la forme :

$$I = I_0 \exp(-\mu_s L)$$

L'imagerie en diffusion apporte des informations disponibles sur les séquences d'imagerie conventionnelle.

3.2.4.2.4 Réflexion :

La réflexion est une technique spectroscopique qui consiste à récupérer le rayonnement électromagnétique réfléchi par la surface d'un objet. Il existe deux grands groupes de réflexion : les réflexions internes et les réflexions externes. Les réflexions internes intègrent la réflexion totale atténuée qui se manifeste sur des matériaux fortement réfractaires, facilitant la transmission des radiations infrarouge. Les réflexions externes, en revanche, intègrent la réflexion spéculaire et la réflexion diffuse qui diffèrent l'une de l'autre par l'état de la surface de l'échantillon à étudier [39]. Du point de vue spectrale, ce rayonnement réfléchi, par la surface d'un objet, renseigne sur la nature de l'objet. Ceci implique sa constitution, ses propriétés physiques et optiques à travers son indice de réfraction. En imagerie de réflexion, les rayonnements réfléchis produisent un signal transcrit en image visualisable. L'image obtenue en niveau de gris présente des contrastes permettant une vue en relief, avec un grand degré de détails caractérisant l'état de la surface de l'objet. En général, la plupart des surfaces sont assujetties à ces deux types de réflexion. Certaines effectuent principalement la réflexion spéculaire, et d'autres principalement la réflexion diffuse. En ce qui concerne les organismes biologiques, les deux types se manifestent, mais la réflexion diffuse est plus prépondérante [40].

Dans le cas général d'un corps constitué de différentes couches successives, le coefficient de réflexion peut être évalué selon la relation de Fresnel [41] qui est de la forme :

$$R(\theta_i) = \frac{1}{2} \left[\frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} + \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} \right]$$

Où θ_i est l'angle d'incidence et θ_t l'angle de transmittance. Ces deux angles sont liés par la loi de Snell Descartes, $n_i \sin \theta_i = n_t \sin \theta_t$ où n_i et n_t sont respectivement les indices de réfraction des milieux incidents et transmis.

3.2.4.2.5 Réflexion spéculaire et réflexion diffuse :

Comme préalablement définie, l'état de la surface de l'échantillon à étudier donnera, soit une réflexion spéculaire, ou une réflexion diffuse (Figure 7). La plus connue des deux est la réflexion spéculaire. Selon cette réflexion, la radiation incidente est réfléchiée selon la loi de Snell Descartes pour la réflexion [42]. Ceci est vrai pour des surfaces polies (polies du point de vue où la taille des irrégularités est inférieure à la longueur d'onde de la radiation incidente). Lorsque les irrégularités de la surface augmentent comparativement à la longueur d'onde de la radiation incidente, la réflexion s'effectue de plus en plus de manière aléatoire, on parle alors de réflexion diffuse. Dans ces conditions, si la radiation est transmise à travers une première interface, une partie de celle-ci est absorbée et le reste est alors diffusé. Les lois de la réflexion diffuse découlent d'un traitement empirique, le chemin optique exact est mathématiquement difficile à décrire. Comme conséquence, aucune théorie rigoureuse n'a été proposée à ce jour. Cependant, plusieurs théories approchées décrivant le phénomène ont été développées.

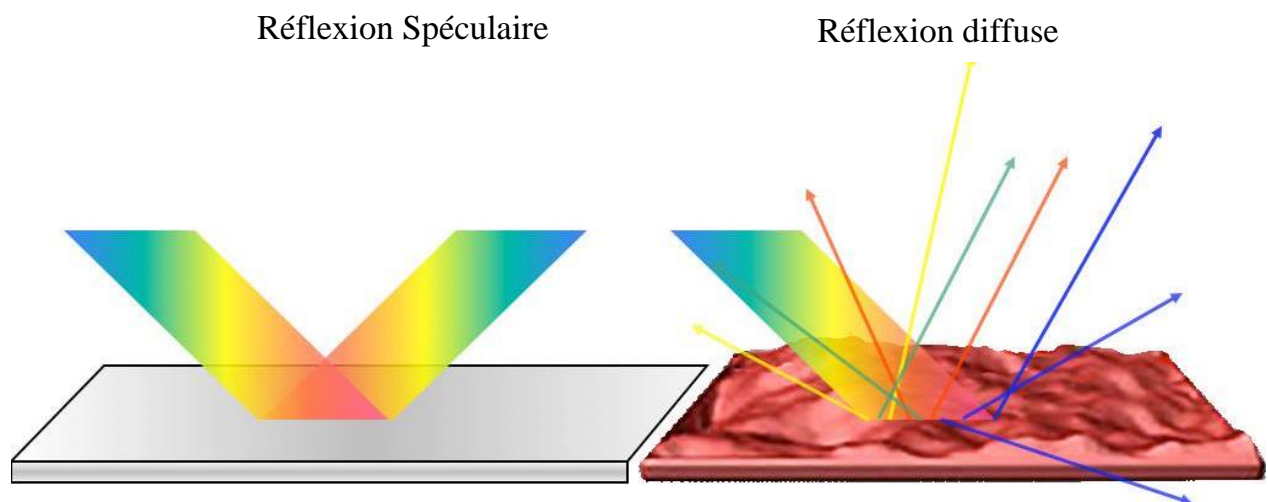


Figure 7: Représentation schématique des phénomènes de réflexion spéculaire (à gauche) et de réflexion diffuse (à droite)

En effectuant les mesures, il devient important de considérer les deux types de mesures qui sont en fait complémentaires pour pouvoir obtenir le plus grand nombre d'informations sur l'échantillon à étudier. La contribution relative de ces deux fonctions varie avec la surface géométrique. Des études ont montré que l'expression de la réflexion

diffuse est fonction des propriétés géométriques et optiques de l'échantillon [40], [43]. Shoji Tominaga et al ont aussi montré que l'expression de la réflexion spéculaire met en exergue la géométrie de la substance, par contre celle de la réflexion diffuse fait ressortir les caractéristiques chimiques et optiques du corps [39].

3.2.5 Les longueurs d'onde et leur intérêt :

Le paludisme ou malaria est due à un protozoaire de type Plasmodium transmis à l'homme par un moustique, anophèle femelle hématophage. Le cycle d'évolution du plasmodium comporte une phase sexuée ou sporogonique qui se déroule chez l'anophèle et une phase asexuée ou schizogonique qui se déroule chez l'homme. Cette phase schizogonique se déroule en deux étapes (la pré-érythrocytaire ou hépatique et érythrocytaire). A l'étape la pré-érythrocytaire, le parasite au stade de sporozoïte inoculé par le moustique, va se loger dans le foie (Figure 8).

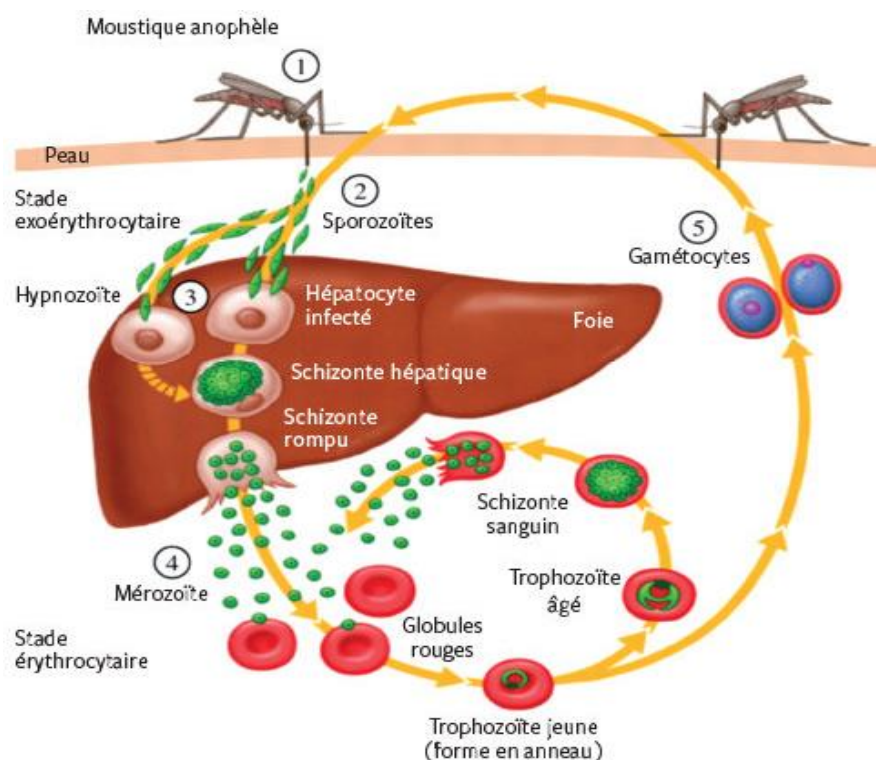


Figure 8: Schéma de reproduction du parasite dans le foie

Chaque sporozoïte se multiplie à l'intérieur d'un hépatocyte en milliers de mérozoïtes, forme du parasite qui infecte les érythrocytes. Lorsque le mérozoïte pénètre à l'intérieur du globule rouge, il baigne dans une mer d'hémoglobine qu'il utilise pour assurer sa croissance. Le parasite ingère l'hémoglobine et l'introduit dans sa vacuole qui est à pH acide. Il la décompose ainsi en globine qui apporte des acides aminés, et en hème qui apporte le fer et qui est transformée en hémozoïne par une hème-polymérase. La formation de l'hémozoïne, résidus constituant un pigment noir aussi appelé « pigment malarique » caractérise la présence de plasmodium dans le globule rouge.

Ce dernier également appelé hématie désigne des cellules sans noyau qui circulent dans le sang et servent à transporter l'oxygène et rejeter le gaz carbonique. Chaque hématie contient de l'hémoglobine, protéine riche en fer qui est l'agent de transport de l'oxygène dans l'hématie des poumons vers les différents organes du corps et du gaz carbonique des organes vers les poumons. Cette hémoglobine apporte la couleur rouge du sang. L'ingestion de celui-ci par le plasmodium entraîne une réduction de sa concentration dans l'érythrocyte et confère à ce dernier un noyau dû à la présence du parasite. Aussi le pigment noir de l'hémozoïne affecte la couleur de la cellule qui peut devenir foncée (Figure 9).

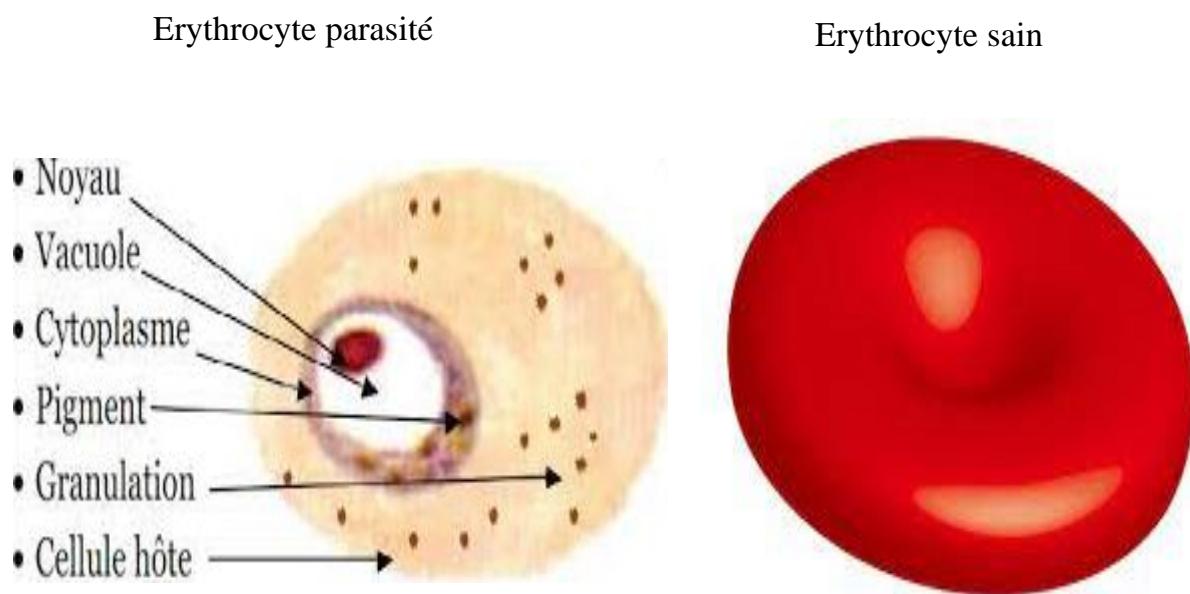


Figure 9: Caractérisation du globule rouge infectée par le plasmodium

Le diagnostic du paludisme est essentiellement basé sur l'étude des caractéristiques de contraste entre le globule rouge sain et le globule rouge infecté. En particulier, il est question dans ce travail, de l'analyse spectrophotométrique des caractéristiques du globule rouge à travers les intensités lumineuses transmises, réfléchies et diffusées à travers l'échantillon de frottis éclairé. Les longueurs d'onde allant de 375 à 940 nm ont été utilisées pour illuminer l'échantillon et la transmittance, la réflectance et la diffusion puis les intensités captées à travers l'échantillon ont été enregistrées. Ceci permet de quantifier la concentration en hémoglobine. Afin de convertir mathématiquement les intensités détectées aux différentes longueurs d'onde et les concentrations en hémoglobine, il faut prendre en compte le coefficient d'extinction de l'hémoglobine à ces longueur d'onde selon la loi de Beer Lambert [44] [45] :

$$A = \log \left(I_0 / I \right) = \epsilon c l$$

- l Épaisseur traversée en cm,
- c Concentration en mol/l ,
- ϵ Absorbance

Les globules rouges non parasités sont essentiellement composés d'hémoglobine et leurs spectres sont dominés par l'empreinte spectrale de l'hémoglobine. Les propriétés optiques de l'hémoglobine ont été bien analysées sur une large gamme spectrale [46] [47] [34]. Son spectre est caractérisé par la bande Soret (412 nm) et leurs deux bandes apparentées à 541 et 576 nm, en mode de transmission [46] [48] [49]. La longueur d'onde 660 nm correspond à sa faible absorption. Le proche infrarouge (700 à 940 nm) sera caractérisé essentiellement par une faible absorption de la lumière par l'hémoglobine [50].

Les cellules sanguines parasitées sont caractérisées par les propriétés optiques de l'hémozoïne qui diffère de l'hémoglobine par un décalage vers le rouge des bandes spectrales et une diminution de la concentration en hémoglobine. L'hémozoïne présente une bande d'absorption particulière à 630 et 660 nm. Wilson et coll. ont mesuré une diminution globale de la probabilité de diffusion de l'UV au proche infrarouge [34] [51].

3.3 CONCLUSION

L'analyse spectrophotométrique des cellules biologiques fait l'objet, ces dernières années d'un intérêt croissant pour sa capacité d'exploration substantielle des caractéristiques chimiques, physiques et physiologique des cellules. Une telle caractérisation sans radiation ionisante à partir de mesures simples ouvre la voie à des méthodes non invasives pour leur analyse, avec une instrumentation accessible à moindre coût. Propulsant ainsi les méthodes de diagnostic optique sur la scène en médecine et en biologie. Le Microscope multispectral et multimodal, instrument versatile offre cette possibilité de caractérisation spectrale à travers la mise en œuvre de trois géométries et treize longueurs d'onde définis de façon adéquate.

**A Novel Multispectral Images
Normalization Application to
Unstained Malaria Infected
Blood Smear**

Résumé :

Le laboratoire L2IS à Yamoussoukro a construit ces dernières années un microscope multispectrale capable d'acquérir suivant 13 (treize) longueurs d'ondes en 3 (trois) modes, un total de 39 images en niveau de gris. Cet instrument trouve ses applications dans divers domaines de recherche allant de l'agriculture à la santé. En particulier, le diagnostic automatique, rapide et précoce du paludisme, pour ce qui concerne la santé.

Pour atteindre l'objectif du diagnostic rapide, l'un des obstacles importants est le temps d'acquisition des images et leurs traitements. En effet, après l'acquisition des images, les bruits optiques et électroniques sont soustraits de celles-ci ; puis, un deuxième niveau de pré-traitement consistera à normaliser les images pour assurer l'homogénéité de l'éclairage, selon l'équation (Equation 1*) ci-après :

$$I_{\lambda} = \frac{I_S - I_B}{I_R - I_B} \quad \text{Avec} \quad \begin{array}{l} I_S \text{ Image de l'échantillon} \\ I_R \text{ Image de référence} \\ I_B \text{ Image du bruit résiduel} \\ I_B \text{ Image suivant } \lambda \end{array} \quad (1^*)$$

Cette opération nécessite l'acquisition 39 images au lieu de 13 pour chacun des 3 modes. Etant donné que l'image d'un échantillon de frottis sanguin est constituée de 39 images en niveau de gris, son acquisition exige 3 fois plus de temps pour répondre à ce besoin de normalisation. Une autre alternative consiste à estimer les bruits et la fonction d'éclairage à partir de la seule image contenant les cellules sanguines.

Cette première contribution répond à cette problématique par la reconstitution de fond de l'image du frottis sanguin. En effet, après avoir extraire les cellules sanguines de l'image de l'échantillon, nous reconstruisons les espaces vides laissés par celles-ci à partir du fond de l'image. Deux méthodologies ont été mise en œuvre dans ce travail :

- ✓ La reconstruction par la méthode géostatistique basée sur les paramètres spatiaux du fond de l'image, lesquels paramètres sont calculés à partir d'une analyse variographique,
- ✓ Le rééchantillonnage par Bootstrap par la méthode de Monté Carlo.

Ces deux méthodes en plus de réduire le temps d'acquisition, améliore la qualité de l'image. Toutefois, la méthode de reconstruction géostatistique prend plus de temps du fait de l'analyse variographique.

A Novel Multispectral Images Normalization Application to Unstained Malaria Infected Blood Smear

Solange Doumun^{abc,1}, Sophie Dabo-Niang^a, and Jérémie Zoueu^b

^aUniv, Lille, CNRS, UMR 8524-Laboratoire Paul Painlevé, INRIA-MODAL, F-59000 Lille, France

^bLaboratoire d'Instrumentation, Image et Spectroscopie, Institut National Polytechnique Félix Houphouët-Boigny, BP 1013, Yamoussoukro, Côte d'Ivoire

^cUMR 9221- Lille Économie Management, INRIA-MODAL, F-59000 Lille, France

Abstract:

Multispectral and multimodal unstained blood smear images are acquired and assessed to provide computer-assisted automated diagnostic evidence for malaria disease. Because of acquisition system, these images are non-uniformly illuminated and exhibit contrast variability and local luminosity. This limitation affects the diagnostic process and its outcome. In order to overcome this limitation, a normalization of these multispectral images is an essential pre-processing step in the diagnostic. We introduce here a new method to normalize luminosity and contrast in Multispectral and multimodal images. The method is based on estimation of the Bright reference image which represents luminosity and contrast variability function from the background part of the image. This is accomplished by two different re-sampling methodologies namely gaussian random field simulation by variogram analysis and Bootstrap re-sampling. In order to avoid intensity saturation problem of some pixels, an outlier impute method is also suggested. Both proposed image normalization approaches are shown to outperform state-of-the-art methods of multispectral and multimodal unstained blood smears images in terms of the structural similarity index (SSIM), mean squared error (MSE). They enhance the contrast of the image and preserve its spectral footprint and natural appearance more precisely. The normalization using Bootstrap re-sampling reduces multimodal and

¹ To whom all correspondence should be addressed. Email: mekapeusolange.doumunepseoulai.etu@univ-lille.fr

multispectral images acquisition time by 66%. Beyond speeding up acquisition time, Bootstrap re-sampling, processing time is less than 4% of gaussian random field simulation processing time.

Keywords: Normalization; Multispectral imaging; Gaussian random field; Bootstrap Re-sampling; Malaria

1. Introduction:

Microscopic examination of blood smears is a “gold standard” diagnostic tool widely used to gather important clinical information about malaria patient in the developing world. The precise identification of the malaria parasite, the parasitemia and its staging, are important parameters in diagnostic of this disease. They lead to the therapeutic treatment with suitable drugs. However, it is a time-consuming process and required extensive training. Computer-aided diagnosis methods based on blood smears images are now available to assist biologists in they work.

A simple multispectral and multimodal microscope using a CMOS camera and a set of quasi-monochromatic Light Emitting Diodes (LEDs) ranging from ultraviolet to near-infrared wavelengths as illumination sources was constructed [1]. It allows detecting malaria in blood smear images without using markers for parasites as it is the case of thick and thin blood smears.

Multispectral and multimodal unstained blood smear images are acquired with this LED-based multispectral microscope [2]. It records megapixel images in thirteen spectral bands

from ultraviolet to near-infrared in each mode. Because of the nonuniform light illumination these images, local luminosity and contrast variability are present. Seriously the diagnostic accuracy and its outcome are affected. Images with large luminosity and contrast variability, are very difficult to analyze with such automatic systems and the obtained results may be of poor quality. In this case, the normalization step is a necessary pre-requisite, aimed at obtaining images with a common standardized value for luminosity and contrast.

Several techniques have been used to improve nonuniform luminosity and contrast levels. Especially, by estimation of the variability of contrast and luminosity using the image background and the subsequent compensation of this variability in the whole image [3]; by the sub-histograms equalization [4]; by exploiting the underlying color vector space’s geometry [5], by exploring bivariate histogram of gradient and image intensity [6]. The technique used for the specific application to the multispectral and multimodal blood smear images has previously been described in [1]. Normalization was applied by subtracting the dark reference image ($I_{D\lambda}$) from the sample ($I_{S\lambda}$) and the bright

reference image ($I_{R\lambda}$) reference images. Then the sample image is divided by the bright reference image according to Equation 1.

$$I_{\lambda} = \frac{I_{S\lambda} - I_{D\lambda}}{I_{R\lambda} - I_{D\lambda}}. \quad (1)$$

The main reason that calls for the development of a new and different normalization algorithm is the complexity of the acquisition procedure and the duration of acquisition process. This normalization process requires the acquisition of 3 images (the dark reference image, the bright reference image and sample image) for a grayscale image in the same configuration. Knowing that, a multispectral and multimodal image of a blood smear sample consists of 39 grayscale images, its acquisition takes three times longer to meet this need of normalization. Therefore, the duration of a multispectral and multimodal image acquisition may be several minutes. Usually, 117 images in three different modes are acquired further increasing the overall diagnosis time. To meet the demands of real-time diagnosis, acquisition and pre-processing time must be significantly shortened. In this work, we present a novel normalized method for both luminosity and contrast in multispectral and multimodal images, based on

a reconstruction of the bright reference image using the sample image. The background part of the sample image is estimated by the extraction of red blood cells from the sample image and then is used to reconstruct observed fields and predict pixels values of cell locations. The whole background image is thus reconstructed as a bright reference. Reconstruction is performed using two methodologies. A first methodology based on geostatistical gaussian random field simulation, by variogram analysis using the statistical and spatial features of the sample image without the cells and a second methodology using bootstrap re-sampling algorithm. Finally, only the sample image is acquired for the overall pre-processing, according to Equation 2:

$$I_{\lambda} = \frac{I_{S\lambda}}{I_{R\lambda}}. \quad (2)$$

where

I_{λ} grayscale image in wavelength λ ,

$I_{S\lambda}$ Image of sample,

$I_{R\lambda}$ Reconstructed bright reference image.

Even if it has been developed specifically for multimodal and multispectral images normalization, the proposed techniques can be applied to achieve background reconstruction

by random field simulation to any image where a background portion is present.

2. Materials and methods:

2.1. Material:

The multimodal and multispectral image consists of a set of 39 grayscale images of label-free blood smear have been studied. The images were captured using a multimode, multispectral imaging system developed by our group and has been previously described by [1]. It is a modified commercial metallurgical microscope (Brunel SP80). The ocular was substituted by a 5 Mpx (2592×1944) monochromatic CMOS camera (Guppy-503B, Allied Vision Technology, using a MT9P031 sensor from Micron/ Aptina) equipped with $2.2 \times 2.2 \mu\text{m}$ size individual pixel, each pixel having 12-bits depth. This system is constituted of a set of 13 LEDs lighting arrangement according to 13 wavelengths used to illuminate the sample. The wavelengths are ranged from Ultra-Violet to near-infrared (375 to 935 nm) [7]. The blood smear samples used in this study were provided, by the physicians of regional

hospital center of Yamoussoukro (Ivory Coast), without chemical further.

2.2. Image acquisition:

The samples are the unstained blood smear from sick patient prepared by using a capillary blood. The process of smearing consists in spreading a drop of blood uniformly over a glass slide, in order to obtain a single layer of cells without any staining. Our data are the grayscale images of those samples made in Transmission, Reflection and Scattering. Then the dark field measurements are made straight on the sample, and an empty glass slide is processed as reference images in those same geometries to satisfy the need of normalization pre-pressing step [8]. The images are obtained for each of the three modes from 375 to 940 nm (13 wavelengths).

The same scene is automatically captured at each wavelength and in each geometry through an interface (DAQ) developed on LABVIEW 8.6 which controls the current intensities [9], [7], [10]. The spectral image I_{xy} is represented by a 2D matrix:

$$I_{xy} = \int_0^{+\infty} S_{sy\lambda} R_{sy\lambda} D_{sy\lambda} d\lambda. \quad (3)$$

where

I_{xy} spectral image,

$S_{sy\lambda}$ emission spectrum,

$R_{sy\lambda}$ sample response,

$D_{sy\lambda}$ detector sensibility.

