

UNIVERSITÉ DE LAUSANNE
ÉCOLE DES SCIENCES CRIMINELLES

PROJET DE MASTER EN SCIENCES FORENSIQUES
ORIENTATION INVESTIGATION ET IDENTIFICATION NUMÉRIQUES

Interprétation des attributions de caméra source

Interpretation of Source Camera Attribution

Hannes Spichiger

Sous la supervision de
Prof. Franco Taroni

21 août 2017

Résumé L'identification de la caméra source, c'est-à-dire celle qui a été utilisée pour prendre une image, est une question qui se pose de plus en plus devant les tribunaux. Afin de répondre à cette question, les tribunaux peuvent faire recours à des experts qui ont à disposition toute une gamme d'approches pour analyser et comparer des images numériques. Basé sur les connaissances des sciences forensiques classiques, ce travail propose une approche Bayésienne pour l'évaluation des résultats obtenus. Plusieurs situations et types de traces sont traitées à l'aide des réseaux Bayésiens.

Mots-clés : Interprétation, caméra source, réseaux Bayésiens

Abstract Source camera identification, the identification of the camera that has been used to take a given image, is a question asked more