If we consider, the spectral dimension we have a 3D image $I_{xy\lambda}$ for each geometry.

2.3. Pre-processing methodologies:

2.3.1. Existing methodology:

Once all images were acquired and saved, the dark reference image (I_{D_λ}) was subtracted from the sample (I_{S_λ}) and the bright (I_{R_λ}) reference images to remove optical and electronic noise from the image. Then the luminosity and contrast correction were performed by dividing the sample image (Figure 1) by the bright reference image to get the normalized image according to Equation 1.

To obtain 13 spectral images for each mode a total of 39 grayscale images (13 sample images, 13 bright reference images, and 13 dark

reference images) are taken. Since the final multimodal and multispectral image of a blood smear sample consists of 39 grayscale images (13 per mode), it required 117 grayscale images to meet the need for normalization Equation 1. Requiring 2 times extra acquisition time for the additional images. Finally, to remove the impulse noise due to the acquisition system appearing in the image as high intensity pixels, a 2-D median filter was applied to the normalized images (Figure 2). The median filter replaces all the data value in the image using the middle value of the window around each sample point. It can be useful to reduce noise. However, it can unfortunately degrade the image instead of improving by smoothing it. Therefore, it should be used with great caution, in particularly, for an image-based computer-aided diagnosis system.



Figure 1: Sample image in transmission at 435nm

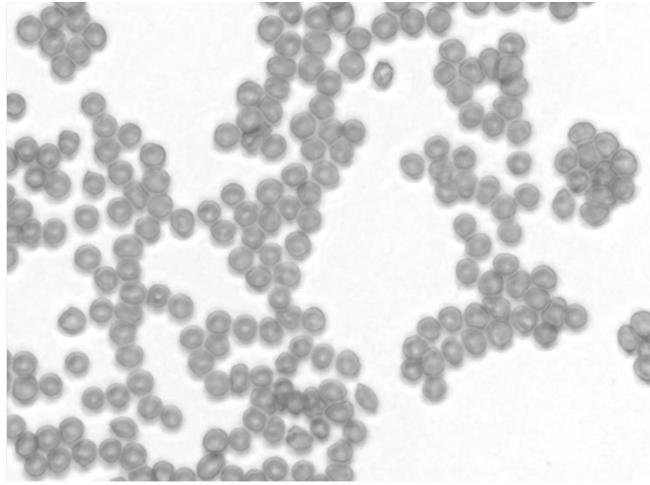
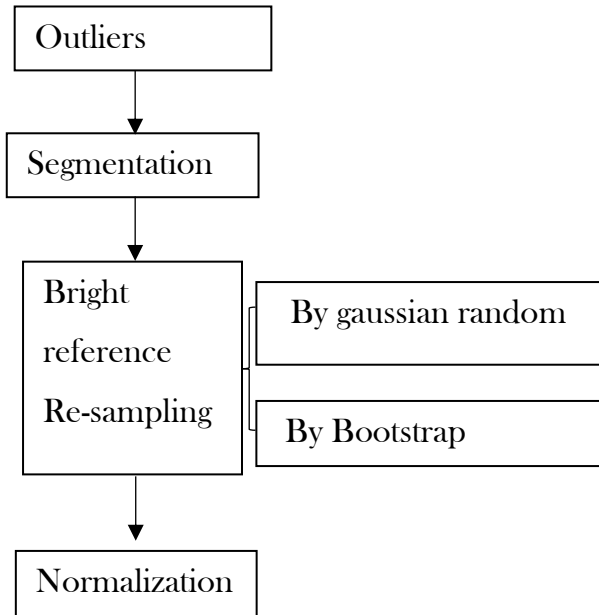


Figure 2: The existing normalized and filtered image from the sample image Fig. 1

2.3.2. Proposed methodologies:

Algorithm



2.3.2.1. Outliers imputing:

The first step in the pre-processing is to address the high intensity pixels in the image. Usually, a median filtering pre-processing is applied on the normalized images. We address those outliers by Hampel’s method [11], [12], [13]. The Hampel outlier detection procedure is robust and performs well in the context of multiple outlier’s data [14] [15]. The process used median \tilde{x} of dataset X of interest, a threshold t and Median Absolute Deviation MAD defined by:

$$MAD = median|x - \tilde{x}|. \tag{4}$$

Let a sample $X = \{x_s\}_{s \in S}$ composed of $N = n * m$ spatial observations indexed in $S = \{s = (s_1, s_2), 1 \leq s_1 \leq m, 1 \leq s_2 \leq n\}$.

An index s is a pixel of a given image. For simplicity, we denote in the following the components of X by x_1, \dots, x_N in an arbitrary order.

An observation x_s is defined as outlier as following:

$$outlier = \begin{cases} 1 & \text{if } |x_s - \tilde{x}| > t * \hat{S} \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

Considering our data which are images, the components x_s are pixels intensity values, \tilde{x}

median of pixels values. In the literature, a moving window is used [16]. It is usual to take the values, $t = 3$ as the best estimation of the threshold and $\hat{S} = 1,4826 * MAD$ [17],[18]. In the state-of-art those outliers are cleaning by imputing the median of the moving window [19] [14] [13] [20] [15]. Since we are concerned by the high pixel intensity values, also to have data closed to our original image data, we chose to impute the outlier values by $\tilde{x} + t * S$ as show below:

$$y_s = \begin{cases} \tilde{x} + t * S & \text{if } x_s - \tilde{x} > t * S \\ x_s & \text{otherwise} \end{cases}$$

2.3.2.2. Sample image segmentation process:

The segmentation mask is designed by Otsu algorithm(Otsu, 1979), applied after a low-pass filtering. We used the filter2 function of the image processing package "EBImage" of R language. This function uses a 2D Fast Fourier Transform which is a convolution product equivalent to the following equation:

$$(w * f)(x, y) = \sum_{s=-\infty}^{+\infty} \sum_{t=-\infty}^{+\infty} w(s, t)f(x + s, y + t) \tag{6}$$

The objective is to remove the cells but also the near-field deformations at the edges of the cells.

This mask is used for segmentation of the images Figures 3 & 4.

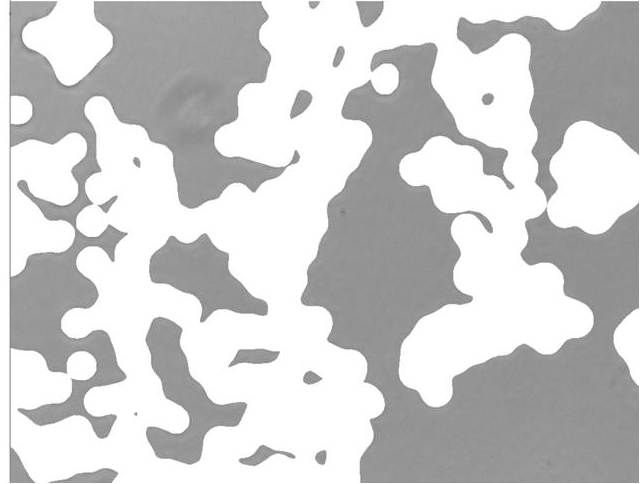


Figure 3: Segmentation mask on the Figure 1

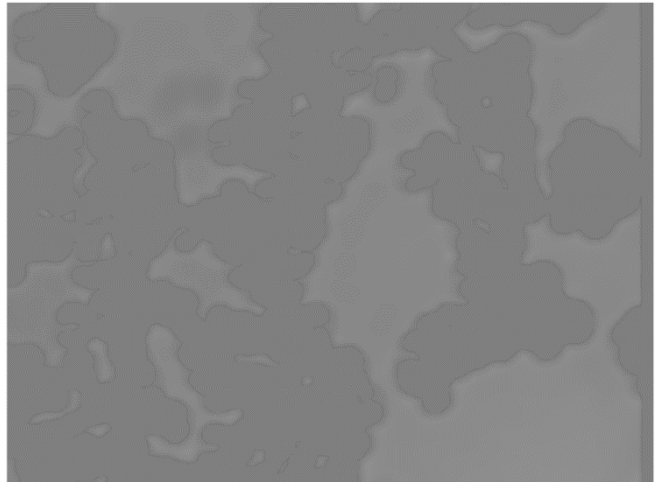


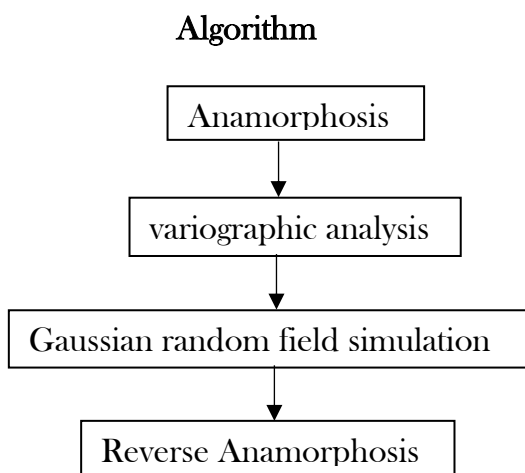
Figure 4: Segmented background from Figure 1 based on Figure 2

2.3.2.3. Bright reference Resampling:

Here, we have developed two algorithms to reconstruct random data. In those algorithms,

instead of imputing the missing pixel values corresponding to the deleted cells (Figure 3), we resample the entire image, by reconstructing observed fields and predicting pixels values of cell locations.

2.3.2.3.1 Geostatistical reconstruction:



Our approach uses the geostatistical techniques of gaussian random field simulation. It takes advantage of the variogram analysis method of kriging theory [21]. Kriging is a spatial stochastic interpolation method that predicts the value of a phenomenon at non-observed sites. It is the best unbiased minimum variance estimator in a gaussian context [22]. In the statistical literature, these methods, initially applied in the mining field to predict deposits, have a very diversified field of application such as: geology,

oceanography, econometrics, epidemiology, renewable energy, environmental sciences, forestry, image processing and many other fields [23], [24], [25], [26], [27], [28].

Geostatistics:

Geostatistics is a branch of spatial statistics that aims to provide a description of spatially or spatio-temporally distributed observations. It is characterized by an approach that derives its originality from the fact that it is based on data. Initially, intended for the study of the evolution of mining deposits, geostatistics became one of the best-known methods of spatial statistics for analyzing spatial autocorrelation. Such phenomenon appears in a variety of fields, including geology, oceanography, soil science, econometrics, image processing, environmental science, forestry, epidemiology and many others, [29], [28] [9]. In geostatistics, observed data and position information are used to improve estimation in a spatial context. This estimation is coupled with an assessment of its uncertainty, which ensures the reliability of the estimator.

Geostatistical modelling consists of two fundamental steps: the identification of the

spatial structure of observations using variographic analysis and prediction, i.e. interpolation by Kriging. As far as we are concerned in this work, we are interested in simulating the background image of a blood smear image using variographic analysis of a gaussian random field (spatial process). Some of those concepts are briefly described below:

As developed by [9] [30], a random field is a stochastic process taking values in Euclidean space and defined over a parameter space of dimensionality at least one.

Let S be a spatial set, that is $S \subset \mathbb{R}^d$, $d > 1$. A random field Z on S is a random spatial function (collection of random variables) $Z = \{Z_s, s \in S\}$. We suppose that the random field is second order stationary. This means, the first moments of Z_s exist and are invariant by translational.

$$E(Z(s+h)) = E(Z(s)) = \mu, \quad (7)$$

$$Cov(Z(s), Z(s+h)) = C(h), \quad (8)$$

$$s, s+h \in S,$$

where $C(\cdot)$, the stationary covariance function is positive definite. If $C(h)$ is isotropic, it can be written as $C(h) = C(\|h\|)$. Otherwise, the field is said anisotropic.

Usually, it is assumed that $Z = (Z_s, s \in S)$ is gaussian and observed at some n fixed locations $\{s_1, s_2, \dots, s_n\} \subset S, Z_s \in \mathbb{R}$.

This gaussian hypothesis is fundamental. However, when the field is not normal a transformation or anamorphosis should be made. In our following methodology, the data of interest (blood smear images) are transformed to follow a normal distribution in order to apply a geostatistical methodology detailed below.

Let be μ and $\sigma \in \mathbb{R}; \sigma > 0$. It is said that a random variable Z_s follows a normal distribution of parameters μ and σ^2 (noted $Z_s \sim \mathcal{N}(\mu; \sigma^2)$) if its probability density f is defined by:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}. \quad (9)$$

With the mean μ of Z_s and its standard deviation σ (and therefore its variance σ^2). A random variable Z_s follows the reduced centered normal distribution of \mathbb{R} (noted $\mathcal{N}(0; 1)$) if its probability density f is defined by:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \quad (10)$$

Anamorphosis [31]:

For a geostatistical simulation, an anamorphic transformation is necessary to switch from any field to a gaussian field. Anamorphosis is a non-linear transformation that modifies the statistical properties of our variable of interest. It allows to express any probability distribution to a standard gaussian distribution. The preliminary approach we use, transforms the image data into data coming from a gaussian variable upstream, then a variogram analysis is performed and we do a simulation method. The obtained simulated gaussian field is re-transformed to obtain a field of the image data. In order to justify the anamorphosis, a normality test is performed on the image data before applying the appropriate anamorphosis equation.

Variogram:

As mentioned above, the main tool for the spatial dependence analysis is the variogram function 2γ defined by:

$$\gamma(h) = \frac{1}{2}Var(Z_{s+h} - Z_s). \tag{11}$$

It is such that:

- $\gamma(h) \geq 0, \gamma(0) = 0,$
 $\gamma(-h) = \gamma(h)$ (12)

It is an even function, defined as positive.

- $\lim_{\|h\| \rightarrow \infty} \frac{\gamma(h)}{\|h\|^2} = 0,$ (13)

- $\gamma(h) = C(0) - C(h),$ (14)

- $\lim_{\|h\| \rightarrow \infty} \gamma(h) = l, l = C(0),$ (15)

When the process of interest is stationary.

- $\gamma(h)$ is conditionally negative definite:
 $\forall s_1, \dots, s_n; \forall \lambda_1, \dots, \lambda_n$ with $\sum_{i=1}^n \lambda_i = 0.$
 $\sum_{i,j} \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0,$ (16)

The semi-variogram is estimated empirically from the data as:

$$\hat{\gamma}_n(h) = \frac{1}{2N(h)} \sum_{s_i, s_j \in N(h)} (Z_{s_i} - Z_{s_j})^2, \tag{17}$$

$h \in \mathbb{R}^d$

Where

$$N(h) = \{(s_i, s_j): s_i - s_j = h; i, j = 1, \dots, n\}$$

and N_h is the number of elements of $N(h)$

The empirical variogram permits to fit a parametric variogram on the data among several parametric models, for instance:

- Exponential:
 $\gamma(h) = C \left(1 - \exp \left(-\frac{\|h\|}{a} \right) \right),$ (18)

$$C > 0, a > 0,$$

- Spherical:

$$\gamma(h) = \begin{cases} C \left(\frac{3}{2} \frac{\|h\|}{a} + \frac{1}{2} \frac{\|h\|^3}{a^3} \right) & \text{if } 0 \leq \|h\| \leq a \\ C & \text{if } \|h\| \geq a \end{cases} \quad (20)$$

- Gaussian:

$$\gamma(h) = C \left(1 - \exp \left(-\frac{\|h\|^2}{a^2} \right) \right), \quad (21)$$

$$C > 0, a > 0.$$

- Power:

$$\gamma(h) = C(\|h\|^a); a < 2, \quad (22)$$

- Matern:

$$\gamma(h) = C \left[1 - \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{2\nu^{\frac{1}{2}} \|h\|}{\rho} \right) K_{\nu} \left(\frac{2\nu^{\frac{1}{2}} \|h\|}{\rho} \right) \right]. \quad (23)$$

where K_{ν} is a Bessel function of the second kind of order ν .

The three fundamental parameters of the variogram are nugget, sill and range presented in the Figure 5. The nugget noted $C_0 = \lim_{h \rightarrow 0} \gamma(h)$ the discontinuity at the origin which reflects an irregularity of process to local variations as a result of the presence of error terms.

The sill variance $C_1 = \lim_{h \rightarrow \infty} \gamma(h)$ is the value of the tray: the estimated maximum of variance to

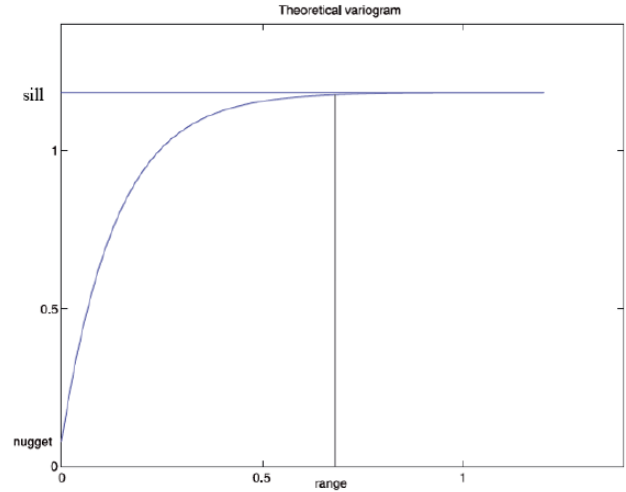


Figure 5: Example of variogram with Sill, Nugget and range Parameters (Dabo-Niang and Zoueu, 2012)

which the variogram curve increases and remains thereafter. The distance from which it is reached is called range a . It indicates the limit of spatial dependence, that is, two points with distance greater than the range are no longer correlated [9].

$$\|h\| \geq a \Rightarrow C_1 - \gamma(h) \leq \varepsilon. \quad (24)$$

In accordance with the chosen parametric model, several inferences methods can be used to estimate the parameters of the variogram, based on the underlying process observations $(Z_{s_i})_{i=1, \dots, n}$. We can refer to, as generalized least squares, maximum likelihood or the ordinary

least squares approximation. The least squares approximation of $\hat{\theta}$ is the solution of the problem:

$$\min_{\theta} \sum_{k=1}^K \left(\gamma_{\theta}(h_k) - \hat{\theta}(h_k) \right)^2. \quad (25)$$

Based on the observed blood smear image data, the choice of a parametric variogram model $\gamma_{\hat{\theta}}$ help to simulate in the following section, at a given location an observation of a gaussian random field with variogram $\gamma_{\hat{\theta}}$.

Spatial Simulation:

The spatial simulation used is the generation of data from a stationary gaussian random field with known mean and variogram function [32]. This produces a smoother reconstruction of the field at locations where data are not available. Unconditional simulation method is used. It aims to reproduce fields that have the same spatial structure (variogram) and histogram as those inferred from the observed data. The simulation is then performed using the RFSimulate method of the Random field package of R; see [33]. After this simulation, our grayscale image can be normalized according to Equation 2.

2.3.2.3.2. Simulation by Bootstrap re-sampling:

Bootstrapping is a statistical technique resampling that employs the method of Monte Carlo applied to observed data. This computer-intensive technique is usually used to determine a summary statistic of the sampling distribution such as median, mean and standard deviation or relationship (regression coefficient and correlation) even if these sampling distributions are extremely difficult to obtain analytically [34]. The estimation of standard error and bias for any statistical parameter of interest can be generated using Bootstrapping [35], [36]. For spatial analysis in geostatistical (i.e. autocorrelation, clustering...) confidence intervals as well as statistical significance are calculated based on bootstrapping methodology.

The bootstrap technique can be applied in two main ways. For the parametric method, a knowledge of the data's distributional form is required (e.g., Gaussian and exponential). A greater precision is provided using such distributional information when it is available. In the nonparametric version of bootstrapping, which is more common, no distributional

assumption is needed. According to this methodology, an estimate of the statistical distribution of the sample or the relationship of interest is calculated from the bootstrap samples. These samples are formed by random drawing with replacement of a series of random samples based on the parameters of the original sample.

Nonparametric bootstrapping:

Following Efron [37], we apply the nonparametric bootstrapping to simulate all pixels values in the image. The observed data which is constituted of the background image pixels values, is randomly sampled, and new datasets (image with same size as the original image) are constructed from the events drawn (and replaced). The process of bootstrapping involves drawing random samples from the original dataset. Here's how it works: The bootstrap method has an equal probability of randomly drawing each original data point for inclusion in the resampled datasets. By the property of "with replacement" aspect of the process, the procedure can select a data point more than once for a resampled dataset. Based on the observed data, the procedure creates

resampled datasets that are the same size as the original dataset. In this statistical modeling the independence between pixels is assumed. The simulated random field is uniformly distributed (uniform contrast).

Simulation with Monte Carlo method:

Monte Carlo Simulation is a powerful mathematical technique that used to generate random variables for modeling processes in a system such as define potential risk or uncertainty in finance. Those processes are performed using the theory of probability in order to obtain an estimated solution to the problem. Monte Carlo method was originally developed in order to study nuclear reactions and is now used in various domains [32] [28]. It can be applied to address any mathematical model or problem that is very complex, time consuming, or resource intensive to solve analytically. Instead of tackling the numerical problem directly, Monte Carlo allows the researcher to obtain an approximation of the solution through setting up an experiment of statistical sampling. To model a large number of random variables this method can be very useful

unlike the ordinary methods whose analysis can be much more complex.

As the name indicates, the method borrows from games of chance such as those played at the famous casinos of Monte Carlo in Monaco. The Monte Carlo method relies on realizations from a probability density function. Ideally, to correctly apply the Monte Carlo method and obtain valid results, the sampling method employed should be completely random. The number of realizations must be sufficiently large to accurately represent the distribution of the input variables. The Monte Carlo method is both interesting and useful. Application areas seem limited only by imagination and computing power.

Based on this principle of bootstrapping, we have developed an algorithm to resample random data. In this algorithm, instead of imputing the missing pixel values corresponding to the deleted cells (Figure 3), we reconstruct the entire image based on the statistical bootstrap method according to the principle of the Monte Carlo algorithm. This resampling technique based on stochastic simulation. It does not require any other information than that available

on the individuals (pixels) of the original sample (image).

Let Z be the dataset from our image (Figure 1), it is assumed to follow some distribution f within a population p (pixel's distribution). Let N denote the entire population size (the image pixels total number). We extract from the image the cells. The cells' locations are the missing data, and the background pixels values obtained are considered to be the observed data of Z (with size k): $z_1, z_2, \dots, z_k, k < N$ (Figure 3). Our goal then is to reconstruct the total background pixels using the observed pixels values. i.e., reconstruct the observed data and predict the missing data values.

3. Results and discussion:

The suggested approach is investigated using the 16 bits grayscale images. The resampling is performed on a core i5 machine running with 8GB RAM on Windows 10 using R software. In order to better appreciate the results obtained with the proposed technique, results on both geostatistical re-sampling and Bootstrap re-sampling are reported. Using the sample original images layer, each test image was segmented into background and the

foreground image (with is constituted of the blood cells) and the whole background image is reconstituted using the observed pixels values in the background image and then the normalized image is obtained by dividing the sample original image by the reconstituted background. The results are presented first for the

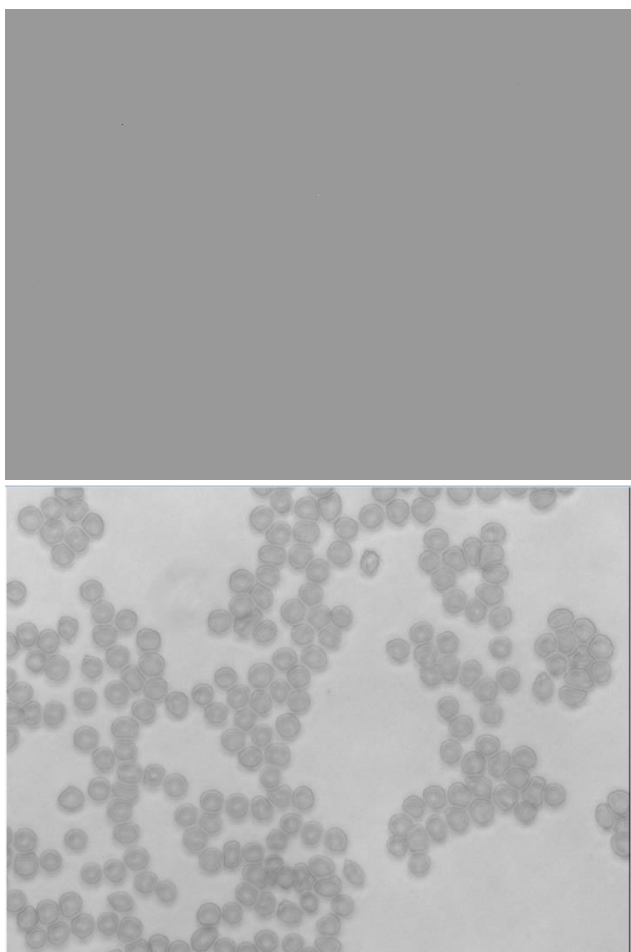


Figure 6: Geostatistical Computed images. *a(top)* – reconstituted bright reference image from segmented image Fig 3; *b(bottom)*- Normalized image

geostatistical re-sampling and then for the Bootstrap re-sampling.

3.1. Geostatistical re-sampling

Figure 6 at the top is the reconstructed bright reference image based on the background image threshold (Figure 3). In fact, based on the observed background pixel values, we computed the spatial dependency by variographic analysis using autofitVariogram method of automap package. Geostatistical simulation (using RFSimultate method from the Random field package) has then been apply to the spatial structure parameters of the variogram i.e. nugget, sill, range and variogram model.

The normalized image (Figure 6- bottom) is obtained by dividing the original sample image (Figure 1) by the reconstructed bright reference image (Figure 6-top).

After Normalizing, we estimated the quality of the normalized image using the image assessment parameters i.e. MSE (Mean Squared Error), PSNR (Peak Signal to Noise Ratio)[38], entropy and structural similarity index (SSIM) (Table 1) [39], [40]. Those Image Quality Assessment parameters are defined by:

$$SSSI = \frac{4\sigma_{xy} \bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)} \quad (26)$$

where

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i,$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i,$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2,$$

$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2,$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}).$$

N denotes the total number of pixels,

x_i and y_i are the gray level of pixel i in images

X and Y respectively. Let the

$$entropy = -\sum_{i=1}^N p_i \log_{10} p_i \quad (27)$$

p_i is the probability associated with gray level i .

$$PSNR = 10 \log_{10} \frac{1}{\sqrt{MSE}} \quad (28)$$

Let us use the spatial pixel index $s = (s_1, s_2)$,

$$MSE = \frac{1}{mn} \sum_{s_1=1}^m \sum_{s_2=1}^n (x_{s_1, s_2} - y_{s_1, s_2})^2 \quad (29)$$

where m and n are respectively the rows and columns number of the input image. The MSE

values are calculated on the observed pixel values by comparing the two normalized images

The improvement can also be appreciated by observing entropy and the SSIM values in Table

1. The lowers values for the normalized

reference image shows that there is a lost information. This can be explained by the median filtering process on the whole image which has a smoothing effect resulting in a heavily reduced contrast. The effect of this can also see in Figure 7. This represents the profile curve along a horizontal line of the images.

This graph gives us a chance to see how the

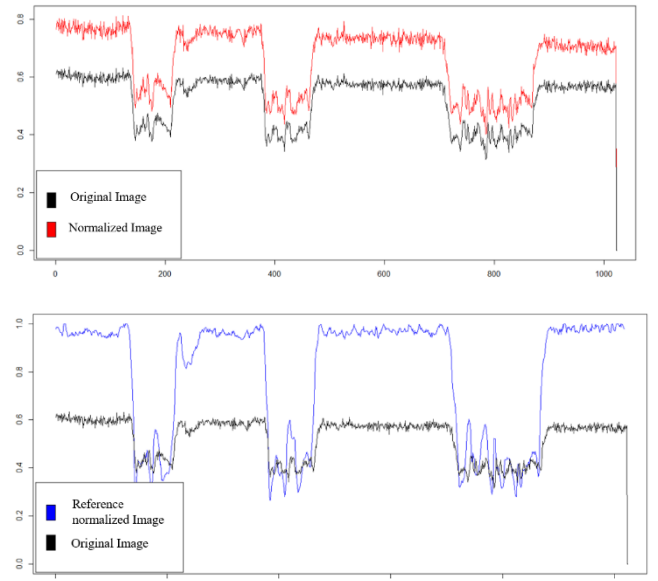


Figure 7: Comparison of greyscale intensity profiles; Top: Original image VS proposed normalized image; Bottom: Original image VS Usual normalized image

proposed normalization procedure images are most closely match the original data (Figure 7-top). i.e., Usual technique normalized using (Equation 1), normalized using reconstructed

bright reference (Equation 2) and the original image (Figure 1).

Table 1: Image Quality Assessment for geostatistical process

	Reference Normalized image	Geostatistical based normalized image
MSE	0.125744	0.01040432
PSNR	9.005127	19.82786
Entropy	13.5453	13.55848
SSIM	0.9861848	0.9996384

One can note the normalized using reconstructed bright reference image markedly exhibits the visual quality and the contrast with more details (information) in the image. In fact, it can be observed in Table 1, that the MSE of our normalized image (Equation. 2) is lower than that of the normalized reference image (Equation 1) and the PSNR is higher. This shows that our image has better quality, according to [38]. In contrary, the reference normalized image (Equation 1) has the variations which are not conform to the shape of the profile curve of the original image. This

confirms the loss of information mentioned. This loss is also underlined by the similarity indices and the entropy. On the contrary, high intensity pixels values imputation method is not affect the whole image.

3.2. Bootstrap re-sampling:

For this Bootstrap re-sampling process, we assume, the background observed pixel intensities are iid (independent identically distributed) random variables. The observed pixels in Figure 3 have been re-sampling by bootstrap according to Monte Carlo's algorithm (Figure 8-top). The reconstructed data have been obtained by re-sampling with replacement. The reconstructed bright reference image is then used for normalization of the sample original image. The bootstrap based normalized image can be seen in bottom of Figure 8.

We can notice that the bootstrap normalization procedure significantly improves contrast and appears to most closed to the original data according to Table 2. Its MSE is lowest and the PSNR is highest. The improvement can also be appreciated by observing the values of entropy and SSIM. In fact, the high values for the Bootstrap re-sampling-based normalization,

explains the preservation of information details using this normalization process. Furthermore, as meant above, the usual normalization process (Equation 1) as well as the applied median filtering deteriorate the information contained in the image.

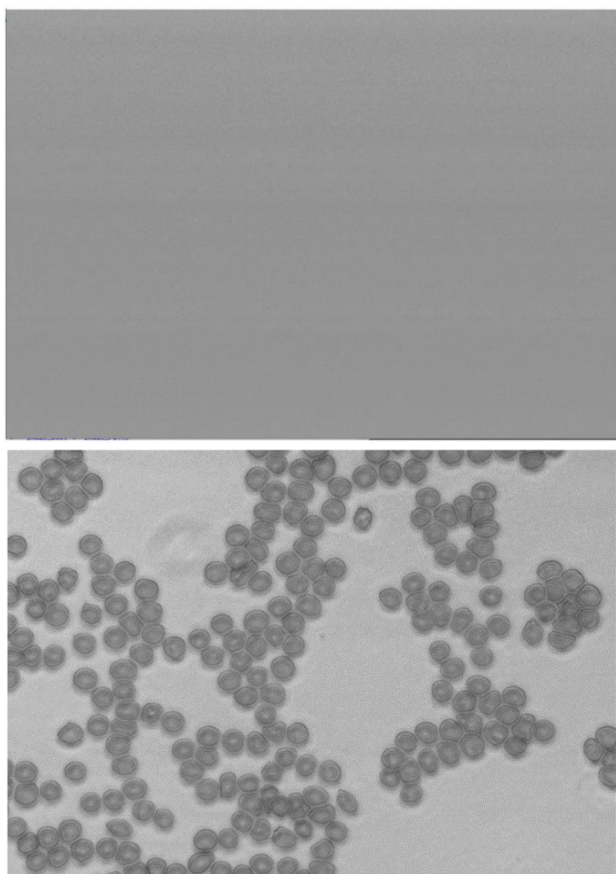


Figure 8: Bootstrap Computed images. top – Bootstrap re-sampling reconstituted bright reference image from segmented image fig 4; bottom-- Bootstrap based Normalized image

Table 2: Image Quality Assessment for bootstrap process

	Reference Normalized image	Bootstrap re-sampling based normalized image
MSE	0,125744	0,008041
PSNR	9,005127	20,94697
Entropy	13,5453	13,5576
SSIM	0,986185	0,999822

The stability assessment of Bootstrap method is important in this methodology. Bootstrap stability analysis was carried out using the Mean Squared Error (MSE) and Standard Deviation (SD) on the observed pixel values. The Table 3 shows the values of the MSE and the SD for first, fiftieth, hundredth and for five hundred processing. By analysis this table, it can be seen the stability of our bootstrap processing by decreasing of MSE values and observing the SD values are negligible. These parameters confirm the stability of our method.

Table 3: stability assessment of Bootstrap

Number of Bootstrap	1	50	100	500
MSE (sd)	7,429.10 ⁻⁴	7,412.10 ⁻⁴ (1,677.10 ⁻⁶)	7,4.10 ⁻⁴ (1,583.10 ⁻⁶)	7,393.10 ⁻⁴ (1,596.10 ⁻⁶)

3.3. Geostatistical re-sampling vs Bootstrap re-sampling:

The follow Table 4 compares ours two reconstructed images. The analysis of this table shows us that the geostatistical method has a good predictive capability and preserve the spatial dependence of the observed pixels of the background image. It can be seen that normalization using geostatistical reconstruction improves the normalized image, since the value of MSE is very low. However, it is time-consuming time compared to the Bootstrap reconstruction. The Bootstrap reconstruction

has the advantage of being stable, easy to set up, very fast and having a very low MSE. For the above reasons, the Bootstrap based normalization process is recommended in order to reduce computational burden and save diagnosis time.

Table 4: Methodologies comparison

Methodologies	MSE	Entropy	SSIM	Processing time	Other Advantage
Geostatistical Re-sampling	1,04.10 ⁻²	13,55848	0,9996	8 mn	preserve the spatial dependence
Bootstrap re-sampling	8,04.10 ⁻³	13,5576	0.9998	20 s	stability

3.4. Bootstrap re-sampling assessment:

This last step aims at assessing the bootstrap-based normalization images (Equation 2) on our multispectral and multimodal image compared to reference normalization (Equation 1). The table below compares the values of the image quality assessment parameters between these two types of normalized images. Knowing that it is made up of 39 grayscale images acquired with 13 wavelengths in 3 geometries. We focus on the 13 wavelengths in transmission mode for the purpose of this evaluation. By analyzing the summary table of image quality assessment metrics Table 5 i.e., SSIM, Entropy, MSE and PNSR, we can note that the proposed normalization procedure improves images luminosity and contrast with better results than the usual normalization procedure. The improvement can be appreciated by observing the results SSIM and Entropy values in Table 5 with are higher for the proposed normalization procedure comparing with the usual normalization procedure values. The lower values of MSE for bootstrap-based normalization shows us that our approach preserve more closely the original image

information. This can explain the high value of the PSNR in Table 5.

By comparing the two normalized images i.e., Bootstrap based normalized image (Figure 8-bottom) and that the usual normalized image (Figure 2), one can note that the usual normalized image exhibits a smoothing effect resulting in a heavily reduced contrast. This can be explained by the median filtering which affect the image. On the contrary, our normalization process is not significantly affected by the outliers imputing procedure. It does not smooth the normalized image. Ultimately, our method improves image quality and preserves information optimally compared to conventional method.

4. Conclusion:

The algorithm proposed here mainly originated from our need of obtaining normalized images, to be processed for the automatic and rapid diagnosis of malaria using multispectral and multimodal images of unstained red blood smears. Therefore, the preservation of information and the improvement of the resulted image are important without losing sight of the ultimate

objective of time saving. As such, the proposed approach enhances the contrast of those images and preserves the images information, in order to facilitate there, the future processing. This was based on simulations of the bright reference image using the sample image. Two methods were developed, the geostatistical random field simulation using the variogram analysis and bootstrap resampling based on Monte Carlo method. Both approaches are shown to outperform state-of-the-art method for multispectral and multimodal unstained blood smears images in terms of the structural similarity index (SSIM), the mean squared error (MSE). Mainly the bootstrap-based normalization, which speeds up the

normalization time (18 s) is recommended. This may save time for diagnosis and treatment.

In the perspective, we propose to exploit the fast-spatial bootstrap for taking in to account the spatial dependency of pixel's distribution in the background image.

Acknowledgements:

EMS-Simons for Africa and MOBLILEX scholarships of Lille university are acknowledged for financial support.

		SSIM		Entropy		MSE		PSNR	
	wavelengt h	Reference Normalize d image	Bootstrap Normalize d	Reference Normalize d image	Bootstrap Normalized	Reference Normalize d image	Bootstrap Normalize d	Reference Normalized image	Bootstrap Normalize d
image 1	375 mn	0,9792	0,9980	13,55	13,56	0,1782	0,0206	7,49	16,86
	400 mn	0,9855	0,9999	13,54	13,56	0,1344	0,0075	8,72	21,26
	435 mn	0,9862	0,9998	13,55	13,56	0,1257	0,0080	9,01	20,95
	470 mn	0,9780	0,9999	13,55	13,56	0,1864	0,0058	7,30	22,34
	525 mn	0,9746	0,9994	13,55	13,56	0,2177	0,0099	6,62	20,05
	590 mn	0,9848	0,9994	13,54	13,56	0,1392	0,0131	8,56	18,83
	625 mn	0,9750	0,9979	13,55	13,56	0,2221	0,0258	6,53	15,88
	660 mn	0,9660	0,9992	13,55	13,56	0,2961	0,0112	5,28	19,51
	700 mn	0,9771	0,9998	13,55	13,56	0,2018	0,0068	6,95	21,67
	750 mn	0,9715	0,9997	13,55	13,55	0,2492	0,0095	6,03	20,22
	810 mn	0,9755	0,9997	13,55	13,56	0,2171	0,0081	6,63	20,91
	850 mn	0,9552	0,9996	13,55	13,56	0,3907	0,0090	4,08	20,48
940 mn	0,9526	0,9998	13,55	13,56	0,4131	0,0087	3,84	20,59	
image 2	375 mn	0,9792	0,9998	13,55	13,56	0,1838	0,0066	7,36	21,78
	400 mn	0,9862	0,9996	13,53	13,56	0,1357	0,0071	8,68	21,47
	435 mn	0,9867	0,9985	13,54	13,56	0,1304	0,0150	8,85	18,25
	470 mn	0,9786	0,9999	13,55	13,56	0,1844	0,0054	7,34	22,64
	525 mn	0,9735	0,9998	13,55	13,56	0,2248	0,0069	6,48	21,61
	590 mn	0,9819	0,9983	13,55	13,55	0,1538	0,0190	8,13	17,22
	625 mn	0,9706	0,9988	13,55	13,55	0,2433	0,0159	6,14	17,97
	660 mn	0,9657	0,9999	13,55	13,56	0,2920	0,0062	5,35	22,08
	700 mn	0,9791	0,9999	13,55	13,56	0,1813	0,0045	7,42	23,48
	750 mn	0,9753	0,9981	13,55	13,56	0,2145	0,0161	6,68	17,94
	810 mn	0,9776	0,9994	13,55	13,57	0,1977	0,0071	7,04	21,51
	850 mn	0,9622	0,9997	13,55	13,57	0,3278	0,0063	4,84	22,02
940 mn	0,9618	0,9998	13,55	13,56	0,3335	0,0064	4,77	21,91	
Imag e3	375 mn	0,9763	0,9999	13,54	13,56	0,2113	0,0077	6,75	21,13
	400 mn	0,9847	0,9981	13,53	13,56	0,1493	0,0166	8,26	17,81

	435 mn	0,9861	0,9994	13,53	13,56	0,1375	0,0094	8,62	20,28
	470 mn	0,9775	0,9999	13,55	13,56	0,1984	0,0064	7,02	21,95
	525 mn	0,9748	0,9998	13,55	13,56	0,2196	0,0077	6,58	21,14
	590 mn	0,9799	0,9996	13,55	13,55	0,1742	0,0121	7,59	19,18
	625 mn	0,9695	0,9974	13,55	13,55	0,2625	0,0245	5,81	16,10
	660 mn	0,9586	0,9967	13,55	13,56	0,3579	0,0261	4,46	15,84
	700 mn	0,9763	0,9968	13,55	13,56	0,2074	0,0253	6,83	15,97
	750 mn	0,9716	0,9980	13,55	13,56	0,2481	0,0173	6,05	17,62
	810 mn	0,9758	0,9997	13,55	13,56	0,2147	0,0063	6,68	22,00
	850 mn	0,9623	0,9998	13,55	13,56	0,3301	0,0058	4,81	22,39
	940 mn	0,9598	0,9992	13,55	13,56	0,3539	0,0104	4,51	19,83
	375 mn	0,9808	0,9999	13,54	13,56	0,1674	0,0079	7,76	21,01
	400 mn	0,9896	0,9981	13,53	13,56	0,1105	0,0174	9,57	17,60
	435 mn	0,9898	0,9980	13,53	13,55	0,1061	0,0191	9,74	17,19
	470 mn	0,9805	0,9997	13,55	13,56	0,1691	0,0077	7,72	21,15
	525 mn	0,9747	0,9997	13,55	13,56	0,2159	0,0081	6,66	20,91
	590 mn	0,9825	0,9987	13,55	13,55	0,1513	0,0178	8,20	17,49
Imag	625 mn	0,9756	0,9983	13,55	13,55	0,2096	0,0196	6,79	17,08
e4	660 mn	0,9736	0,9996	13,55	13,56	0,2265	0,0074	6,45	21,30
	700 mn	0,9819	0,9998	13,55	13,56	0,1602	0,0054	7,95	22,67
	750 mn	0,9790	0,9987	13,55	13,56	0,1861	0,0129	7,30	18,89
	810 mn	0,9808	0,9999	13,54	13,56	0,1745	0,0055	7,58	22,56
	850 mn	0,9726	0,9997	13,55	13,56	0,2418	0,0061	6,16	22,12
	940 mn	0,9717	0,9983	13,55	13,56	0,2470	0,0159	6,07	17,97
	375 mn	0,9779	0,9996	13,55	13,56	0,1819	0,0079	7,40	21,01
	400 mn	0,9819	0,9986	13,55	13,56	0,1530	0,0129	8,15	18,88
	435 mn	0,9829	0,9990	13,55	13,56	0,1465	0,0106	8,34	19,75
image	470 mn	0,9787	0,9996	13,55	13,56	0,1788	0,0063	7,48	22,04
5	525 mn	0,9752	0,9999	13,55	13,56	0,2073	0,0050	6,83	22,98
	590 mn	0,9790	0,9999	13,55	13,56	0,1773	0,0050	7,51	23,05
	625 mn	0,9761	0,9996	13,55	13,56	0,2010	0,0071	6,97	21,46
	660 mn	0,9764	0,9999	13,55	13,57	0,1980	0,0041	7,03	23,84

	700 nm	0,9782	0,9999	13,55	13,57	0,1849	0,0038	7,33	24,26
	750 nm	0,9759	0,9993	13,55	13,57	0,2055	0,0076	6,87	21,18
	810 nm	0,9765	0,9993	13,55	13,57	0,2023	0,0079	6,94	21,03
	850 nm	0,9722	0,9999	13,55	13,57	0,2378	0,0044	6,24	23,55
	940 nm	0,9707	0,9991	13,55	13,56	0,2512	0,0093	6,00	20,33
Image e6	375 nm	0,9793	0,9999	13,55	13,57	0,1698	0,0037	7,70	24,36
	400 nm	0,9820	0,9993	13,55	13,56	0,1516	0,0091	8,19	20,42
	435 nm	0,9821	0,9988	13,55	13,56	0,1511	0,0126	8,21	18,98
	470 nm	0,9796	0,9999	13,55	13,56	0,1704	0,0060	7,69	22,24
	525 nm	0,9803	0,9999	13,55	13,56	0,1659	0,0054	7,80	22,71
	590 nm	0,9812	0,9997	13,55	13,57	0,1617	0,0050	7,91	23,04
	625 nm	0,9793	0,9999	13,55	13,56	0,1791	0,0047	7,47	23,29
	660 nm	0,9781	0,9996	13,55	13,57	0,1879	0,0057	7,26	22,47
	700 nm	0,9777	0,9989	13,55	13,56	0,1925	0,0106	7,16	19,73
	750 nm	0,9745	0,9986	13,55	13,56	0,2192	0,0131	6,59	18,83
	810 nm	0,9752	0,9995	13,55	13,56	0,2151	0,0076	6,67	21,18
	850 nm	0,9712	0,9998	13,55	13,56	0,2467	0,0067	6,08	21,75
	940 nm	0,9713	0,9990	13,55	13,56	0,2457	0,0107	6,10	19,70
	image 7	375 nm	0,9851	0,9998	13,54	13,55	0,1408	0,0099	8,51
400 nm		0,9919	0,9986	13,53	13,56	0,0943	0,0144	10,26	18,42
435 nm		0,9911	0,9977	13,54	13,56	0,0933	0,0204	10,30	16,90
470 nm		0,9812	0,9984	13,55	13,56	0,1604	0,0157	7,95	18,03
525 nm		0,9779	0,9992	13,55	13,56	0,1873	0,0102	7,28	19,92
590 nm		0,9766	0,9990	13,55	13,54	0,2021	0,0191	6,94	17,20
625 nm		0,9737	0,9996	13,55	13,55	0,2266	0,0128	6,45	18,93
660 nm		0,9717	0,9999	13,55	13,56	0,2445	0,0055	6,12	22,62
700 nm		0,9775	0,9997	13,55	13,56	0,1988	0,0065	7,02	21,89
750 nm		0,9749	0,9997	13,55	13,56	0,2209	0,0073	6,56	21,36
810 nm		0,9770	0,9993	13,55	13,56	0,2043	0,0097	6,90	20,14
850 nm		0,9683	0,9996	13,55	13,56	0,2759	0,0083	5,59	20,80
940 nm		0,9682	0,9986	13,55	13,56	0,2795	0,0146	5,54	18,35
		375 nm	0,9708	0,9998	13,55	13,56	0,2410	0,0054	6,18

	400 nm	0,9816	0,9988	13,54	13,56	0,1654	0,0113	7,82	19,48
	435 nm	0,9830	0,9989	13,54	13,56	0,1528	0,0118	8,16	19,30
	470 nm	0,9722	0,9999	13,55	13,57	0,2337	0,0052	6,31	22,85
	525 nm	0,9655	0,9998	13,56	13,57	0,2902	0,0055	5,37	22,58
	590 nm	0,9773	0,9988	13,55	13,56	0,1856	0,0133	7,31	18,77
image	625 nm	0,9619	0,9997	13,56	13,56	0,3223	0,0085	4,92	20,69
8	660 nm	0,9524	0,9999	13,56	13,57	0,4091	0,0041	3,88	23,88
	700 nm	0,9750	0,9999	13,55	13,57	0,2068	0,0039	6,85	24,12
	750 nm	0,9695	0,9993	13,55	13,57	0,2604	0,0077	5,84	21,16
	810 nm	0,9742	0,9999	13,55	13,57	0,2226	0,0040	6,52	23,94
	850 nm	0,9433	0,9998	13,55	13,57	0,4970	0,0057	3,04	22,46
	940 nm	0,9341	0,9988	13,55	13,56	0,5861	0,0116	2,32	19,37

- [1] M. Brydegaard, A. Merdasa, H. Jayaweera, J. Ålebring, and S. Svanberg, "Versatile multispectral microscope based on light emitting diodes," *Rev. Sci. Instrum.*, vol. 82, no. 12, p. 123106, Dec. 2011, doi: 10.1063/1.3660810.
- [2] J. T. Zoueu and S. T. Zan, "Trophozoite stage infected erythrocyte contents analysis by use of spectral imaging LED microscope.," *J. Microsc.*, vol. 245, no. 1, pp. 90–99, 2012, doi: 10.1111/j.1365-2818.2011.03548.x.
- [3] M. Foracchia, E. Grisan, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Med. Image Anal.*, vol. 9, no. 3, pp. 179–190, Jun. 2005, doi: 10.1016/j.media.2004.07.001.
- [4] M. F. Khan, E. Khan, and Z. A. Abbasi, "Image contrast enhancement using normalized histogram equalization," *Optik*, vol. 126, no. 24, pp. 4868–4875, Dec. 2015, doi: 10.1016/j.ijleo.2015.09.161.
- [5] A. Gupta *et al.*, "GCTI-SN: Geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images," *Med. Image Anal.*, vol. 65, p. 101788, Oct. 2020, doi: 10.1016/j.media.2020.101788.
- [6] I.-M. Sintorn, L. Bischof, P. Jackway, S. Haggarty, and M. Buckley, "Gradient based intensity normalization," *J. Microsc.*, vol. 240, no. 3, pp. 249–258, Dec. 2010, doi: 10.1111/j.1365-2818.2010.03415.x.
- [7] A. J. Merdasa, M. Brydegaard, S. Svanberg, and J. T. Zoueu, "Staining-free malaria diagnostics by multispectral and multimodality light-emitting-diode microscopy," *J. Biomed. Opt.*, vol. 18, no. 3, p. 036002, Mar. 2013, doi: 10.1117/1.JBO.18.3.036002.
- [8] O. K. Bagui and J. T. Zoueu, "Red Blood Cells Counting by Circular Hough Transform Using Multispectral Images," *Science Alert*, 2014. <https://scialert.net/fulltext/?doi=jas.2014.3591.3594> (accessed Jan. 09, 2020).
- [9] S. Dabo-Niang and J. T. Zoueu, "Combining kriging, multispectral and multimodal microscopy to resolve malaria-infected erythrocyte contents: KRIGING, MULTISPECTRAL AND MULTIMODAL MICROSCOPY TO RESOLVE MALARIA-INFECTED ERYTHROCYTE CONTENTS," *J. Microsc.*, vol. 247, no. 3, pp. 240–251, Sep. 2012, doi: 10.1111/j.1365-2818.2012.03637.x.
- [10] S. Doumun, S. Dabo, and J. Zoueu, "Detection and segmentation of erythrocytes in multispectral label-free blood smear images for automatic cell counting," *J. Spectr. Imaging*, vol. 9, Sep. 2020, doi: 10.1255/jsi.2020.a10.
- [11] F. R. Hampel, "A General Qualitative Definition of Robustness," *Ann. Math. Stat.*, vol. 42, no. 6, pp. 1887–1896, Dec. 1971, doi: 10.1214/aoms/1177693054.
- [12] L. Davies and U. Gather, "The Identification of Multiple Outliers," *J. Am. Stat. Assoc.*, vol. 88, no. 423, pp. 782–792, Sep. 1993, doi: 10.1080/01621459.1993.10476339.
- [13] S. Bhowmik, B. Jelfs, S. P. Arjunan, and D. K. Kumar, "Outlier removal in facial surface electromyography through Hampel filtering technique," in *2017 IEEE Life Sciences Conference (LSC)*, Dec. 2017, pp. 258–261. doi: 10.1109/LSC.2017.8268192.
- [14] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj, "Generalized Hampel Filters," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, p. 87, Aug. 2016, doi: 10.1186/s13634-016-0383-6.
- [15] C.-H. Park and J.-H. Chang, "Robust Shrinkage Range Estimation Algorithms Based on Hampel and Skipped Filters," *Wireless Communications and Mobile Computing*, Jan. 01, 2019. <https://www.hindawi.com/journals/wcmc/2019/6592406/> (accessed Jun. 14, 2020).
- [16] R. K. Pearson, "Outliers in Process Modeling and Identification," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 1, pp. 55–63, Jan. 2002, doi: 10.1109/87.974338.
- [17] R. K. Pearson, "Data cleaning for dynamic modeling and control," in *1999 European Control Conference (ECC)*, Aug. 1999, pp. 2584–2589. doi: 10.23919/ECC.1999.7099714.
- [18] R. K. Pearson, "Exploring process data," *J. Process Control*, vol. 11, no. 2, pp. 179–194, Apr. 2001, doi: 10.1016/S0959-1524(00)00046-9.
- [19] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj, "The class of generalized hampel filters," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug. 2015, pp. 2501–2505. doi: 10.1109/EUSIPCO.2015.7362835.
- [20] F. A. Ghaleb, M. B. Kamat, M. Salleh, M. F. Rohani, and S. Abd Razak, "Two-stage motion artefact reduction algorithm for electrocardiogram using weighted adaptive noise cancelling and recursive Hampel filter," *PLOS ONE*, vol. 13, no. 11, p. e0207176, Nov. 2018, doi: 10.1371/journal.pone.0207176.
- [21] T. D. Pham, "Geostatistical Simulation of Medical Images for Data Augmentation in Deep Learning," *IEEE Access*, vol. 7, pp. 68752–68763, 2019, doi: 10.1109/ACCESS.2019.2919678.
- [22] N. Cressie, "Geostatistics," *Am. Stat.*, vol. 43, no. 4, pp. 197–202, 1989, doi: 10.1080/00031305.1989.10475658.
- [23] T. D. Pham, "Geostatistical Entropy for Texture Analysis: An Indicator Kriging Approach," *Int. J.*

- Intell. Syst.*, vol. 29, no. 3, pp. 253–265, Mar. 2014, doi: 10.1002/int.21639.
- [24] T.-T. Le, J. Guilleminot, and C. Soize, “Modélisation stochastique continue et identification inverse d’interphases aléatoires à partir de simulations atomistiques,” Aug. 2015, Accessed: Jun. 21, 2019. [Online]. Available: <https://hal-upec-upem.archives-ouvertes.fr/hal-01192569>
- [25] S. Dabo-Niang, C. Ternynck, and A.-F. Yao, “Nonparametric prediction of spatial multivariate data,” *J. Nonparametric Stat.*, vol. 28, no. 2, pp. 428–458, Apr. 2016, doi: 10.1080/10485252.2016.1164313.
- [26] X. Guyon, *Random Fields on a Network: Modeling, Statistics, and Applications*. New York: Springer-Verlag, 1995. Accessed: May 30, 2020. [Online]. Available: <http://www.springer.com/gp/book/9780387944289>
- [27] C. Gaetan and X. Guyon, Eds., “Modèle spatial du second ordre et géostatistique,” in *Modélisation et statistique spatiales*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–53. doi: 10.1007/978-3-540-79226-0_1.
- [28] L. Anselin and R. Florax, Eds., *New Directions in Spatial Econometrics*. Berlin Heidelberg: Springer-Verlag, 1995. doi: 10.1007/978-3-642-79877-1.
- [29] G. Biau and B. Cadre, “Nonparametric Spatial Prediction,” *Stat. Inference Stoch. Process.*, vol. 7, no. 3, pp. 327–349, Oct. 2004, doi: 10.1023/B:SISP.0000049116.23705.88.
- [30] S. Dabo-Niang and A.-F. Yao, “Kernel regression estimation for continuous spatial processes,” *Math. Methods Stat.*, vol. 16, no. 4, pp. 298–317, Dec. 2007, doi: 10.3103/S1066530707040023.
- [31] A. Caseri, “Apport de la simulation conditionnelle géostatistique pour la prévision immédiate d’ensemble de pluies et l’alerte aux crues rapides,” PhD Thesis, Tese de Doutorado, AgroParisTech, 2017.
- [32] W. J. K. Aldworth, “Spatial prediction, spatial sampling, and measurement error,” Doctor of Philosophy, Iowa State University, Digital Repository, Ames, 1998. doi: 10.31274/rtd-180813-10766.
- [33] T. Gneiting, “Strictly and non-strictly positive definite functions on spheres,” *Bernoulli*, vol. 19, no. 4, pp. 1327–1349, Sep. 2013, doi: 10.3150/12-BEJSP06.
- [34] J. J. Harden, “A Bootstrap Method for Conducting Statistical Inference with Clustered Data,” *State Polit. Policy Q.*, vol. 11, no. 2, pp. 223–246, Jun. 2011, doi: 10.1177/1532440011406233.
- [35] P. M. Dixon, “Bootstrap Resampling,” in *Encyclopedia of Environmetrics*, American Cancer Society, 2006. doi: 10.1002/9780470057339.vab028.
- [36] W. Sauerbrei and M. Schumacher, “A bootstrap resampling procedure for model building: Application to the cox regression model,” *Stat. Med.*, vol. 11, no. 16, pp. 2093–2109, 1992, doi: <https://doi.org/10.1002/sim.4780111607>.
- [37] B. Efron, “The Bootstrap and Modern Statistics,” *J. Am. Stat. Assoc.*, vol. 95, no. 452, pp. 1293–1296, Dec. 2000, doi: 10.1080/01621459.2000.10474333.
- [38] R. Kumar, “A Real Time Approach to Compare PSNR and MSE Value of Different Original Images and Noise (Salt and Pepper, Speckle, Gaussian) Added Images,” p. 4, 2018.
- [39] U. Sara, M. Akter, and M. S. Uddin, “Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study,” *J. Comput. Commun.*, vol. 7, no. 3, Art. no. 3, Mar. 2019, doi: 10.4236/jcc.2019.73002.
- [40] S. Md. R. Islam, X. Huang, and K. Le, “A Novel Image Quality Index for Image Quality Assessment,” in *Neural Information Processing*, Berlin, Heidelberg, 2013, pp. 549–556. doi: 10.1007/978-3-642-42051-1_68.

**Detection and Segmentation of
Erythrocytes in Multispectral
Label-Free Blood Smear
Images for Automatic Cell
Counting**

Résumé :

Dans ce travail, nous nous sommes intéressés à la segmentation non supervisée d'images multispectrales de frottis sanguin non marqué, appliquée au diagnostic du paludisme. Notre méthode tire profit de la loi de Beer Lambert en transmission par application de l'équation de la standardisation statistique suivie d'une segmentation adaptative locale puis de la fermeture des contours par hystérésis et enfin de l'algorithme de ligne de partage des eaux. Ne nécessitant aucun prétraitement, il capitalise non seulement le temps dû à ce dernier mais également un gain de $2/3$ du temps d'acquisition. Aussi, la validation de nos résultats a été effectuée par l'analyse des paramètres de performance tels que la (Accuracy =justesse), la précision, le rappel et la F-Mesure. Ce qui nous donne, une sensibilité (Précision) de 98,47% et de spécificité (Rappel) 98,23%, un degré de précision (F-Mesure) de 98,34% et une justesse (Accuracy) de 96,75%.

Detection and Segmentation of Erythrocytes in Multispectral Label-Free Blood Smear Images for Automatic Cell Counting

Solange Doumun^{a,b,c,1}, Sophie Dabo-Niang^b, and Jérémie ZOUEU^c

^a Univ, Lille, CNRS, UMR 8524-Laboratoire Paul Painlevé, INRIA-MODAL, F-59000 Lille, France

^b Univ, Lille, CNRS, UMR 9221-Laboratoire Lille Economie Management, INRIA-MODAL, F-59000 Lille, France

^cLaboratoire d'Instrumentation, Image et Spectroscopie, Institut National Polytechnique
Félix Houphouët-Boigny, BP 1013, Yamoussoukro, Côte d'Ivoire

Abstract:

In this work we propose an efficient approach of image segmentation for multispectral images of unstained blood film and erythrocytes automatic counting. Our method takes advantage of Beer Lambert's law by using first statistical standardization equation applied to transmittance images, followed by the local adaptive threshold to detect the blood cells, and hysteresis contour closing to obtain the complete blood cells boundaries and finally the watershed algorithm is used. With this method, images pre-processing is not required leading to time savings. We obtained the following results that show that our technique is effective, efficient and fast: Precision of 98.47% and recall of 98.23%, a degree of precision (F-Measurement) of 98.34% and an accuracy of 96.75%.

Keywords: Multispectral imaging, segmentation, malaria, automatic diagnosis, image analysis

¹ To whom all correspondence should be addressed. Email: mekapeusolange.doumunepseoulai.etu@univ-lille.fr

1. Introduction:

Malaria continues to be a common and critical public health problem in developing countries. It is caused by plasmodium parasites that are transmitted through the bite of mosquitoes [1]. In 2018, 228 million cases of malaria were recorded and 405 000 deaths from this disease, mostly children under 5 years, and mostly in Sub-Saharan Africa [2]. Early and reliable detection is needed for quick action and is considered as an effective way to fight this pandemic. Gold standard diagnosis is made by blood collection and visual inspection by clinic staff using simple, visible light-based microscopes. This manual process is time-consuming and unreliable. In this article, we discuss detection and segmentation of red blood cells in multispectral and multimodal unstained blood smear images for automatic erythrocytes counting in computer-assisted malaria diagnosis effort. Automatic malaria diagnosis based image processing is today an active field of research since the advent of computer-aided diagnosis [3]. Generally, the analysis begins with images pre-processing, followed by segmentation and finally classification. According to the state-of-the-art, several works propose various methods for

segmentation in malaria diagnosis. We can mention the following methods:

Otsu [4]-[7] is well established technique based on [8] a global thresholding. The principle is built on the calculation of an optimal threshold (which maximizes the interclass variance) from the image histogram. This approach is simple, fast, and well adapted to bimodal histograms (background and foreground) such as blood smear images. However, it does not consider the spatial position of pixels in the image. Uneven illumination images are also missed in the processing, making it difficult to find the optimal threshold. Watershed segmentation [9] describes the image as a topographic landscape with ridges and valleys [10]. Considering images as reliefs, the gray levels are identified as altitudes. The image is thus subdivided into watersheds, corresponding to the geographical space in which all the waters converge towards the same minimum. These basins are then delimited by ridge lines or watersheds. They make it possible to identify and extract continuous and homogeneous regions of the image (objects such as cells). Nonetheless sensitive to noise, it has shown a tendency towards over segmentation because each minimum generates a watershed. Kmeans algorithm [11], [12] partitions the

points in the data matrix X into K clusters. This partition is iterative and minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances [12]. The algorithm is initialized by randomly chosen a set of centers. Afterwards, it iteratively searches the optimal partition. Each pixel is assigned to their nearest cluster centroid. After all pixels have been assigned, the new cluster centroids are recalculated by the average of each cluster. The algorithm stops when no changes are made again. This is the steady state. Its simplicity and its ability to process large datasets makes it effective. It works well with uneven illumination images. However, the result is highly dependent on the starting centers. Therefore, it is necessary to run the algorithm several times before reaching the optimal solution. A prior knowledge of the data is also essential. Applied to blood cells, the algorithm leads to a partial segmentation of the blood cells. In fact, it does not consider the spatial dependencies of the pixels. Moreover, the boundaries of the partitions are different from the cell contours. Hough's transform [11], [12] is a pattern recognition method that uses the spatial information of the characteristic points of image. In its implementation, it allows to pass from the space of the acquired image to

the space of the parameters that represent the geometrical shapes searched for in the image [13]. Robust to noise, the Hough transform is also robust curve detector especially circular Hough Transform for the circular shapes as red blood cells. The drawback of this method is the difficulty in identifying and classifying unwanted contours, as well as discontinuities in contours. Changes in the slope (generally closed) around the area of interest towards an equilibrium position at the edges of the object must be detected under the action of several forces (internal energy, potential energy and external energy). This equilibrium position is the one that minimizes the energy. This technique called active contour [14]-[16] is suitable for several imaging applications. But the result depends on the initial setting which leads to false contours if it is incorrectly fixed. It also encounters difficulties in detecting several objects in the same image as well as digital instability and unperforms to segment highly overlapping cells [17]. Fuzzy Logic Segmentation [18] implements the concept of degree of membership. It defines the membership function which lies at the heart of any fuzzy logic system. The capability of fuzzy rule-based techniques significantly depends upon it. The fuzzy logic is flexible enough to consider inaccuracies and uncertainties.

Imitating human reasoning, it is easy to understand even by non-specialists. Though, the accuracy of the segmentation is not very high and the number of classes of objects in the image must be known in advance.

Thus, these methods depend on the characteristics of the image, that are affected by a number of factors such as the acquisition system, the red blood staining type, etc. As result of the described approaches above, it is challenging to provide a universal method, especially for the type of images used in this work. Various methods have been proposed with detection rates higher than 98% compared to manual counting [19]. However, these techniques are very sensitive to noise and involve pre-processing steps which require the acquisition of additional images making them time consuming. To counter this, we employ front-end processing using image standardization in transmission mode aiming to isolate the red blood cells and facilitate the segmentation.

2. Materials and methods:

2.1 Multispectral and Multimodal Microscope:

The image acquisition device consists of a modified Brunel optical microscope. Zoueu et

al. [12], [20] introduced the concept of unstained blood smear malaria diagnosis using spectral imaging with an initial microscope version and in 2011, Brydegaard et *al.* [21], built the model used in this work. The system is composed of 3 acquisition modes (Transmission, Reflection and Scattering), using a set of 13 wavelengths ranging from ultraviolet to near infrared (375 to 940 nm). This device has been previously described by Brydegaard et *al.* [21]. It is a modified commercial metallurgical Microscope (BRUNEL SP80). The ocular was substituted by a 5MPix (2592*1944) monochromatic CMOS camera (Guppy-503B, Allied Vision Technology, with a MT9P031 sensor from Micron/Aptina) equipped with individual pixel size of 2.2*2.2 μm , each having a 12 bit depth. [22]

The original filament light sources were substituted by a set multiplexing module of LEDs. The whole system was controlled by a PC running a custom-made LabVIEW code (National Instruments NI).

2.2 Image Acquisition:

The program controls camera exposure time and gain, LED multiplexing and current through the NI data acquisition board DAQ [21]. Bright reference is used to calibrate the exposure times and gains and sample images

are captured and saved in 12-bits unsigned integer images TIFF format (Figure1). The result is a low contrast greyscale image. The blood cells cannot obviously be differentiated from the background contribution. The proposed method aims to process these acquired images for detection, segmentation and counting of blood cells.



Figure 1: Original sample image acquired with LED light at 435nm in transmission mode

All samples for all geometries (39 images) were acquired keeping the same region of interest [19]-[21]. The spectral image is represented by a 2D matrix:

$$I_{xy} = \int_0^{+\infty} S_{sy\lambda} R_{sy\lambda} D_{sy\lambda} d\lambda \quad (1)$$

where I is a spectral image,

S emission spectrum,

R sample response

D detector sensibility.

If we consider, the spectral dimension we have a 3D image $I_{sy\lambda}$ for each geometry.

2.3 Pre-processing:

Initially, the pre-process is performed using a customized algorithm in MATLAB (MathWorks). Once acquired, the images are saved, and background are subtracted from the sample image, followed by images normalization to ensure homogeneous illumination. Spectral images are obtained as below:

$$I_{\lambda} = \frac{I_{S_{\lambda}} - I_{B_{\lambda}}}{I_{R_{\lambda}} - I_{B_{\lambda}}} \quad (2)$$

I_{λ} grayscale image in wavelength λ

$I_{S_{\lambda}}$ Sample Image

$I_{R_{\lambda}}$ Bright reference Image

$I_{B_{\lambda}}$ Dark reference Image.

To obtain 13 spectral images for each mode, a total of 39 images (13 samples, 13 Brights references Images, and 13 Darks references Images) are taken. The final spectral image of a blood smear sample consists of 39 grayscale images (13 per mode) requiring extra acquisition time for the additional images needed in the normalization Equation 2. Finally, a 2D-median filter is applied to the normalized images in order to remove the noise.

2.4 Beer Lambert's Law in Transmission:

Beer Lambert's law establishes a proportionality between the concentration of a

chemical entity and the length path (thickness) of light in each media. According to this law the absorbance is:

$$A = \epsilon lc = -\log(T),$$

where l is the optical path length (thickness)

c concentration

ϵ absorptivity.

T transmittance

Transmittance is therefore inversely proportional to the thickness and the concentration:

$$T = e^{-\epsilon lc} \quad (3)$$

According to this law, in the transmission mode image, the pixel values of the red blood cells are lower than the pixel values of the background image (Figure 1), and therefore lower than the average pixels values.

Standardization:

Standardization allows to reduce data and to have zero mean and a standard deviation of 1. The aim of this process is to overcome the sensitivity of the data to overly large dispersions, while maintaining the shape of joint distributions.

Let x_j be the grey level of the pixel in a given image, the standardized grey level X_m is

$$X_m = \frac{x_j - \bar{x}}{\bar{\sigma}} \quad (4)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, is the sample grey level mean of all pixels of the image, while the

$$\bar{\sigma} = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2}$$

sample standard deviation is:

Normalization:

With the same intent as standardization, normalization allows scaling data in $[0,1]$.

Let x_{min} and x_{max} be respectively the minimum and the maximum values of the pixels grey level in a given image. A given grey level x_j is normalized as X_n :

$$X_n = \frac{x_j - x_{min}}{x_{max} - x_{min}}$$

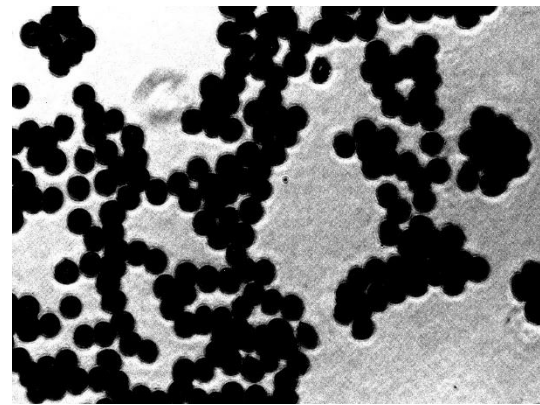


Figure 2: Standardized image of Figure 1

Applying the standardization equation on (Figure 1) gives the result in (Figure 2) below.

The image contrast is then enhanced, and the blood cells can be clearly distinguished from the background. However, the background illumination is not homogenized.

In the following, the proposed automatic images analysis method is performed through a customized algorithm in R software 3.5.3.

2.5 Adaptive thresholding:

Unlike global thresholding, the adaptive one allows the threshold to be different depending on the region of the image [23]. Therefore, spatial dependencies of the background signal due to spurious signals or uneven illumination can be anticipated. The value of pixel intensity is determined from the local neighborhood. We obtain a new smoothed image after applying to all pixels an adaptive thresholding procedure. The thresholding is performed by comparing each pixel's intensity to a smoothed background image. For the smoothing, we used the `filter2` function of the image processing package "EImage" [23] of R language. This function uses a 2D Fast Fourier Transform which is a convolution product equivalent to the following equation:

$$(w * f)(x, y) = \sum_{s=-\infty}^{+\infty} \sum_{t=-\infty}^{+\infty} w(s, t) f(x + s, y + t)$$

where w is a weighting function whose value is highest in the middle of the window ($s = t = 0$) and decreases at the border.

2.6 Hysteresis contour closing:

The adaptive thresholding algorithm allows to detect the red blood cells contours. However, some of these contours are not closed. The hysteresis contour closing is performed to obtain the complete blood cells boundaries. It

consists in detecting the extreme points of the unclosed contours and their extension.

Two different thresholds are used on I which is a greyscale image. A high threshold $T1$ and a low threshold $T2$ such that:

$$T1 > T2, \min(I) \leq T1 \text{ and } T2 \leq \max(I)$$

Let $S1$ and $S2$ be two sets (Binary Mask Images) such as:

$$S1 = \{x \in \mathbb{R}^n | I(x) \geq T1\},$$

$$S2 = \{x \in \mathbb{R}^n | I(x) \geq T2\}.$$

Any edges with intensity greater than $T2$ are the sure edges. Any edges with intensity less than $T1$ are sure to be non-edges.

The edges between $T2$ and $T1$ thresholds are classified as edges only if they are connected to a sure edge otherwise discarded.

This amounts to find C where $S2 \subset C \subset S1$. C is the binary image representing the contours, equal to 1 for pixels such as $|\nabla I(x)| \geq T1$ and 0 elsewhere [24]. As far as we are concerned, the set $S1$ is our standardized, normalized and segmented image with threshold 0, corresponding to the value for which Equation 4 is zero. The set $S2$ is the result of a convolution of the image obtained in $S1$ by the magnitude gradient of the FreiChen [25], [26] defined by:

$$M = \sqrt{D_x^2 + D_y^2},$$

where

$$D_x = \frac{1}{2+\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{pmatrix} \text{ and}$$

$$D_y = \frac{1}{2+\sqrt{2}} \begin{pmatrix} -1 & \sqrt{2} & -1 \\ 0 & 0 & 0 \\ 1 & \sqrt{2} & -1 \end{pmatrix}.$$

Performance Assessment

The proposed approach effectiveness was measured using different parameters such as, *TP*: True Positive, *FP*: False Positive, *FN*: False Negative, to evaluate the performance indexes which are: Precision, Recall, Accuracy and F-Measurement. In addition, we use a parameter *DR*: Detection Rate which is the coverage rate of the manually counted cells. A comparison measure, namely, the positive agreement (P_{pos}), alternative to the well-known kappa index, is also computed to compare the manual method to automatic methods. These indices are calculated according to the following equations:

$$P_{pos} = \frac{2 * P_{11}}{P_{.1} + P_{1.}}$$

$$DR = 100 * \frac{MC - CelDel}{MC},$$

$$Precision = \frac{TP}{TP + FP},$$

$$Accuracy = \frac{TP}{TP + FP + FN},$$

$$Recall = \frac{TP}{TP + FN},$$

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall},$$

where: $TP = AC - FP$,

$$FP = CES + FD$$

$$FN = ME + CelDel$$

$$P_{.1} = P_{21} + P_{11}$$

$$P_{1.} = P_{11} + P_{12}$$

$$P_{11} = \text{Min}(\text{Rate A}, \text{Rate B})$$

$$P_{12} = \begin{cases} \text{Rate B} - \text{Rate A} & \text{if Rate B} > \text{Rate A} \\ 0, & \text{otherwise} \end{cases}$$

$$P_{21} = \begin{cases} 0 & \text{if Rate B} > \text{Rate A} \\ \text{Rate B} - \text{Rate A}, & \text{otherwise} \end{cases}$$

$$P_{22} = 1 - \text{Max}(\text{Rate A}, \text{Rate B})$$

P_{pos} : The index of positive agreement between two cell counting methods (Rate A and Rate B).

CelDel: Number of cells deleted by segmentation.

CES: Counting error due to cell segmentation

FD: out-of-cell detection

FN: number of cells omitted from the automatic counting

ME: Merging Error; counting error due to overlapping cells.

MC: number of cells manually counted

AC: number of cells automatically counted

CE: Counting error; $EC = 100 * \left| \frac{MC - AC}{MC} \right|$

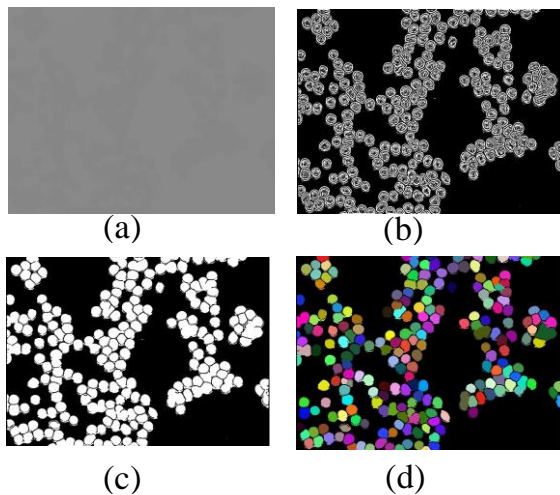
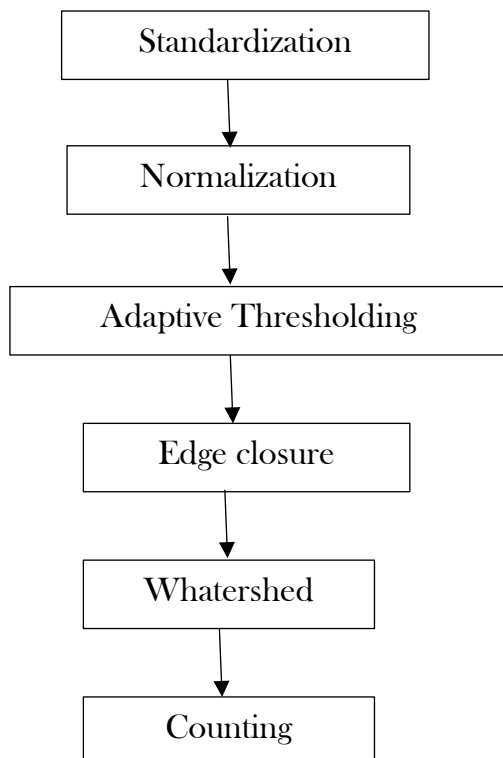


Figure 3: Segmentation process illustrated with a 435 nm image; a- Sample image, b- Adaptive thresholding, c- post-processing, and d- Counting

3. Results and discussions

Algorithm



3.1 Segmentation and counting process

The proposed segmentation approach used in this work can be divided in three basic steps.

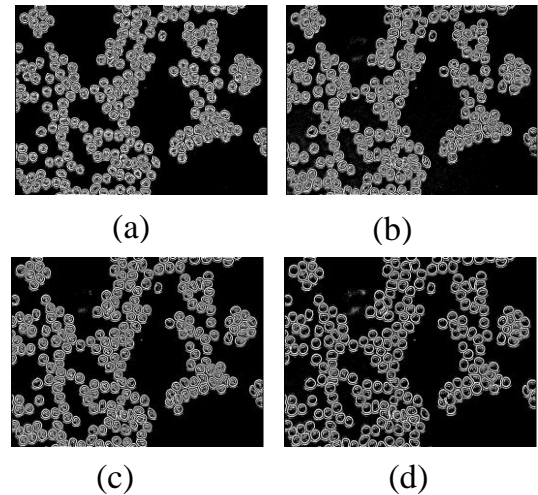


Figure 4: Adaptive thresholding in Transmission according to wavelengths: a- Image at 435 nm, b- Image at 525 nm, c- Image at 590 nm, d- Image at 700 nm.

(Figure 3) summarizes these steps. (Figure 3-a) is the original sample image from which the processing starts. It's a low contrast image where the blood cells are not clearly identified.

- The first step consists in standardization (Figure 2) and normalization.
- Step 2 realizes an adaptive thresholding which aims to separate the erythrocytes from the background (Figure 3-b).
- In step 3, the inhomogeneity of segmented cells is overcome by post-treatment which consists of contour closing by hysteresis and cell filling (Figure 3-c). The watershed algorithm is thus performed to isolate the red blood cells. This leads to counting process (Figure 3-d). The (Figure 3) reflects the whole

process using LED light at wavelength 435 nm, in transmission mode.

3.2 Geometry selection:

As indicated previously, the image used in this work is acquired thanks to 3 modalities and 13 wavelengths ranging from 370 nm to 940 nm. This section aims to explore both geometries and wavelengths, using the proposed methodology, to determine the best erythrocytes contrast over the background to achieve suitable segmentation.

(Figure 4) shows images acquired in transmission mode using wavelengths 435 nm, 525 nm, 590 nm and 700 nm respectively. The choice of these wavelengths reflects the main absorption peaks of heme and their potential for best contrast. The original sample images were segmented using adaptive segmentation applied to (Figure 2) followed by normalization.

In each image (Figure 4), we observe a black background which corresponds to both glass slide and the plasma contribution; the cells with higher pixels intensities correspond to the erythrocytes. The contrast is obtained from the variability of hemoglobin concentration, the absorbencies and the relative thicknesses of the plasma and erythrocytes. Contrasts

variations from one wavelength image to another comes from the absorptivity.

In (Figure 5), the mode is switched in reflection while maintaining both region of interest (ROI) and wavelength at the same value. (Figure 5-a) is the image with 435nm, which represents the highest absorption peak for haemoglobin and gives the higher contrast of the blood smear images. For reflection, the contrast derives from the relative reflection coefficients between the background and the erythrocytes which is composed

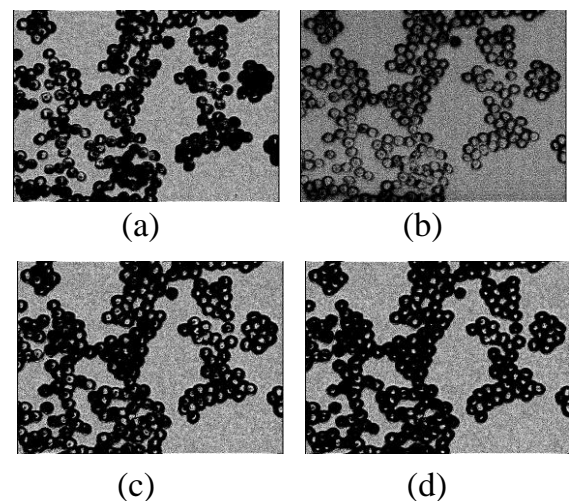


Figure 5: Adaptive thresholding in Reflection according to wavelengths: a- Image at 435 nm, b- Image at 525 nm, c- Image at 590 nm, d- Image at 700 nm.

predominantly of haemoglobin.

By comparison, the segmentation with transmission images leads to higher contrast and reveals more clearly erythrocytes edges while resulting to simple post-processing by cell filling algorithm, edge closing algorithm

and watershed closing algorithm. The cells isolation and counting processing is then shortened. Transmission mode provides the best results. The system is then switched to the scattering mode for above-mentioned ROI and wavelengths (Figure 6).

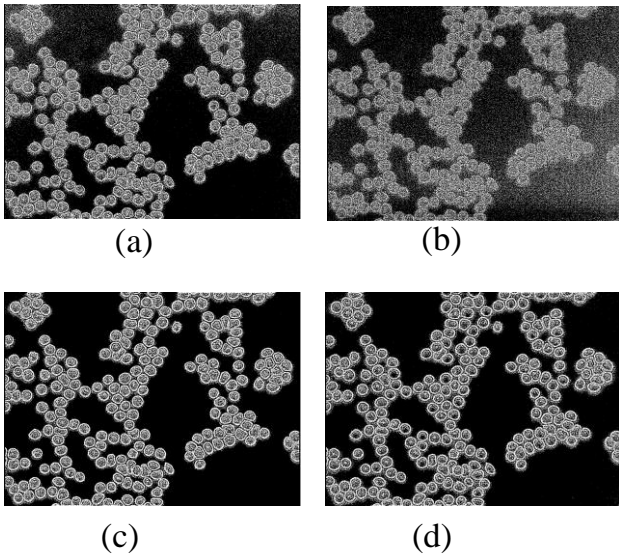


Figure 6: Adaptive thresholding in Scattering according to wavelengths: a- Image at 435 nm, b- Image at 525 nm, c- Image at 590 nm, d- Image at 700 nm

Automatic counting resulting from the proposed segmentation approach applied to the 13 wavelengths in the Scattering, Reflection and Transmission modes are presented in (Table 1. The examined ROI contains 225 erythrocytes according to a manual counting. The automatic counting deviations from the manual one are estimated to 8.58%, 43.59%, 77.88%, respectively in transmission, reflection and Scattering modes. These findings lead to the selection of

transmission mode which is compatible with Beer Lambert's law equation.

By comparing the counting accuracy of all wavelengths in transmission mode, the best result is obtained with 435 nm which is associated with an automatic counting error of 1.98% and an Accuracy of 98%. Furthermore, the images at 525 nm, 590 nm, 625 nm, 660 nm, 700 nm, 750 nm and 810 nm also introduce smaller automatic counting errors (less than 10%) and an Accuracy higher than 90% as shown in (Table 2). Also, all these wavelengths detect 100% of the cells.

3.3 Result on non-pretreated images:

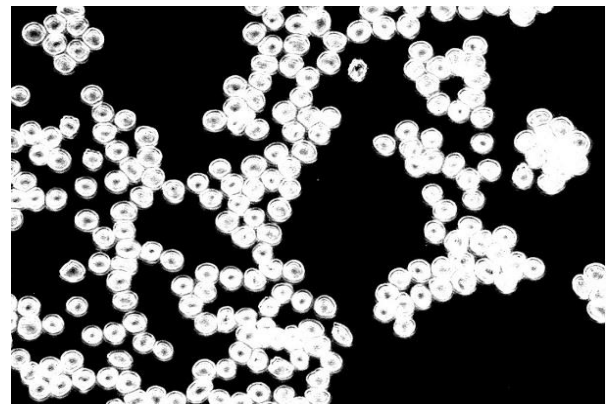


Figure 7: Standardized and inverted image of Figure 2

Here we describe the results of our method on images acquired from the unstained blood smear without the normalization pre-processing step which requires the acquisition of Dark reference images and Bright reference images Equation 2.

As illustration is given with the wavelength 435 nm image, see (Figure 3-a). By applying the standardization Equation 4, we get negative values for the red blood cell pixels, due to Beer Lambert's law Equation 3. This is consistent with the inverted standardized image as depicted in (Figure 7).

All the cells are effectively identified. As stated above, the counting effectiveness depends on the cell's isolation algorithm.

Adaptive thresholding is used to isolate red blood cells. However, depending on the wavelength, some features (contour details) can clearly be seen and others not. In this way, the blood cells uncorrectly segmented are removed. All cells are detected by the combination of several wavelengths. We can obtain and improve this result by using a proposed hysteresis contour closing algorithm applied to a single image (one of these wavelengths) as shown in (Figures 3-c and 3-d).

3.4 Result with pre-processed image:

Applying the standardization equation to the pre-processed image produces a different result. In fact, only the contours are shown in (Figure 8-a). This is due to the pre-processing algorithm used for the image smoothing. Because of the concave shape of the red blood cells, the effect of Beer Lambert's law in

transmission Equation 3 is more accentuated on the contours which are thick and less affected by the smoothing algorithm. However, the cells are segmented by closing the contours, as shown in (Figure 8).

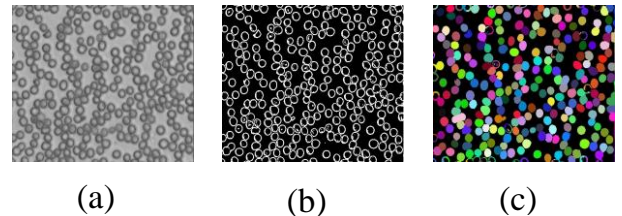


Figure 8: Pre-processed image segmentation; a- Image at 435nm after normalization and filtering, b- Image after adaptive thresholding, c- Counting

3.5 Discussion

In the summary (Table 3), we compare the performance of the proposed automatic counting with the manual one on 30 raw images. The proposed algorithm is applied to images with two different sizes. A first batch of seven (7) samples of size is 1020*768 and a second batch of 23 images of size 2048*1088. The thresholding algorithm reveals erythrocytes that can be overlapped or attached and cannot be isolated spatially by the isolation algorithm, this leads to counting error. However, the result can be improved by considering the magnification parameter of the acquisition system in order to define the relative size of the erythrocyte.

As shown in (Table 3), the acquisition time is one third of that of the method previously used in our laboratory. Secondly, we capitalize the execution time of the normalization and pre-processing algorithms that are usually performed [11], [19], [21], [26]-[28]. Finally, the speed average is 30 seconds for 1020*768 size images and less than 2 minutes for 2048*1088 size images. This meets the real-time processing demands.

The numerical results show also the detection rates of 100% for the first batch and 99.65% for the second with an average of 99.82%, indicating the capacity that can be obtained if the isolation and counting post-processing is improved.

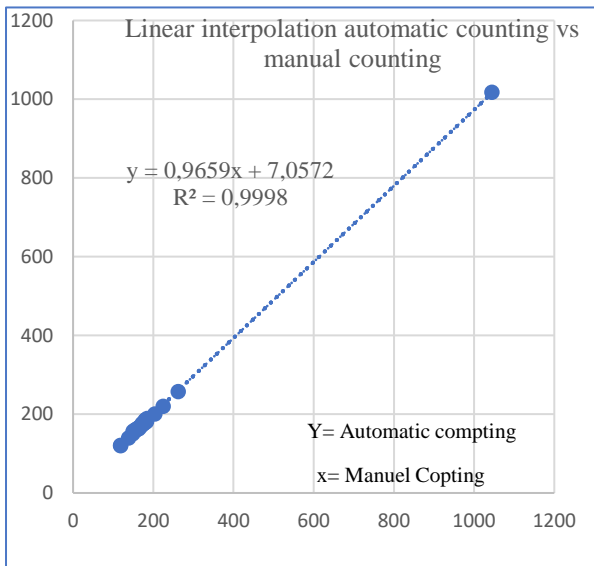


Figure 9: Automatic counting vs Manuel counting (Figure 9) gives a determination coefficient R^2 of 0.9998 (when performing a simple linear regression) which shows the closeness between

the automatic and manual counting methods. We also obtained an error rate of 1.6% (Table 3) and root-mean-square error, RMSE of 5.86 (Table 4). These results show a clear improvement over those of Bagui et al. [19].

Our method was compared to five authors who reported on the segmentation and counting of red blood cells [30].

Analysis of (Table 5) gives us an Accuracy of 96.8% behind Loddo [31]. This is due to an inadequate of our cell isolation algorithm which could be improved by considering the magnification features of the acquisition system in order to parameterize the relative size of the erythrocyte. However, we have a better result according to Precision 98.5% and Recall 98.2%. We obtained a degree of precision (F-Measurement) of 98.3% which is a key indicator of performance for a diagnostic test. The positive agreement index of 0.99 shows the high agreement [32]-[35] level of the proposed method with manual counting compare to other's.

4. CONCLUSION AND PERSPECTIVES:

In this study, a method was introduced for detection and counting of RBCs from spectral images without using references images. By using this approach, the time saved can be

critical in emergency situation and when the peak demand of test is highest; taking into consideration that each extra batch of 13 images will add 2.6 minutes of acquisition time and about the same for pretreatment. Moreover, the novel algorithm runs in less than 2 minutes for large images (2048*1088). The performances evaluation of the proposed approach results in a Precision of 98.47% and Recall 98.23%, which gives a degree of precision (F-Measurement) of 98.34% and Accuracy of 96.75%. These results show that the proposed method is efficient and can lead to a real-time diagnosis.

In perspective, we plan to improve the counting rate by merging over-segmented cells and disconnecting the ones that are overlapping.

TABLE 1 - Geometry comparaison

Wavelength (nm)	Scattering		Reflexion		Transmission	
	A C	Error	A C	Error	A C	Error
		Rate (%)		Rate (%)		Rate (%)
375	34	84,89	80	64,44	160	28,89
400	25	88,89	182	19,11	196	12,89
435	28	87,56	216	4,00	221	1,78
470	154	31,56	165	26,67	200	11,11
525	18	92	110	51,11	225	0
590	48	78,67	203	9,78	224	0,44
625	49	78,22	110	51,11	217	3,56
660	33	85,33	80	64,44	227	0,89
700	58	74,22	150	33,33	211	6,22
750	98	56,44	109	51,56	227	0,89
810	44	80,44	130	42,22	211	6,22
850	29	87,11	65	71,11	190	15,56
940	29	87,11	50	77,78	173	23,11
Average	49,8	77,88	126,92	43,59	206,31	8,58

TABLE 2 -Wavelength comparison in Transmission mode

A C	Error Rate(%)	Accuracy (%)
160	28,89	82,47
196	12,89	90,74
221	1,78	98,22
200	11,11	90,91
225	0,00	90,68
224	0,44	96,89
217	3,56	94,71
227	0,89	81,60
211	6,22	94,95
227	0,89	93,91
211	6,22	97,22
190	15,56	94,97
173	23,11	89,58
206,31	8,58	92,07

TABLE 3 - Counting synthesis table.

No								C	E	Precisio			F-	P _{tot}
Image	MC	AC	NOC	CES	FD	ET (s)	DR(%)	(%)	n	Recall	Accuracy	measure		
1	225	221	5(12)	1	0	26,34	100	1,77	99,55	96,92	96,49	98,21	0,99	
2	150	156	7(14)	1	0	22,06	100	4,00	99,36	95,54	95,09	97,48	1,00	
3	149	150	5(10)	2	3	33,59	100	0,67	96,67	96,67	93,55	96,67	0,99	
4	204	200	5(10)	1	0	26,38	100	1,96	99,50	97,55	97,07	98,51	0,99	
5	262	257	5(10)	0	0	15,11	100	1,90	100	98,09	98,09	99,04	0,99	
6	119	120	3(6)	2	0	44,78	100	0,84	98,33	97,52	95,93	97,93	1,00	
7	119	121	3(6)	3	0	48,34	100	1,68	97,52	97,52	95,16	97,52	1,00	
Partial Average1						30,94	100	1,89	98,70	97,12	95,91	97,91	0,99	
8	1045	1017	20(44)	0	0	234,27	99,61	2,68	100	97,31	97,31	98,64	0,99	
9	139	139	1(4)	4	0	101,76	99,28	0,00	97,12	97,12	94,41	97,12	0,99	
10	155	160	0	3	0	136	99,37	3,23	98,13	99,38	97,53	98,75	1,00	
11	178	177	0	3	0	124,18	100	0,56	98,31	100	98,31	99,15	0,99	
12	182	182	1(2)	2	0	102,92	100	0,00	98,90	99,45	98,36	99,17	0,99	
13	171	171	0	0	0	76,7	100	0,00	100	100	100	100	1,00	
14	183	181	0	4	0	86,09	100	1,09	97,79	100	97,79	98,88	0,98	
15	185	189	0	4	0	68,93	100	2,16	97,88	100	97,88	98,93	1,00	
16	181	183	0	4	0	93,22	98,87	1,10	97,81	98,87	96,73	98,34	0,99	
17	172	175	0	4	0	90,41	99,41	1,74	97,71	99,41	97,15	98,55	1,00	
18	164	163	0	3	0	107,81	99,39	0,61	98,16	99,38	97,56	98,77	0,99	
19	153	156	0	5	0	114,72	100	1,96	96,79	100	96,79	98,37	0,99	
20	175	177	0	3	0	133,12	98,28	1,14	98,31	98,30	96,66	98,30	1,00	
21	180	185	0	5	0	114,62	100	2,78	97,30	100	97,30	98,63	1,00	
22	182	183	0	2	0	108,53	99,45	0,55	98,91	99,45	98,37	99,18	1,00	
23	156	159	0	3	0	94,83	100	1,92	98,11	100	98,11	99,05	1,00	
24	165	165	0	0	0	89,73	100	0,00	100,00	100	100	100	1,00	
25	159	162	0	3	0	84,06	100	1,89	98,15	100	98,15	99,07	1,00	
26	160	161	1(2)	3	0	87	99,37	0,63	98,14	98,75	96,93	98,44	0,99	
27	164	166	0	3	0	83	100	1,22	98,19	100	98,19	99,09	1,00	
28	175	177	1(3)	5	0	105,55	99,42	1,14	97,18	98,28	95,55	97,72	0,99	
29	174	177	1(2)	3	0	90,94	100	1,72	98,31	99,43	97,75	98,86	1,00	
30	181	185	0	3	0	95,89	99,42	2,21	98,38	99,43	97,82	98,90	1,00	
Partial Average2						105,40	99,65	1,32	98,24	99,33	97,59	98,78	1,00	
Total Average							99,82	1,60	98,47	98,23	96,75	98,34	0,99	

TABLE 4 - Performance Indicator

Performance Indicator	Value
RMSE	5,86
MAE	3,03
MAPE %	1,30
R ² adjust	0,9998

TABLE 5 - Blood cells counting performance compared with the state-of-the-art

	Alomari and al[36]	Loddo and al[31]	Moallem and al[16]	Poostchi and al[17]	Cecilia and al[30]	Our Appr
Images type	Stained thin blood smears RGB color images from Dhruv Pathology Lab	Stained thin blood smears JPG 24 bits color depth (2592*194 4) images from ALL-IDB dataset (canon powershote G5 camera)	Stained thin blood smears RGB color images from regular light microscope by smartphone	Stained thin blood smears RGB color images acquired by a Zeiss Axio Imager	Stained thin blood smears JPG 24 bits color depth (2592*194 4) images from ALL-IDB and MP-IDB dataset (canon powershote G5 camera)	unstained blood smears grayscale 12-bit depth images acquired by a 5MPix (2592*1944) monochromatic CMOS camera
Methods used	Iterarative structured circle	Nearest Neighbour and Support	Adaptative meanshift clustering	Laplacien of Gaussian filter and Geodesic	Modified Edge Boxes	Standization, Adaptive thresholding and the

	detection algorithm	Vector Machine	and algorithm	snake contour	active	watershed algorithm	
Accuracy	95,3	98			92,5	95,6	96,8
Precision	95	89	98		99,3	98,4	98,5
Recall	98	98	93		90,3	95	98,2
F- Measure	96	93	95		94,5	96,6	98,3
P _{pos}	98,99	98,99	96,37		94,90	97,44	99,09

MC : number of cells manually counted

AC : number of cells automatically counted

N O C: Number of overlapping cells

C E S: Counting error due to cell segmentation

F P: False positive; the number of cells which should not be included in the counting

E T(s): Execution Time in seconds

D R: Detection rate; the manual counting cells coverage rate

C E: Counting error; $EC = 100 * \left| \frac{MC-CA}{MC} \right|$

Partial Average1: 1020*768 sizes images average

Partial Average2: 2048*1088 sizes images average

Total Average: Average on all the 30 images

RMSE: Root Mean Square Error

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

R² Adjust: Determination coefficient

References

- [1] “Paludisme.” <https://www.who.int/fr/news-room/fact-sheets/detail/malaria> (accessed Nov. 05, 2019).
- [2] “WHO | World malaria report 2019,” *WHO*. <http://www.who.int/malaria/publications/world-malaria-report-2019/en/> (accessed Jul. 30, 2020).
- [3] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, “Image analysis and machine learning for detecting malaria,” *Transl. Res.*, vol. 194, pp. 36–55, Apr. 2018, doi: 10.1016/j.trsl.2017.12.004.
- [4] S. S. Savkare and S. P. Narote, “Automated system for malaria parasite identification,” in *2015 International Conference on Communication, Information Computing Technology (ICCICT)*, Jan. 2015, pp. 1–4. doi: 10.1109/ICCICT.2015.7045660.
- [5] S. S. Devi, S. A. Sheikh, A. Talukdar, and R. H. Laskar, “Malaria Infected Erythrocyte Classification Based on the Histogram Features using Microscopic Images of Thin Blood Smear,” *Indian J. Sci. Technol.*, vol. 9, no. 45, Dec. 2016, doi: 10.17485/ijst/2016/v9i45/94119.
- [6] S. S. Devi, R. H. Laskar, and S. A. Sheikh, “Hybrid Classifier Based Life Cycle Stages Analysis for Malaria-infected Erythrocyte Using Thin Blood Smear Images,” *Neural Comput Appl*, vol. 29, no. 8, pp. 217–235, Apr. 2018, doi: 10.1007/s00521-017-2937-4.
- [7] G. P. Gopakumar, M. Swetha, G. Sai Siva, and G. R. K. Sai Subrahmanyam, “Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner,” *J. Biophotonics*, vol. 11, no. 3, Mar. 2018, doi: 10.1002/jbio.201700003.
- [8] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
- [9] J.-D. Kim, K.-M. Nam, C.-Y. Park, Y.-S. Kim, and H.-J. Song, “Automatic detection of malaria parasite in blood images using two parameters,” *Technol. Health Care Off. J. Eur. Soc. Eng. Med.*, vol. 24 Suppl 1, pp. S33–39, 2015, doi: 10.3233/THC-151049.
- [10] S. Derivaux, S. Lefevre, . Wemmert, and J. Korczak, “Segmentation par ligne de partage des eaux basée sur des connaissances texturales.” <https://icube-publis.unistra.fr/5-DLWK07> (accessed Feb. 06, 2020).
- [11] A. Nanoti, S. Jain, C. Gupta, and G. Vyas, “Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear,” in *2016 International Conference on Inventive Computation Technologies (ICICT)*, Aug. 2016, vol. 1, pp. 1–6. doi: 10.1109/INVENTIVE.2016.7823258.
- [12] J. T. Zoueu, S. Ouattara, A. Toure, S. Safi, and S. T. Zan, “Spectroscopic approach of multispectral imaging of plasmodium falciparum - infected human erythrocytes,” in *2009 3rd ICTON Mediterranean Winter Conference (ICTON-MW)*, Dec. 2009, pp.

- 1-7. doi: 10.1109/ICTONMW.2009.5385598.
- [13] H. MAITRE, "Survey on Hough transform," 1985. <http://hdl.handle.net/2042/2334>
- [14] D. Bibin, N. Madhu S., and P. Punitha, "Malaria Parasite Detection From Peripheral Blood Smear Images Using Deep Belief Networks - IEEE Journals & Magazine." <https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/7931565> (accessed Jan. 11, 2020).
- [15] D. K. Das, A. K. Maiti, and C. Chakraborty, "Automated system for characterization and classification of malaria-infected stages using light microscopic images of thin blood smears," *J. Microsc.*, vol. 257, no. 3, pp. 238-252, Mar. 2015, doi: 10.1111/jmi.12206.
- [16] G. Moallem *et al.*, "Detecting and segmenting overlapping red blood cells in microscopic images of thin blood smears," in *Medical Imaging 2018: Digital Pathology*, Mar. 2018, vol. 10581, p. 105811F. doi: 10.1117/12.2293762.
- [17] M. Poostchi *et al.*, "Malaria parasite detection and cell counting for human and mouse using thin blood smear microscopy," *J. Med. Imaging*, vol. 5, no. 4, p. 044506, Dec. 2018, doi: 10.1117/1.JMI.5.4.044506.
- [18] J. T. Zoueu and S. T. Zan, "Trophozoite stage infected erythrocyte contents analysis by use of spectral imaging LED microscope.," *J. Microsc.*, vol. 245, no. 1, pp. 90-99, 2012, doi: 10.1111/j.1365-2818.2011.03548.x.
- [19] O. K. Bagui and J. T. Zoueu, "Red Blood Cells Counting by Circular Hough Transform Using Multispectral Images," *Science Alert*, 2014. <https://scialert.net/fulltext/?doi=jas.2014.3591.3594> (accessed Jan. 09, 2020).
- [20] S. Dabo-Niang and J. T. Zoueu, "Combining kriging, multispectral and multimodal microscopy to resolve malaria-infected erythrocyte contents: KRIGING, MULTISPECTRAL AND MULTIMODAL MICROSCOPY TO RESOLVE MALARIA-INFECTED ERYTHROCYTE CONTENTS," *J. Microsc.*, vol. 247, no. 3, pp. 240-251, Sep. 2012, doi: 10.1111/j.1365-2818.2012.03637.x.
- [21] M. Brydegaard, A. Merdasa, H. Jayaweera, J. Ålebring, and S. Svanberg, "Versatile multispectral microscope based on light emitting diodes," *Rev. Sci. Instrum.*, vol. 82, no. 12, p. 123106, Dec. 2011, doi: 10.1063/1.3660810.
- [22] A. J. Merdasa, M. Brydegaard, S. Svanberg, and J. T. Zoueu, "Staining-free malaria diagnostics by multispectral and multimodality light-emitting-diode microscopy," *J. Biomed. Opt.*, vol. 18, no. 3, p. 036002, Mar. 2013, doi: 10.1117/1.JBO.18.3.036002.
- [23] G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber, "EBImage—an R package for image processing with applications to cellular phenotypes," *Bioinformatics*, vol. 26, no. 7, pp. 979-981, 2010, doi: 10.1093/bioinformatics/btq046.

- [24] M. Bergounioux, *Introduction au traitement mathématique des images - méthodes déterministes*. Springer, 2015.
- [25] Dicky Apdilah, Muhammad Yasin Simargolang, and Robbi Rahim, “A Study of Frei - Chen Approach for Edge Detection,” *Int. J. Sci. Res. Sci. Eng. Technol. IJSRSET*, vol. 3, no. 1, pp. 59–62, Jan. 2017, doi: 10.5281/zenodo.239730.
- [26] M. Al-khassaweneh, “Robust and Invisible Watermarking Technique Based on Frei-Chen Bases,” in *2019 IEEE International Conference on Electro Information Technology (EIT)*, May 2019, pp. 1–4. doi: 10.1109/EIT.2019.8833875.
- [27] J. T. Zoueu and S. Zan, “Trophozoite stage infected erythrocyte contents analysis by use of spectral imaging LED microscope.,” *J. Microsc.*, vol. 245, no. 1, pp. 90–99, 2012, doi: 10.1111/j.1365-2818.2011.03548.x.
- [28] O. K. BAGUI and J. T. ZOUEU, “(1) Automatic malaria diagnosis by the use of multispectral contrast imaging | Request PDF,” *ResearchGate*, 2014. https://www.researchgate.net/publication/302584015_Automatic_malaria_diagnosis_by_the_use_of_multispectral_contrast_imaging (accessed Oct. 04, 2018).
- [29] O. K. BAGUI, W. YAVO, D. TANO, and J. T. ZOUEU, “Etude de l’effet de l’amodiaquine sur les globules rouges infectés par le paludisme dans les images multispectrales,” <http://www.afriquescience.info>, Nov. 01, 2014. <http://www.afriquescience.info/document.php?id=3968> (accessed Sep. 25, 2018).
- [30] C. Di Ruberto, A. Loddo, and L. Putzu, “Detection of red and white blood cells from microscopic blood images using a region proposal approach,” *Comput. Biol. Med.*, vol. 116, p. 103530, Jan. 2020, doi: 10.1016/j.compbiomed.2019.103530.
- [31] A. Loddo, L. Putzu, C. D. Ruberto, and G. Fenu, “A Computer-Aided System for Differential Count from Peripheral Blood Cell Images,” in *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, Nov. 2016, pp. 112–118. doi: 10.1109/SITIS.2016.26.
- [32] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, Oct. 2012, doi: 10.11613/BM.2012.031.
- [33] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, Mar. 1977.
- [34] A. R. Feinstein and D. V. Cicchetti, “High agreement but low Kappa: I. the problems of two paradoxes,” *J. Clin. Epidemiol.*, vol. 43, no. 6, pp. 543–549, Jan. 1990, doi: 10.1016/0895-4356(90)90158-L.
- [35] D. V. Cicchetti and A. R. Feinstein, “High agreement but low kappa: II. Resolving the paradoxes,” *J. Clin. Epidemiol.*, vol. 43, no. 6, pp. 551–558, Jan. 1990, doi: 10.1016/0895-4356(90)90159-M.
- [36] Y. M. Alomari, S. N. H. Sheikh Abdullah, R. Zaharatul Azma, and K. Omar, “Automatic Detection and Quantification of WBCs and

RBCs Using Iterative Structured Circle
Detection Algorithm,” *Computational and
Mathematical Methods in Medicine*, 2014.
<https://www.hindawi.com/journals/cmmm/2014/979302/> (accessed Mar. 19, 2020).

**Unsupervised Classification of
Erythrocytes in Multispectral
Unstained Blood Smear Images
for Malaria Automatic
Diagnosis**

Résumé :

Dans ce travail nous avons développé un algorithme de classification non supervisé des globules rouges selon qu'ils soient parasités ou non dans le cadre du diagnostic automatique du paludisme. Il commence par l'extraction de 11 descripteurs de textures d'Haralick basés sur la matrice de co-occurrence des niveaux de gris par cellules pour chaque longueur d'onde et géométrie. Étant donné que notre image est composée de 39 longueurs d'onde (images en niveaux de gris), ceci nous donne au total de 429 descripteurs par cellule.

Une analyse en composantes principales fonctionnelles multivariées est donc appliquée afin de réduire la dimension. Nous pouvons alors mettre en œuvre les algorithmes de classification. Deux algorithmes ont été utilisés à savoir le K-moyennes en 3 classes et la Classification Ascendante Hiérarchique. L'analyse spectrale nous permet de caractériser les classes de cellules. Ainsi, pour une caractérisation rigoureuse, nous croisons les résultats des 2 méthodes de classification afin d'en faire une classification « hybride » en sélectionnant les cellules appartenant à la fois à chacune des classes parasitée ou saine des deux algorithmes de classification.

Le but ici étant d'appliquer à ces classes de cellules, un algorithme de classification supervisée pour apprendre d'elle et par la suite classer le reste des cellules afin d'établir la parasitémie.

Unsupervised Classification of Erythrocytes in Multispectral Unstained Blood Smear Images for Malaria Automatic Diagnosis

Solange Doumun^{a,b,c,3}, Sophie Dabo-Niang^b, and Jérémie ZOUEU^c

^a Univ, Lille, CNRS, UMR 8524-Laboratoire Paul Painlevé, INRIA-MODAL, F-59000 Lille, France

^b Univ, Lille, CNRS, UMR 9221-Laboratoire Lille Economie Management, INRIA-MODAL, F-59000 Lille, France

^cLaboratoire d'Instrumentation, Image et Spectroscopie, Institut National Polytechnique Félix Houphouët-Boigny, BP 1013, Yamoussoukro, Côte d'Ivoire

Abstract:

Malaria is an infectious and potentially fatal disease which affects nearly half of the world's population. It kills one person every 30 seconds in endemic area. An effective surveillance through early and accurate diagnosis is essential to control this pandemic disease.

This work proposed a multispectral image processing algorithm for automatic malaria diagnosis on label free blood smears. The erythrocytes unsupervised classification process is developed for malaria-infected RBCs (Red Blood Cells) identification in multispectral unlabeled blood smears images. Multispectral images are acquired using the multispectral and multimodal microscope according to 13 wavelengths and 3 modes. A novel segmentation methodology is used to identify and threshold erythrocytes. For each wavelength, 11 texture-based features are generated per RBC according to the geometries. Then, a multivariate functional Principal component analysis (MFPCA) is done on the generated data. The new features obtained (Scores of the MFPCA) are used to characterize the erythrocytes. Two unsupervised clustering algorithms are performed for classification of the RBCs, namely K-means and Hierarchical Ascendant Classification (HAC). We combined the classes of both algorithms to enhance the overall performance of the classification. The original data of the cell's classes determine the spectral signatures of the cells classes as parasitized or health cells. In addition, we explore different way to analyze our data as functional data.

The experimental results showed that the proposed methods have good robustness to detect malaria parasitized cells.

³ To whom all correspondence should be addressed. Email: mekapeusolange.doumunepscoulai.etu@univ-lille.fr

Keywords: Unsupervised Classification; Multispectral imaging; Automatic Malaria Diagnosis, Multivariate analysis, functional data.

I. Introduction:

Malaria disease is caused by parasites of the genus *Plasmodium*. It is a common and critical public health problem which remain prevalent in developing countries. It leads to highest rate of mobility and mortality in endemic areas. Pregnant women and children under age of 5 years are mostly vulnerable [1].

Early and reliable detection of malaria parasite is a way to overcome the disease by an efficient treatment. Also, an accurate count of cells is needed for classifying disease severity as well as drug effectiveness measurement. Gold standard malaria diagnosis tool remains blood smears microscopic examination. However, this manual process of diagnosis is time-consuming, error-prone and depends heavily on extensive training.

However, delay in diagnosis leading to treatment delay can cause death in malaria patients. Therefore, automatic malaria diagnosis system should be developed for early and accurate diagnosis in-order-to effective surveillance of the pandemic.

Nowadays, biologists are assisted by Computer-based systems which improve the diagnosis.

Spectroscopy and imagery are combined in a same technology name multispectral imaging.

The advantages of both technologies are then combined to have robustness in object characterization.

This work describes a computer assisted malaria blood cells characterization and classification based on this multispectral imaging technology.

A LED (Light emitting diode) based multispectral and multimodal microscope used a CMOS (Complementary Metal-Oxide-Semiconductor) camera [2]-[5] has been developed and allows to automatic malaria diagnosis through markers free blood smear images.

Several diagnostic techniques have been used to classify RBCs in those images, mostly based on principal component analysis. Especially, combined hierarchical clustering and Principal Component analysis for differentiation of RBCs by pixels analysis [5]; parasitemia measurement strategies by counting and differentiation parasitized from healthy red blood cells is done [6]; the hemozoin spectral fingerprint is identified ranging from 590 to 700 nm as a malaria diagnosis proof [7]. However, those technique used pixels to differentiate cells or spectral mean of RBCs for the classification therefore they are noise sensitives and the result need to be improved.

Improving the malaria diagnosis using the multispectral and multimodal images that overcomes the limitation of images low contrast quality and taking into account the complexity of those images is the fundamental reason that leads to development of a new algorithm of cells differentiation and classification.

The contribution of this paper is to provide a methodology to differentiate the parasitized from healthy RBCs using unsupervised clustering algorithm.

The proposed methodology consists of three steps of data mining using in computer-based image analyze process. Descriptor's extraction using a set of 11 Haralick texture features for each RBCs according to the wavelengths and the geometries. Those RBCs segmentation process have previously be described [8]. Knowing that, our image is consist of 39 grayscales images, we have for $39 * 11$ descriptors for each of the 236 cells. Therefore, a dimension reduction methodology was implemented through multivariate analysis of functional data. Finally, a so-called hybrid classification algorithm using the two unsupervised classification algorithms is applied on the new cells features. The malaria parasitized and healthy cells clusters determined may be used for supervised classification

algorithm input. This proposed, multimodal and multispectral images, unsupervised classification process can be applied to achieve classification of data which can be resumed as functional data.

II. Materials and methods:

II.1. Material [8]:

The multimodal and multispectral image consists of a set of 39 grayscales images of label-free blood smear have been studied. The images were captured using a multimode, multispectral imaging system developed by our group and has been previously described by Brydegaard [9]. It is a modified commercial metallurgical microscope (Brunel SP80). The ocular was substituted by a 5 Mpx (2592×1944) monochromatic CMOS camera (Guppy-503B, Allied Vision Technology, with a MT9P031 sensor from Micron/ Aptina) equipped with individual pixel size of $2.2 \times 2.2 \mu\text{m}$, each having 12-bit depth. In this system, 13 LEDs were used to selectively illuminate the sample at 13 different wavelengths ranging from UV to near infrared (NIR) (375 to 940 nm) [10]. Blood samples were prepared and delivered from the local clinic in Yamoussoukro (Ivory Coast) by the physicians. No Chemical features were

added to the sample. The peripheral areas of the smear were observed since they exhibit a single layer of RBCs.

II.2. Image acquisition [8]:

The samples are the unstained blood smear from sick patient prepared by using a capillary blood. The process of smearing consists in spreading a drop of blood uniformly over a glass slide, in order to obtain a single layer of cells without any staining. Our data are the grayscale images of those samples made in Transmission, Reflection and Scattering modes. Then, the dark field measurements are made straight on the sample, and an empty glass slide is processed as reference images in those same geometries to satisfy the need of normalization pre-processing step [11]. The images are obtained for each of the three modes from 375 nm to 940 nm (13 wavelengths).

The same scene is automatically captured at each wavelength and in each geometry through an interface (DAQ) developed on LABVIEW 8.6 which controls the current intensities [8], [10], [12]. The spectral image is represented by a 2D matrix (Equation1):

$$I_{xy} = \int_0^{+\infty} S_{sy\lambda} R_{sy\lambda} D_{sy\lambda} d\lambda. \quad (1)$$

where

I_{xy} is a spectral image,

$S_{sy\lambda}$ emission spectrum,

$R_{sy\lambda}$ sample response,

$D_{sy\lambda}$ detector sensibility.

If we consider, the spectral dimension we have a 3D image $I_{xy\lambda}$ for each geometry.

II.3. Pre-processing methodologies:

II.3.1. Normalization:

As described in our second contribution, once the images are saved, the bright ($I_{R\lambda}$) reference images are reconstructed from the sample ($I_{S\lambda}$). Then, the contrast correction and luminosity were performed by dividing the sample image by the bright reference image to obtain the normalized image according to Equation 2.

$$I_{\lambda} = \frac{I_{S\lambda}}{I_{R\lambda}}. \quad (2)$$

where

I_{λ} grayscale image in wavelength λ ;

$I_{S\lambda}$ Image of sample and

$I_{R\lambda}$ Reconstructed bright reference image.

II.3.2. Segmentation:

The segmentation process was previously developed and described in our first contribution [8]. The transmittance images are transformed by statistical standardization then the algorithms of local adaptive thresholding, hysteresis contour closing, and the watershed algorithm are applied to have the single cells which are used in the work (Figure 1).

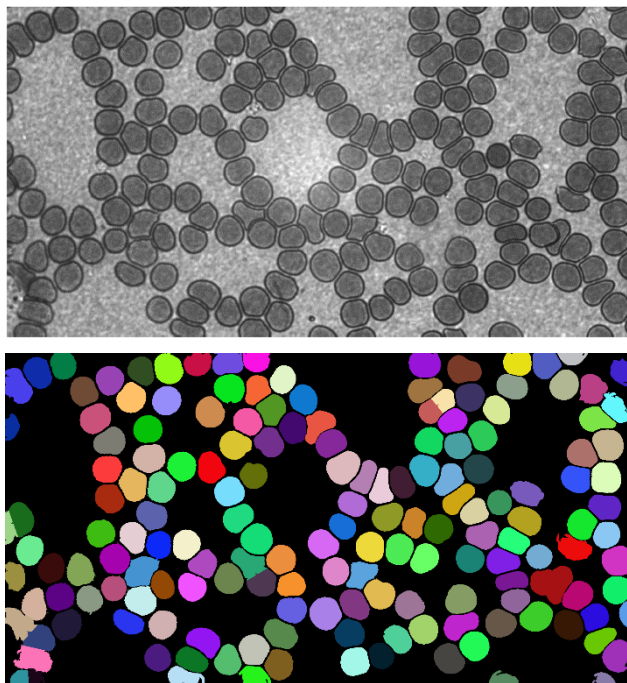
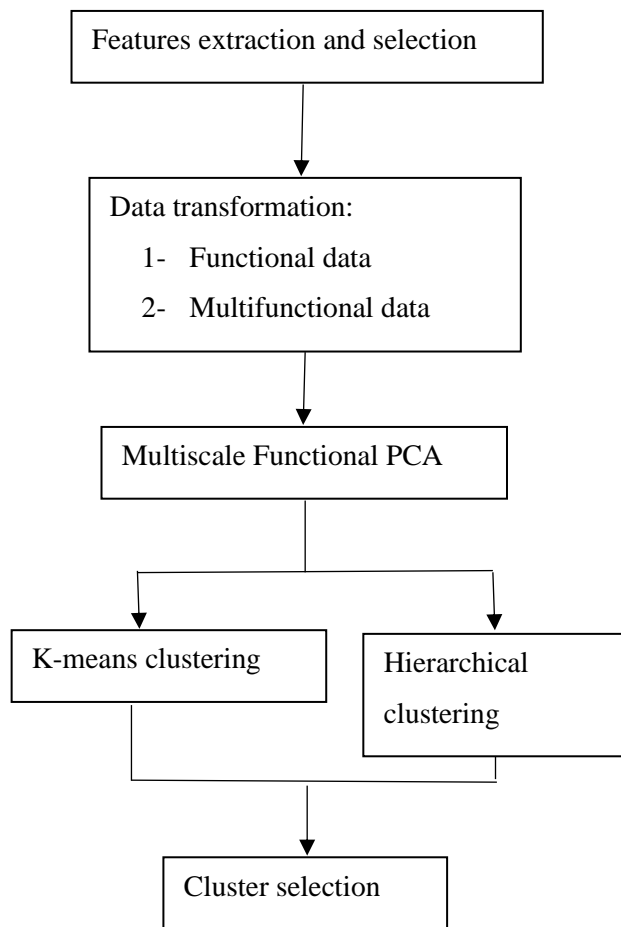


Figure 1: Example of original image top, and segmented and labelled image

II.4. Proposed methodologies:

Algorithm:



II.4.1. Haralick features extraction:

Haralick's features are usually used as texture descriptors in image processing.

The current texture descriptors used in image processing are Haralick features. This set of features is constituted of statistical measurement that consider the spatial dependency of pixels in the image. Known as gray-level spatial

dependence matrix (GLSDM) or gray-level co-occurrence matrix (GLCM), the GLCM functions characterize the RBCs texture by calculating the specified spatial relationship that occur in cell. Those relationship are extracted as statistical measures from this matrix. In this work, we are concerned by 11 invariants in translation and rotation textures features which showed higher accuracies in images analysis [13], [14]-[20]. Those features were computed through EImage package of Bioconductor [21]-[22]; namely Contrast or Inertia, Correlation, Homogeneity or Inverse Difference Moment, Mean, Variance, Entropy, Energy or Angular Second Moment, Difference of Variance, Difference of Entropy, Sum of Entropy and Sum of Variance defined as below:

Let $A = A_{i,j}$; $i, j = 1, \dots, n$, be the unnormalized GLCM elements being the occurrence number of adjacent relations between i and j , two gray scale values of the cell's pixels.

N : Number of gray-levels in cell. 0 is displayed as black and $N-1$ values as white. Let in the following $p_{i,j} = \frac{A_{i,j}}{\sum_{k=1}^N \sum_{l=1}^N A_{k,l}}$ be (i,j) th entry in a normalized gray scale GCLM matrix A of a cell, x denotes the horizontal dimension of A

(x - axis) while y is the vertical dimension (y -axis). We have:

$$p_x(i) = \sum_{j=1}^N p_{i,j},$$

$$p_y(j) = \sum_{i=1}^N p_{i,j},$$

$$\mu_x = \sum_{i=1}^N i * p_x(i),$$

$$\mu_y = \sum_{j=1}^N j * p_y(j),$$

$$\sigma_x^2 = \sum_{i=1}^N (i - \mu_x)^2 * p_x(i),$$

$$\sigma_y^2 = \sum_{j=1}^N (j - \mu_y)^2 * p_y(j),$$

$$p_{x+y}(k) = \sum_{i=1}^N \sum_{j=1}^N \delta_{i+j,k} * p_{i,j},$$

$$k = 2, 3, \dots, 2N.$$

$$p_{x-y}(k) = \sum_{i=1}^N \sum_{j=1}^N \delta_{|i-j|,k} * p_{i,j},$$

$$k = 0, 1, \dots, N-1.$$

where the Kronecker delta function $\delta_{m,n}$ is defined by:

$$\delta_{m,n} = \begin{cases} 1 & \text{when } m = n \\ 0 & \text{when } m \neq n \end{cases}$$

$$\mu_{x+y} = \sum_{k=2}^{2N} k * p_{x+y}(k),$$

$$\mu_{x-y} = \sum_{k=0}^{N-1} k * p_{x-y}(k).$$

The 11 characteristics are:

- 1- Contrast is a second order moment or standard deviation and measures variations in the cell image:

$$Con = \sum_{i=1}^N \sum_{j=1}^N (i-j)^2 p_{i,j}.$$

- 2- Correlation:

$$Cor = \frac{\sum_{i=1}^N \sum_{j=1}^N (i * j) p_{i,j} - \mu_x \mu_y}{\sigma_x \sigma_y}.$$

- 3- Homogeneity

$$Hom = \sum_{i=1}^N \sum_{j=1}^N \frac{1}{1+(i-j)^2} p_{i,j}.$$

- 4- Mean

$$\mu = \frac{\sum_{i=1}^N \sum_{j=1}^N p_{i,j}}{N^2}.$$

- 5- Variance

$$Var = \sum_{i=1}^N \sum_{j=1}^N (i - \mu)^2 p_{i,j}.$$

- 6- Entropy

$$Ent = - \sum_{i=1}^N \sum_{j=1}^N p_{i,j} \log(p_{i,j}).$$

- 7- Difference of Variance

$$Dva = \sum_{i=0}^{N-1} (i - \mu_{x-y})^2 p_{x-y}(i).$$

- 8- Difference of Entropy

$$Den = - \sum_{i=0}^{N-1} p_{x-y}(i) \log(p_{x-y}(i)).$$

- 9- Sum of Entropy

$$Sen = - \sum_{i=2}^{2N} p_{x+y}(i) \log(p_{x+y}(i)).$$

- 10- Sum of Variance

$$Sva = \sum_{i=2}^{2N} (i - \mu_{x+y})^2 p_{x+y}(i).$$

- 11- Energy

$$Energy = \sqrt{(ASM)}.$$

where

ASM= Angular Second Moment

$$ASM = \sum_{i=1}^N \sum_{j=1}^N p_{i,j}^2.$$

II.4.2. Functional data:

We analyze our data within the scope of Multivariate functional data analysis. Functional data analysis (FDA) [24], [25] deals with data that are in the form of functions, curves, images and shapes, or more general complex mathematical objects, thought of as smooth realizations of a stochastic process.

In the last decade, FDA has undergone a large development in a wide variety of fields.

In our monitoring problem, each observation consists of discrete measurements of features extracted at different frequencies.

Consider for n cells denoted i , one observes discrete measurements $x_{i,j,t}$ taken, for a given frequency $t \in T = [T_1, T_2]$ and feature $j = 1, \dots, p$. We assume that these measurements points $x_{i,j,t}$, are noisy observations of a multivariate smooth function $X = (X^1, \dots, X^p)$, $X^k \in L^2(T)$, $k = 1, \dots, p$:

$$x_{i,j,t} = \mu_j(t) + X_i^j(t) + \epsilon_{i,j,t} \quad (3)$$

With μ_j the mean function of the j th feature, $X_i^j(\cdot)$ is a centered squared integrable functional random variable on the frequency

domain T , $\epsilon_{i,j,t}$ are iid zero mean gaussian measurement errors with finite variance. For this representation of discrete measurements in functional dataset, smoothing by expansion in a basis functions may be used.

We aim to classify the cells by unsupervised classification to identify cells clusters with homogeneous characteristics based on the multivariate functional data X_1, \dots, X_n supposed as independent copies of X .

In functional data analysis, it is very practical to perform a cluster analysis based on the functional principal component scores [26] [27] [28]. We use **K**-centers functional clustering based on the calculation of the K first functional principal components scores. These scores are constructed using Multivariate functional principal components [29].

The principle of the method can be described by the following steps by letting for simplicity $p = 1$.

1- Smoothing: the first step when working in FDA is to express raw data by means of

$$X_i^j(t) \approx \sum_{m=1}^P c_{j,m} B_m(t), \quad t \in T$$

Where $B_1(t), \dots, B_p(t)$ are some collection of basis functions (Splines, Fourier, ...).

2- Dimension reduction by principal component analysis.

Let the pointwise empirical functional mean:

$$\bar{X}_{j,n}(t) = \frac{1}{n} \sum_{i=1}^n X_i^j(t)$$

and pointwise empirical covariance function:

$$\hat{c}_{j,n}(t, s) = \frac{1}{n-1} \sum_{i=1}^n (X_i^j(t) - \bar{X}_{j,n}(t)) (X_i^j(s) - \bar{X}_{j,n}(s))$$

For a fixed feature j , we consider the well known standard Karhunen-Loève expansion

$$X_i^j(t) \approx \mu_j(t) + \sum_{k=1}^K \beta_{j,k,i} \phi_{j,k}(t) \quad (3)$$

Where $\phi_{j,k}$'s are the orthonormal eigenfunctions (functional principal components; FPC), in theory $K = \infty$. The functional principal component scores

$\beta_{j,k,i} = \int X_i^j(t) \phi_{j,k}(t) dt'$ are assumed zero-mean and second-order stationary random variables, independent across j .

In practice, the function $\hat{c}_{j,n}(\cdot, \cdot)$ is decomposed in the Eigen basis:

$$\hat{c}_{j,n}(t, s) \approx \sum_{k=1}^K \hat{\lambda}_{k,j} \hat{\Phi}_{j,k}(t) \hat{\Phi}_{j,k}(s),$$

Where $\hat{\lambda}_{1,j} > \dots > \hat{\lambda}_{K,j}$ are the estimated eigenvalues.

One may use the Estimated Functional Principal Components (EFPC's) $\hat{\Phi}_{j,k}$ as basis functions for X_i^j :

$$X_i^j(t) \approx \bar{X}_{j,n}(t) + \sum_{k=1}^K \hat{\beta}_{j,k,i} \hat{\Phi}_{j,k}(t)$$

3- Clustering using the scores:

The $n \times K$ (K to be chosen in practice, here to count 99% of the variation) matrix of estimated scores $\hat{\beta}_{j,k,i} = \int_T (X_i^j(t) - \bar{X}_{j,n}(t)) \hat{\Phi}_{j,k}(t) dt$ are used to do multivariate K-means and hierarchical clustering.

When $p > 1$, then univariate basis expansions for each element of the multivariate functional data is done for each of the p functional components and the matrix scores are derived [24] [29]-[30].

III. Results and discussion:

The proposed methodology is investigated using a set of 39 grayscale images encoded on 16 bits. A core i7 machine running with 16 GB RAM on Microsoft Windows 10 OS is used to perform the process on R software. As presented in our process algorithm, before the classification, the single cell features should be

extracted. Therefore, the red blood cells are segmented and isolate for the feature extraction step. According to the segmentation cell's labelling, the original values of RBCs pixels are extracted from each grey scale image through wavelengths and geometries (Figure 1).

Then, Haralick features extraction method using gray-level spatial dependence matrix of each RBCs is computed. Eleven texture features are selected. We obtain for each of the three geometries and each of the 236 cells, a matrix of features of dimension 11*13 with features in row and frequencies in column. To sum up, a set of 429 invariant texture features is computed per cell. To focus the investigation on the entirety of datapoint from the same curve we choose to analyze our data as a functional data. This is done using funData R [30] package which provides an object-oriented implementation of those type of data. It provides a flexible representation of functional data univariate as well as multivariate. For each mode, the cells are represented as a univariate function of wavelength according to a feature (Figure 2).

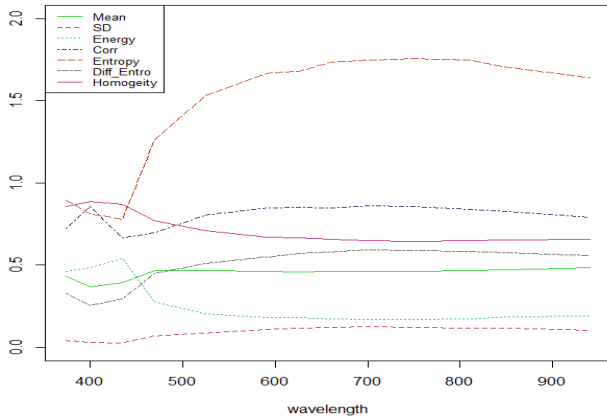


Figure 2: Cell Features spectrum

Consequently, we have a set of 11 functions for each RBCs per mode and wavelength. Features spectrum (Figure 2) shows some of these functions. In object-oriented methodological development, cells are implemented as a “funData” object, where the @argvals slot is a one length list, containing the wavelength which represented observation grid as a vector. The @X slot is a matrix of dimension (number of cells) times (number of wavelengths) containing the features observed values in row-wise format [30]. Multivariate functional principal component analysis is performed on “multiFunData” class of multivariate functional data. Therefore, elements of our “funData” are combined to create this class of data. Three ways of data combination are explored:

- ✓ Individual geometry (transmittance, reflection and scattering)

- ✓ Comparatives analyze combination
- ✓ Conjoined analyze combination

A main goal of this work is the differentiation of RBCs according to whether parasitized or healthy. Hence, K-means and hierarchical clustering algorithms, computed on the result (scores) of multivariate functional principal component are done.

In order to improve the RBCs cluster characterization, the two clustering methods are combined to have a hybrid classification.

To appreciate the results obtained with the proposed technique, results on the three ways of data combination are reported.

III.1. Individual geometry:

Here, geometries are computed individually. Clustering algorithms are performed on the scores of Multivariate Functional Principal Component Analysis. Spectra are projected from the original data of the cells constituted of textures features extracted from the values of cell’s pixels.

A number of 3 clusters by K-means and 2 clusters by HAC are selected since we are looking of parasitized, healthy and intermediate groups of cells.

III.1.1. Transmittance:

A classification tree or dendrogram show us 2 classes of cells in transmittance mode (Figure 3).

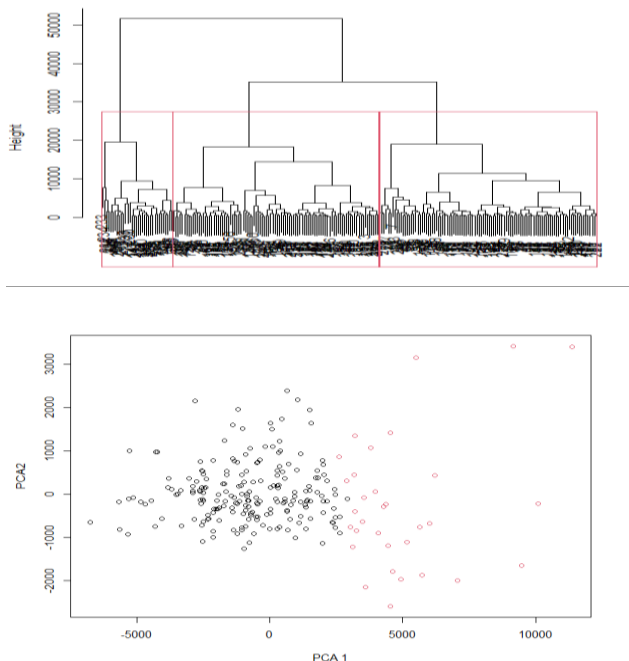


Figure 3: HCA 2 classes for transmittance mode: Dendrogram(top), cell PCA score Plotting(bottom)

This clustering is completed by the scatter plot of K-means groups in the first two functional PCA scores (Figure 4). In the same way, clustering with K-means reveals distinct classes.

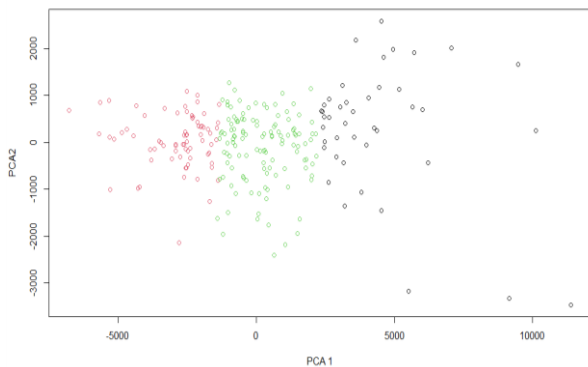


Figure 4: K-means in transmittance mode

However, using K-means in 3 classes reveals interim cluster which should not be overlooked (Figure 4).

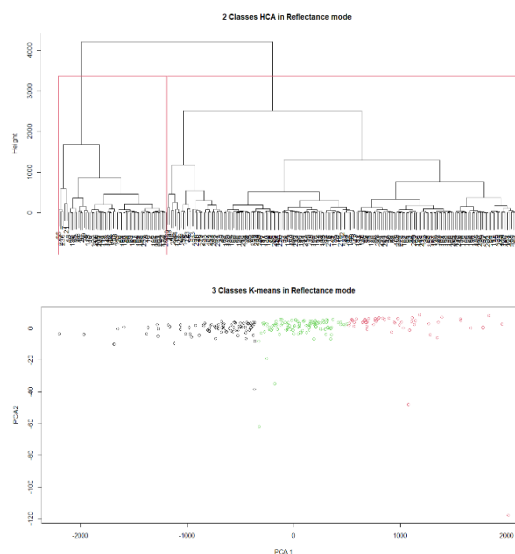


Figure 5: Clustering in Reflectance mode; Top: 2 classes HCA and 3 classes Kmeans (Bottom)

III.1.2. Reflection:

The previews results are confirmed by reflectance which show 3 distinct clusters for K-means classification and 2 clusters for HAC algorithm (Figure 5).

III.1.3. Scattering:

Compared to the other modes, the clustering results in the scattering mode gives less separated groups (Figure 6).

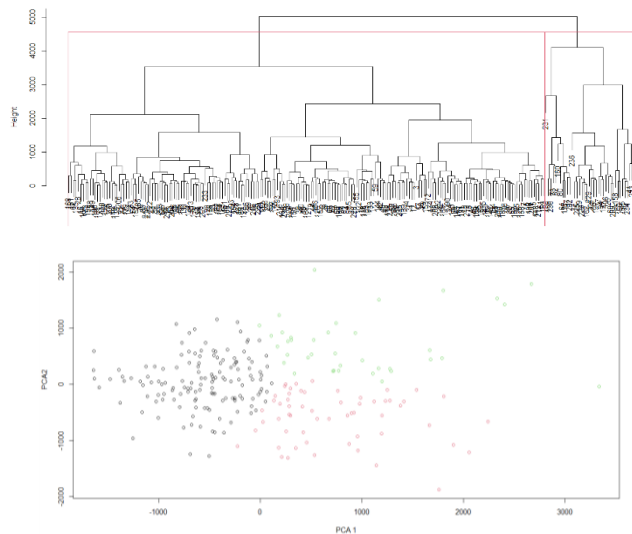


Figure 6: Clustering in scattering mode: HAC (top) and 3 classes K-means (Bottom)

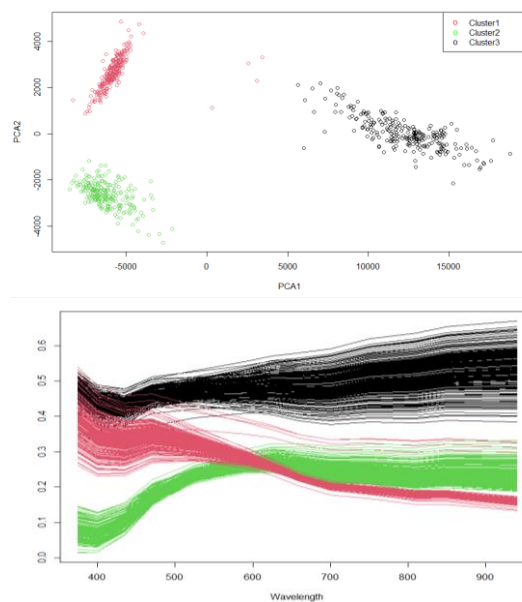


Figure 7: 3 classes K-means for comparative geometries analysis; PCA scores plotting(top); cells spectrum per class

III.2. Comparative analysis of geometries:

Here, each cell is duplicated three times in the sample according to the three geometries. So, we have $236 * 3$ numbers of cells, each cell data is an array of 11 features and 13 frequencies. The sample size is then 708.

A 3 clusters K-means shows 3 distinct clusters corresponding to geometries (Figure 7). This shows that geometries make different and complementary contributions to cells analysis. The 9 clusters K-means, shows 3 distinct clusters per mode (Figure 8). However, the scattering mode clusters are not clearly separate as shown previously in Figure 6.

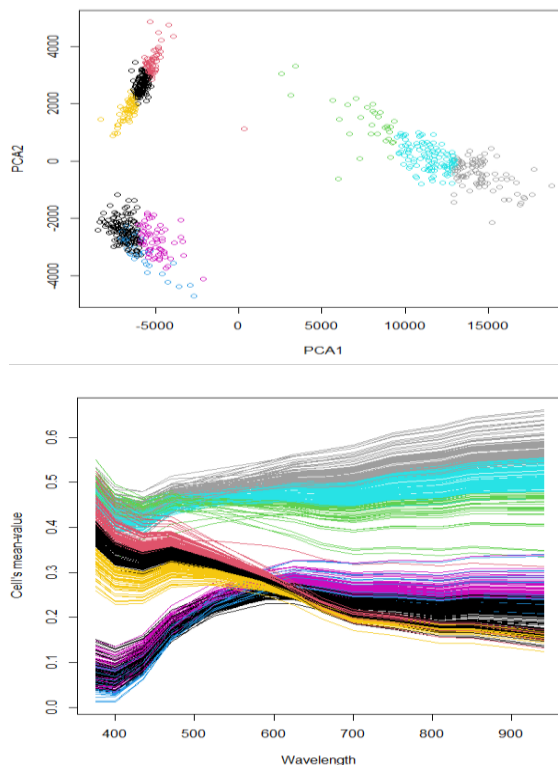


Figure 8: The 9 classes K-means, Cells PCA scores (top), Classes Spectrum (Bottom)

III.3. Joined analyze combination:

In the section, we report the results of the joint analysis of geometries. The sample size is 236. Each cell data consists of an array of $11 \times 13 \times 3$; 11 features, 13 frequencies and 3 geometries. The clustering algorithms, K-means and HAC are computed; Figure 9 shows 3 clusters by K-means for the 3 geometries. In this figure note that:

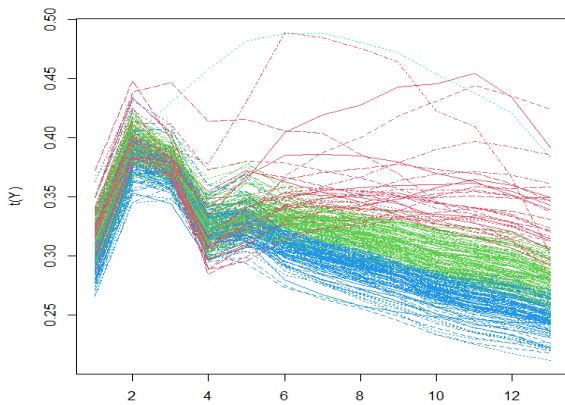


Figure 9: The 3 classes K-means in joints geometries analysis, absorption spectrum

- ✓ The spectra curves in blue and green are characteristics of healthy red blood cells. The differences between these two groups result from the important presence of the scattering. Keeping in mind that there is always a competition between absorption and scattering depending on cell homogeneity and sample preparation. We have blue less

spread, and green more spread but both are healthy.

- ✓ spectra in red characterizes diseased cells. The infection disrupts the homogeneity of the cells and therefore increases the spread. In addition to the spectral parameters characteristic of the infection are found at wavelength 435 nm, 470 nm, and 625 nm. There is the diffusion which disperses the points a little more towards wavelength 700 nm, 750 nm, 810 nm, until 940 nm.

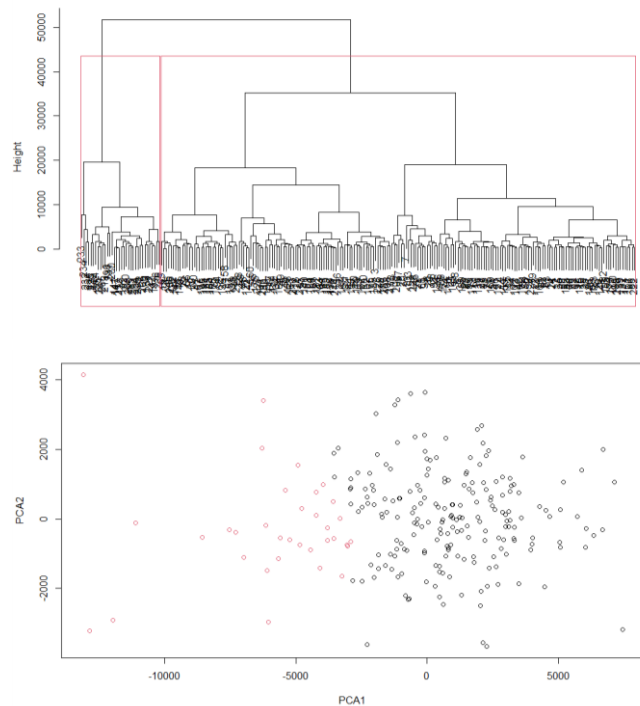


Figure 10: HAC joints geometries

The red cluster of hierarchical clustering (Figure 10 bottom) is similar to K-means red cluster

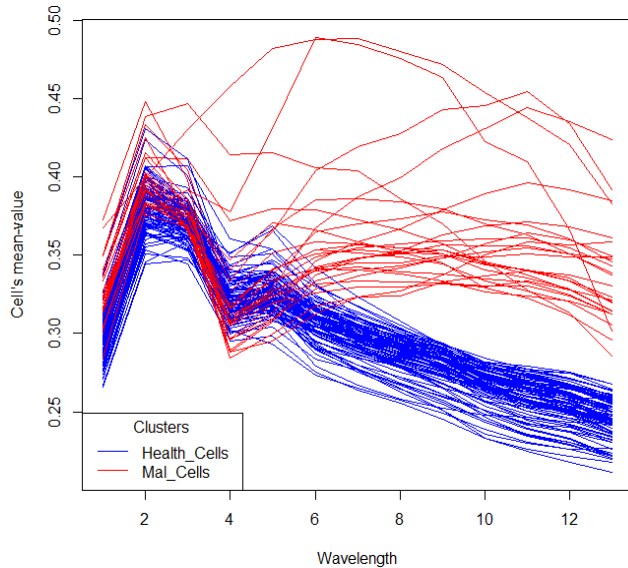


Figure 11: Absorption spectrum of hybrid health and parasitized clusters

(Figure 9). Based on these two types of clustering, a so-called hybrid clustering is chosen to classify the cells in healthy and diseased clusters.

We based our analysis on the healthy cell cluster (blue) which is less spread in K-means clustering (Figure 9) as healthy cells cluster and the red cluster which shows characteristic spectra of diseased cells as parasitized cells cluster. We select cells that are both in the parasitized (red) cluster of K-means (Figure 9) and HAC (bottom of Figure 10). Likewise, the healthy cluster of K-means described previously (curves Blue Figure 9), is combined with the black cluster of hierarchical clustering (bottom of Figure 10) to select cells belonging to both. Our selected healthy and parasitized cells are distinct and

reflect the individual characteristic spectrum of both types of cells according to the cell's mean spectra in absorption mode (Figure 11).

A last aspect of this work not done here is the application of supervised classification algorithms to our clusters (i.e., Partial least squares (PLS), Deep learning, least squares-support vector machines (LS-SVM), back propagation neural network (BPNN)). These prediction methods will be used to classify cells of interim cluster. This will lead determined the parasitemia.

IV. Conclusion and perspectives:

In this work, we explored the richness of multivariate functional analysis apply to multispectral images for blood cells characterization in the context of an automatic and rapid diagnosis of malaria by unsupervised classification.

The parasitized RBCs have been characterized in this work. We can conclude that, malaria diagnosis is possible using our methodology. As perspective, we wish to complete this process of classification by the supervised learning in order to calculate the parasitemia which is essential for the treatment.

Time saving being the goal of an automatic diagnosis, another contribution in this work is to test our algorithm on non-preprocessed images. Of course, a diagnosis without pretreatment of normalization would give rise to an automatic diagnosis in real time (under microscopy).

Acknowledgements

Sida (the Swedish International Development Cooperation Agency) through ISP (the International Science Programme, Uppsala University) is acknowledged for financial support.

- [1] “OMS | Le Rapport de Cette Année En Un Clin d’oeil.” <http://www.who.int/malaria/media/world-malaria-report-2018/fr/>
- [2] O. K. Bagui and J. T. Zoueu, “Red Blood Cells Counting by Circular Hough Transform Using Multispectral Images,” 2014.
- [3] M. Brydegaard, A. Merdasa, H. Jayaweera, J. Ålebring, and S. Svanberg, “Versatile Multispectral Microscope Based on Light Emitting Diodes,” *Rev. Sci. Instrum.*, vol. 82, no. 12, p. 123106, Dec. 2011, doi: 10.1063/1.3660810.
- [4] J. T. Zoueu, S. Ouattara, A. Toure, S. Safi, and S. T. Zan, “Spectroscopic Approach of Multispectral Imaging of Plasmodium Falciparum - Infected Human Erythrocytes,” in *2009 3rd ICTON Mediterranean Winter Conference (ICTON-MW)*, Dec. 2009, pp. 1–7. doi: 10.1109/ICTONMW.2009.5385598.
- [5] J. T. Zoueu and S. T. Zan, “Trophozoite Stage Infected Erythrocyte Contents Analysis by Use of Spectral Imaging LED Microscope,” *J. Microsc.*, vol. 245, no. 1, pp. 90–99, 2012, doi: 10.1111/j.1365-2818.2011.03548.x.
- [6] O. K. Bagui, J. T. Zoueu, and C. Wählby, “AUTOMATIC MALARIA DIAGNOSIS BY THE USE OF MULTISPECTRAL CONTRAST IMAGING,” p. 13, 2015.
- [7] D. L. Omucheni, K. A. Kaduki, W. D. Bulimo, and H. K. Angeyo, “Application of Principal Component Analysis to Multispectral-Multimodal Optical Image Analysis for Malaria Diagnostics,” *Malar J*, vol. 13, p. 485, Dec. 2014, doi: 10.1186/1475-2875-13-485.
- [8] S. Doumun, S. Dabo, and J. Zoueu, “Detection and Segmentation of Erythrocytes in Multispectral Label-Free Blood Smear Images for Automatic Cell Counting,” *J Spectr. Imaging*, vol. 9, Sep. 2020, doi: 10.1255/jsi.2020.a10.
- [9] M. Brydegaard, A. Merdasa, H. Jayaweera, J. Ålebring, and S. Svanberg, “Versatile multispectral microscope based on light emitting diodes,” *Rev. Sci. Instrum.*, vol. 82, no. 12, p. 123106, Dec. 2011, doi: 10.1063/1.3660810.
- [10] A. J. Merdasa, M. Brydegaard, S. Svanberg, and J. T. Zoueu, “Staining-free malaria diagnostics by multispectral and multimodality light-emitting-diode microscopy,” *J. Biomed. Opt.*, vol. 18, no. 3, p. 036002, Mar. 2013, doi: 10.1117/1.JBO.18.3.036002.
- [11] O. K. Bagui and J. T. Zoueu, “Red Blood Cells Counting by Circular Hough Transform Using Multispectral Images,” *Science Alert*, 2014. <https://scialert.net/fulltext/?doi=jas.2014.3591.3594> (accessed Jan. 09, 2020).
- [12] S. Dabo-Niang and J. T. Zoueu, “Combining kriging, multispectral and multimodal microscopy to resolve malaria-infected erythrocyte contents: KRIGING, MULTISPECTRAL AND MULTIMODAL MICROSCOPY TO RESOLVE MALARIA-INFECTED ERYTHROCYTE CONTENTS,” *J. Microsc.*, vol. 247, no. 3, pp. 240–251, Sep. 2012, doi: 10.1111/j.1365-2818.2012.03637.x.
- [13] P. Brynolfsson *et al.*, “Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters,” *Sci. Rep.*, vol. 7, no. 1, Art. no. 1, Jun. 2017, doi: 10.1038/s41598-017-04151-4.
- [14] R. M. Haralick, “Statistical and structural approaches to texture,” *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979, doi: 10.1109/PROC.1979.11328.
- [15] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.
- [16] “Haralick Texture Features.” http://murphy-lab.web.cmu.edu/publication/s/boland/boland_node26.html
- [17] “HaralickTextureExtraction — Orfeo ToolBox 7.2.0 documentation.” https://www.orfeo-toolbox.org/CookBook/Applications/app_HaralickTextureExtraction.html (accessed May 07, 2021).
- [18] T. Löfstedt, P. Brynolfsson, T. Ask Lund, T. Nyholm, and A. Garpebring, “Gray-level invariant Haralick texture features,” *PLOS ONE*, vol. 14, no. 2, p. e0212110, Feb. 2019, doi: 10.1371/journal.pone.0212110.
- [19] A. Wibmer *et al.*, “Haralick Texture Analysis of prostate MRI: Utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason Scores,” *Eur. Radiol.*, vol. 25, no. 10, pp. 2840–2850, Oct. 2015, doi: 10.1007/s00330-015-3701-8.
- [20] N. Zayed and H. A. Elnemr, “Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities,” Oct. 08, 2015.
- [21] “EBImage: Image Processing and Analysis Toolbox for R Version 4.28.1 from

- Bioconductor.”
<https://rdrr.io/bioc/EBImage/>
- [22] G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber, “EBImage—an R Package for Image Processing with Applications to Cellular Phenotypes,” 2010. doi: 10.1093/bioinformatics/btq046.
- [23] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.
- [24] J. Jacques and C. Preda, “Functional data clustering: a survey,” *Adv. Data Anal. Classif.*, vol. 8, no. 3, pp. 231–255, Sep. 2014, doi: 10.1007/s11634-013-0158-y.
- [25] “Functional Data Analysis | James Ramsay | Springer.”
<https://www.springer.com/gp/book/9780387400808> (accessed May 10, 2021).
- [26] “Functional Time Series - ScienceDirect.”
<https://www.sciencedirect.com/science/article/pii/B9780444538581000077> (accessed May 10, 2021).
- [27] S. Hörmann, Ł. Kidziński, and M. Hallin, “Dynamic functional principal components,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 77, no. 2, pp. 319–348, 2015.
- [28] T. Kuenzer, S. Hörmann, and P. Kokoszka, “Principal Component Analysis of Spatially Indexed Functions,” *J. Am. Stat. Assoc.*, pp. 1–13, Feb. 2020, doi: 10.1080/01621459.2020.1732395.
- [29] C. Happ and S. Greven, “Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains,” *J. Am. Stat. Assoc.*, vol. 113, no. 522, pp. 649–659, Apr. 2018, doi: 10.1080/01621459.2016.1273115.
- [30] C. Happ-Kurz, “Object-Oriented Software for Functional Data,” *J. Stat. Softw.*, vol. 93, no. 1, Art. no. 1, Apr. 2020, doi: 10.18637/jss.v093.i05.

Conclusion et Perspectives

7.1 CONCLUSION GENERALE

Le paludisme, ce tueur silencieux n'a cessé de faire des ravages depuis plus 5 décennies qu'il est combattu par une coalition chapeauté par l'OMS sous la coupole de « Roll Back Malaria) constituée d'organismes non gouvernementaux et d'investisseurs privés. Des efforts d'éradication de cette pandémie se sont heurtés à la résistance de l'agent pathogène (l'anophèle). En marge de cette lutte pour l'élimination, La prise en charge rapide par un traitement est salvatrice. En effet, elle permet de réduire l'intensité de la maladie et limite sa transmission. D'où l'importance d'une surveillance efficace par un diagnostic précoce et invariablement exact. Toutefois, des investigations dans la recherche de nouveaux outils de diagnostic n'ont jusque-là pas réussi à surplanter la microscopie qui reste la référence à travers la goutte épaisse et le frottis mince. Ces derniers étant cependant assujettis, à la lenteur d'exécution et à la subjectivité des résultats qui peut varier d'un opérateur à un autre, ainsi qu'aux erreurs d'application des réactifs qui labelisent les cellules malades.

Cette lutte de l'OMS est entièrement partagée par le laboratoire L2IS de Yamoussoukro. Elle est d'ailleurs au cœur de la recherche menée dans ce laboratoire tant au niveau du diagnostic que pour la surveillance en vue de l'éradication de cette maladie. Le microscope multimodal et multispectral fait partie des instruments développés par ce laboratoire et ses partenaires dans le cadre de cette lutte, cette dernière décennie. Le diagnostic par marquage spectral des cellules à travers l'analyse des images multispectrales est le principe de cet instrument. Les travaux réalisés permettent d'établir le diagnostic du paludisme par ce principe. Néanmoins dans le souci de parvenir au diagnostic automatique, rapide et fiable des ajustements ont été nécessaires tant au niveau du temps d'acquisition que de segmentation et de classification. Il était alors impératif d'optimiser ces techniques de traitement et d'analyse des images multispectrales pour parvenir au diagnostic automatique et rapide du paludisme.

L'objectif général de cette thèse est d'établir le diagnostic automatique du paludisme à travers l'analyse d'images multispectrales de frottis non marqué. Plus spécifiquement, nous nous sommes intéressées à la réduction du temps d'acquisition et de traitement, puis

à développer un algorithme de la segmentation d'images multispectrales de faible contraste sans prétraitement, enfin à la classification des cellules selon qu'elles soient saines ou parasités afin d'établir la parasitémie, élément important dont dépend la conduite thérapeutique, dans la prise en charge du patient.

En définitif :

Premièrement, nous avons développé un algorithme de prétraitement de normalisation basé sur la seule image de l'échantillon contrairement de l'équation de normalisation original qui nécessite l'acquisition de trois images (échantillon, référence et le bruit résiduel) pour chacune des images en niveau de gris. Nous avons ainsi réduit le temps d'acquisition des images tout en conservant le processus de prétraitement de normalisation. L'objectif de ce prétraitement étant de soustraire les bruits optiques et électriques, d'assurer l'homogénéité de la luminosité. Nous avons pour cela utilisé au premier abord les méthodes géostatistiques d'analyse variographique et l'estimation de champ aléatoire gaussien pour la reconstitution de l'image de référence basée sur la seule image de l'échantillon. Une deuxième approche de reconstruction a été le rééchantillonnage par la méthode statistique Bootstrap. Aussi, avons-nous adressé dans cette première contribution, les bruits électroniques, se présentant dans l'image acquise sous forme de pixels de hautes intensités, par la méthode de détection et d'imputation des données aberrantes d'Hampel.

Ensuite, nous avons mis en place un processus de détection et de segmentation des érythrocytes, en exploitant la standardisation statistique des images multispectrales en transmission ainsi que les algorithmes de segmentation adaptative, de ligne de partage des eaux et la fermeture de contour par hystérésis.

Et enfin, nous avons classé des cellules selon qu'elles soient parasitées ou non. Ici nous avons utilisé 12 descripteurs de textures des cellules segmentées pour chaque longueur d'onde et chaque géométrie. Une Analyse en Composantes Principales Fonctionnelles Multivariée a ensuite été effectuée sur les données ainsi constituées. Puis une classification non supervisée, par les algorithmes de partitionnement K-moyennes et Classification Ascendante Hiérarchique (CAH) permet d'isoler les classes de cellules parasitées et saines. Lesquelles seront ensuite utilisée pour une classification supervise.

Pour ces traitements le logiciel R a utilisé sur un ordinateur portable de marque HP core i5, 8 Go de RAM sous Windows 10.

7.2- DISCUSSION

Notre première contribution a permis de réduire le temps d'acquisition de 66%, mais encore, notre solution est très rapide et s'exécute en moins de 20 secondes. De plus, la comparaison de nos images ainsi normalisées avec celles de la méthode usuelle montre une nette amélioration de contraste. En revanche, la reconstruction par la méthode géostatistique utilisant l'analyse de variographique qui permet conserver la structure spatiale de l'image de fond est plus gourmand en temps. Cela est dû à l'analyse variographique qui en est l'essence.

S'agissant de notre deuxième contribution, elle s'avère très efficace et efficiente avec une sensibilité de 98,47% et de spécificité 98,23%, un degré de précision de 98,34% devant les méthodes cinq méthodes de l'état de l'art. Par ailleurs, elle est rapide et s'exécute en moins de 2mn ce qui conduirait à un diagnostic en temps réel. Excepté, le fait que les images prétraitées par le filtrage médian présentent un lissage qui rend difficile la fermeture des contours et ainsi leur segmentation. Aussi pour avoir un processus automatique, un paramétrage de la taille des cellules est nécessaire.

Pour ce qui est de la classification, l'utilisation de 11 descripteurs invariants en translation et en rotation nous a permis de caractériser chaque cellule de façon unique. De plus la caractérisation fonctionnelle de ces descripteurs offre une richesse à l'analyse qui est effectué. Une modélisation conjointe avec les algorithmes des K-moyennes et de Classification Ascendante Hiérarchique a mis en évidence les deux classes bien distinctes de cellules parasitées et cellules saines. Ce résultat est bien confirmé par les spectres moyens de ces groupes de cellules puisqu'il s'agit ici d'un diagnostic par analyse spectre.

7.3 QUELQUES PERSPECTIVES :

En perspective à cette thèse, nous envisageons perfectionner la classification par d'autres méthodes de classification non-supervisée en complément des algorithmes de K-moyennes et de Classification Ascendante Hiérarchique. Une autre approche sera de comparer nos résultats aux méthodes de classification multivariées classiques (non fonctionnelles).

L'apprentissage supervisé nous permettra d'évaluer les performances des différents algorithmes et d'effectuer une étude comparative afin d'en retenir la plus performante qui servira à classer les cellules non prises en compte dans la classification non supervisée. Ceci afin pouvoir calculer la parasitémie dont dépend la conduite thérapeutique.

Pour ce qui concerne le prétraitement de normalisation, la prise en compte de la dimension spatiale devra être faite à travers une combinaison des approche Bootstrap et géostatistique.

Pour parachever ce travail, nous envisageons de déployer notre solution sur le microscope multimodal et multispectral. Cette solution sera basée sur notre deuxième contribution qui est algorithme de segmentation en temps réel combinée à une classification à l'aide de la troisième contribution et de l'algorithme de classification supervisée qui sera sélectionné.

Références bibliographiques

- [1] A. Khermach, H. Khalki, L. Louzi, A. Zinebi, K. Moudden, and M. Elbaaj, "Perturbations biologiques au cours du paludisme: à propos de trente cas," *Pan Afr Med J*, vol. 26, Mar. 2017, doi: 10.11604/pamj.2017.26.174.9008.
- [2] "world-malaria-report-2020-briefing-kit-fre.pdf." Accessed: Apr. 10, 2021. [Online]. Available: https://www.who.int/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2020-briefing-kit-fre.pdf?sfvrsn=69c55393_7.
- [3] "OMS | Le Rapport Sur Le Paludisme Dans Le Monde 2019 En Un Clin d'oeil." <http://www.who.int/malaria/media/world-malaria-report-2019/fr/>.
- [4] "L'OMS appelle à relancer la lutte contre le paludisme." <https://www.who.int/fr/news/item/30-11-2020-who-calls-for-reinvigorated-action-to-fight-malaria> (accessed Apr. 08, 2021).
- [5] "RBM_AIM_Report_A4_FRENCH_FINAL.pdf." Accessed: Nov. 28, 2020. [Online]. Available: https://endmalaria.org/sites/default/files/uploads/2017/07/RBM_AIM_Report_A4_FRENCH_FINAL.pdf.
- [6] "World malaria report 2019." <https://www.who.int/publications-detail-redirect/9789241565721> (accessed Nov. 28, 2020).
- [7] "Performance des tests de diagnostic rapide du paludisme: Bilan des résultats d'évaluation des produits par l'OMS: Séries 1-5 (2008-2013)," p. 36.
- [8] L. Karkowski, P. Perez, M. Cabon, G. Cinquetti, and F. Dutasta, "Intérêt de la PCR paludisme en pratique quotidienne. Étude rétrospective dans un service de maladies infectieuses," *La Revue de Médecine Interne*, vol. 36, p. A107, Dec. 2015, doi: 10.1016/j.revmed.2015.10.027.
- [9] M. Brydegaard, Z. Guan, and S. Svanberg, "Broad-band multispectral microscope for imaging transmission spectroscopy employing an array of light-emitting diodes," 2009, doi: 10.1119/1.3027270.
- [10] M. Brydegaard, A. Merdasa, H. Jayaweera, J. Ålebring, and S. Svanberg, "Versatile multispectral microscope based on light emitting diodes," *Review of Scientific Instruments*, vol. 82, no. 12, p. 123106, Dec. 2011, doi: 10.1063/1.3660810.
- [11] « • Coronavirus : nombre d'infections par pays dans le monde 2021 | Statista ». <https://fr.statista.com/statistiques/1091585/morts-infections-coronavirus-monde/> (consulté le avr. 10, 2021).
- [12] « Maladie à virus Ebola ». <https://www.who.int/fr/news-room/fact-sheets/detail/ebola-virus-disease> (consulté le avr. 10, 2021).
- [13] « OMS | Le Rapport sur le paludisme dans le monde 2019 en un clin d'oeil », WHO. <http://www.who.int/malaria/media/world-malaria-report-2019/fr/> (consulté le juill. 30, 2020).
- [14] « WHO | WHO commends the Roll Back Malaria Partnership's contribution to global progress as governing board disbands secretariat », WHO. <https://www.who.int/malaria/news/2015/governing-board-disbands-rbm-secretariat/en/> (consulté le nov. 20, 2020).
- [15] « world-malaria-report-2020-briefing-kit-fre.pdf ». Consulté le: avr. 10, 2021. [En ligne]. Disponible sur: https://www.who.int/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2020-briefing-kit-fre.pdf?sfvrsn=69c55393_7.
- [16] « L'OMS appelle à relancer la lutte contre le paludisme ».

- <https://www.who.int/fr/news/item/30-11-2020-who-calls-for-reinvigorated-action-to-fight-malaria> (consulté le avr. 08, 2021).
- [17] « Guide-Lutte-contre-le-Paludisme-en-entreprise.pdf ». Consulté le: avr. 10, 2021. [En ligne]. Disponible sur: <http://www.pnlp.sn/wp-content/uploads/2018/02/Guide-Lutte-contre-le-Paludisme-en-entreprise.pdf>.
- [18] « MCP_rbm_pi_rpt_6_fr.pdf ». Consulté le: avr. 10, 2021. [En ligne]. Disponible sur: https://path.azureedge.net/media/documents/MCP_rbm_pi_rpt_6_fr.pdf.
- [19] « Epidémiologie du paludisme chez les personnes travaillant sur des sites d'orpaillage illégal en Guyane : Quels enjeux pour la santé publique ? - TEL - Thèses en ligne ». <https://tel.archives-ouvertes.fr/tel-01570446> (consulté le avr. 11, 2021).
- [20] « OMS | Le Rapport Sur Le Paludisme Dans Le Monde 2019 En Un Clin d'oeil ». <http://www.who.int/malaria/media/world-malaria-report-2019/fr/>.
- [21] « Médecine tropicale ». <http://medecinetropicale.free.fr/enseignement.html> (consulté le avr. 11, 2021).
- [22] S. Kasetsirikul, J. Buranapong, W. Srituravanich, M. Kaewthasorn, et A. Pimpin, « The development of malaria diagnostic techniques: a review of the approaches with focus on dielectrophoretic and magnetophoretic methods », *Malar. J.*, vol. 15, no 1, Art. no 1, déc. 2016, doi: 10.1186/s12936-016-1400-9.
- [23] « Staining-Free Malaria Diagnostics by Multispectral and Multimodality Light-Emitting-Diode Microscopy | Semantic Scholar ». <https://www.semanticscholar.org/paper/Staining-free-malaria-diagnostics-by-multispectral-Merdasa-Brydegaard/93f5cd011a7aa6b9bf1a840c232174803446c790>.
- [24] M. Brydegaard, Z. Guan, et S. Svanberg, « Broad-band multispectral microscope for imaging transmission spectroscopy employing an array of light-emitting diodes », 2009, doi: 10.1119/1.3027270.
- [25] H. Yu et al., « Malaria Screener: a smartphone application for automated malaria screening », *BMC Infect. Dis.*, vol. 20, no 1, p. 825, nov. 2020, doi: 10.1186/s12879-020-05453-1.
- [26] A. D. Oliveira et al., « The Malaria System MicroApp: A New, Mobile Device-Based Tool for Malaria Diagnosis », *JMIR Res. Protoc.*, vol. 6, no 4, avr. 2017, doi: 10.2196/resprot.6758.
- [27] « Evaluation of a Novel Magneto-Optical Method for the Detection of Malaria Parasites ». <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0096981> (consulté le avr. 14, 2021).
- [28] « Enhancing malaria diagnosis through microfluidic cell enrichment and magnetic resonance relaxometry detection | Scientific Reports ». <https://www.nature.com/articles/srep11425> (consulté le avr. 14, 2021).
- [29] « Micromagnetic resonance relaxometry for rapid label-free malaria diagnosis | Nature Medicine ». <https://www.nature.com/articles/nm.3622> (consulté le avr. 14, 2021).
- [30] « OMS | Accès universel aux tests diagnostiques du paludisme. Manuel pratique », WHO. <http://www.who.int/malaria/publications/atoz/9789241502092/fr/> (consulté le avr. 17, 2021).
- [31] M. Brydegaard and Lunds universitet, *Aspects of Optical Broad Band Spectroscopy and Information Extraction - Applications in Medicine and Ecology*. 2016.

- [32] M. Brydegaard, Z. Guan, and S. Svanberg, "Broad-band multispectral microscope for imaging transmission spectroscopy employing an array of light-emitting diodes," 2009, doi: 10.1119/1.3027270.
- [33] L. D. Omucheni and U. O. Nairobi, "Multispectral Imaging Of Human Blood Media Applied To Malaria Diagnostics," p. 115.
- [34] A. J. Merdasa, M. Brydegaard, S. Svanberg, and J. T. Zoueu, "Staining-free malaria diagnostics by multispectral and multimodality light-emitting-diode microscopy," *JBO, JBOPFO*, vol. 18, no. 3, p. 036002, Mar. 2013, doi: 10.1117/1.JBO.18.3.036002.
- [35] M. Brydegaard, A. Merdasa, H. Jayaweera, J. Ålebring, and S. Svanberg, "Versatile multispectral microscope based on light emitting diodes," *Review of Scientific Instruments*, vol. 82, no. 12, p. 123106, Dec. 2011, doi: 10.1063/1.3660810.
- [36] O. K. Bagui, J. T. Zoueu, and C. Wählby, "AUTOMATIC MALARIA DIAGNOSIS BY THE USE OF MULTISPECTRAL CONTRAST IMAGING," p. 13, 2015.
- [37] A. J. Cox, A. J. DeWeerd, and J. Linden, "An experiment to measure Mie and Rayleigh total scattering cross sections," *American Journal of Physics*, vol. 70, no. 6, pp. 620–625, May 2002, doi: 10.1119/1.1466815.
- [38] A. J. Cox, A. J. DeWeerd, and J. Linden, "An Experiment to Measure Mie and Rayleigh Total Scattering Cross Sections," *American Journal of Physics*, vol. 70, no. 6, pp. 620–625, May 2002, doi: 10.1119/1.1466815.
- [39] S. Tominaga and B. A. Wandell, "Component estimation of surface spectral reflectance," *J. Opt. Soc. Am. A, JOSAA*, vol. 7, no. 2, pp. 312–317, Feb. 1990, doi: 10.1364/JOSAA.7.000312.
- [40] Y. M. Serebrennikova, J. M. Smith, D. E. Huffman, G. F. Leparc, and L. H. García-Rubio, "Quantitative interpretations of Visible-NIR reflectance spectra of blood," *Opt. Express, OE*, vol. 16, no. 22, pp. 18215–18229, Oct. 2008, doi: 10.1364/OE.16.018215.
- [41] D. Chicea and I. Turcu, "A RANDOM WALK MONTE CARLO APPROACH TO SIMULATE MULTIPLE LIGHT SCATTERING ON BIOLOGICAL SUSPENSIONS," p. 8.
- [42] "Absorption and Scattering of Light by Small Particles | Wiley," Wiley.com. <https://www.wiley.com/en-us/Absorption+and+Scattering+of+Light+by+Small+Particles-p-9780471293408> (accessed Jan. 21, 2021).
- [43] D. E. Huffman, Y. M. Serebrennikova, J. M. Smith, G. F. Leparc, and L. H. García-Rubio, "New method for the detection of micro-organisms in blood: application of quantitative interpretation model to aerobic blood cultures," *J Biomed Opt*, vol. 14, no. 3, p. 034043, Jun. 2009, doi: 10.1117/1.3156837.
- [44] J. G. Kim, M. Xia, and H. Liu, "Extinction coefficients of hemoglobin for near-infrared spectroscopy of tissue," *IEEE Eng Med Biol Mag*, vol. 24, no. 2, pp. 118–121, Apr. 2005, doi: 10.1109/memb.2005.1411359.
- [45] Y. Zhao, L. Qiu, Y. Sun, C. Huang, and T. Li, "Optimal hemoglobin extinction coefficient data set for near-infrared spectroscopy," *Biomed Opt Express*, vol. 8, no. 11, pp. 5151–5159, Oct. 2017, doi: 10.1364/BOE.8.005151.

- [46] E. K. Hanson and J. Ballantyne, "A blue spectral shift of the hemoglobin soret band correlates with the age (time since deposition) of dried bloodstains," *PLoS One*, vol. 5, no. 9, p. e12830, Sep. 2010, doi: 10.1371/journal.pone.0012830.
- [47] "Optical properties of circulating human blood in the wavelength range 400-2500 nm." <https://www.spiedigitallibrary.org/journals/journal-of-biomedical-optics/volume-4/issue-01/0000/Optical-properties-of-circulating-human-blood-in-the-wavelength-range/10.1117/1.429919.full> (accessed Jan. 13, 2021).
- [48] O. K. Bagui, J. T. Zoueu, and C. Wählby, "AUTOMATIC MALARIA DIAGNOSIS BY THE USE OF MULTISPECTRAL CONTRAST IMAGING," p. 13, 2015.
- [49] G. S. Noland, N. Briones, and D. J. Sullivan, "The shape and size of hemozoin crystals distinguishes diverse *Plasmodium* species," *Molecular and Biochemical Parasitology*, vol. 130, no. 2, pp. 91-99, Aug. 2003, doi: 10.1016/S0166-6851(03)00163-4.
- [50] "Optical Absorption of Hemoglobin." <https://omlc.org/spectra/hemoglobin/index.html> (accessed Jan. 14, 2021).
- [51] B. K. Wilson, M. R. Behrend, M. P. Horning, and M. C. Hegg, "Detection of malarial byproduct hemozoin utilizing its unique scattering properties," *Opt. Express*, OE, vol. 19, no. 13, pp. 12190-12196, Jun. 2011, doi: 10.1364/OE.19.012190.
- [52] BAGUI K. OLIVIER. «Developpement d'outils d'analyse d'images de microscope multispectrales et multimodales à diodes électroluminescentes: application à l'étude et au diagnostic automatique du paludisme.» 2015.

Résumé

Titre : Contribution au développement des stratégies d'analyse d'images multispectrales pour l'aide au diagnostic automatique et rapide du paludisme

Mot clés :

Imagerie multispectrale, paludisme, diagnostic automatique, segmentation, Géostatistique, ACP multifonctionnelle, classification

Résumé :

Une surveillance efficace par un diagnostic précoce et exact est cruciale pour la lutte contre la pandémie du paludisme. En effet, le paludisme est une maladie infectieuse qui touche près de la moitié de la population mondiale et tue une personne toutes les 30 secondes dans les zones endémiques.

Dans ce travail, nous présentons une nouvelle méthodologie de diagnostic basée sur l'imagerie multispectrale. La première contribution permet de réduire le temps d'acquisition des images de cellules sanguines tout en conservant le processus de pré-traitement de normalisation. L'objectif de ce pré-traitement est de soustraire les bruits optiques et électriques, d'assurer l'homogénéité de la luminosité. Nous nous appuyons, pour cela sur des méthodes de statistique spatiale (géostatistiques) pour la reconstitution de l'image de référence basée sur la seule image de l'échantillon. Une deuxième approche de reconstruction est le ré-échantillonnage par Bootstrap. Aussi, avons-nous adressé les bruits électroniques par la méthode de détection et d'imputation des données aberrantes d'Hampel. La solution proposée est très rapide et s'exécute en moins de 20 secondes. Nous avons ensuite mis en place un processus de détection des érythrocytes, en exploitant la standardisation statistique des images multispectrales en transmission, des algorithmes de segmentation adaptative, de ligne de partage des eaux et la fermeture de contour par hystérésis.

Afin d'isoler des cellules parasitées, une procédure de classification est proposée sur 12 descripteurs de textures de cellules segmentées pour plusieurs longueurs d'onde et trois géométries (diffusion, réflexion et transmission). L'analyse en composantes principales

pour données fonctionnelles a été effectuée sur les données ainsi constituées avant une classification non supervisée, par les algorithmes de partitionnement K-moyennes et classification ascendante hiérarchique (CAH) permettant d'isoler les classes de cellules parasitées et saines. Lesquelles sont ensuite utilisées pour une classification supervisée des cellules.

Title: Contribution to the development of multispectral image analysis strategies for automatic and rapid diagnosis of malaria

Keywords: Multispectral imaging, segmentation, malaria, automatic diagnosis, Geostatistic, classification

Abstract

Malaria is an infectious and potentially fatal disease which affects nearly half of the world's population. It kills one person every 30 seconds in endemic area. An effective surveillance through early and accurate diagnosis is essential to fight against this pandemic.

This work proposes a new malaria diagnosis methodology based on multispectral imaging. Our first contribution is an algorithm that reduces the acquisition time of blood cell images while maintaining the normalization preprocessing. The normalization is an essential preprocessing step in the diagnostic which ensure uniformity of brightness in the images. The method is based on estimation of the Bright reference image which represents luminosity and contrast variability function from the background part of the image. This is accomplished by two different re-sampling methodologies namely gaussian random field simulation by a geostatistical method (variogram analysis) and Bootstrap re-sampling. To avoid intensity saturation problem of some pixels, we addressed the electronic noise by Hampel's Outlier Detection and Imputation Method. The proposed solution is very fast and takes less than 20 seconds.

The second contribution is about an erythrocyte detection and segmentation process; taking advantage of Beer Lambert's law by using first statistical standardization equation applied to transmittance, the local adaptive threshold algorithms, watershed algorithm and contour closure by hysteresis.

To identify parasitized RBCs, a classification process is performed on 12 textures descriptors of segmented cells for 13 wavelengths and three geometries (Scattering, Reflection and Transmission). A multivariate functional principal component analysis was performed on the data and then unsupervised classification, by K-means partitioning algorithms and hierarchical ascending classification (CAH) allowing to isolate the classes

of parasitized and healthy cells. Finally, the supervised classification is done by machine learning algorithms.

To summarize, our process reduced the acquisition time to a third from 12 to 4 minutes. Also, segmentation and classification processes, both take about 8 minutes, this could lead to real-time diagnosis. In addition, the segmentation process requires any preprocessing. It thus has an advantage in considering low contrast images.

In addition, our methods outperform state-of-the-art methods of multispectral and multimodal unstained blood smears images in terms of contrast enhancement, similarity index (SSIM) and mean squared error (MSE) according to normalization methodology. Segmentation recorded a precision of 98.47% and Recall of 98.23%, a degree of precision (F-Measurement) of 98.34% and an accuracy of 96.75%.

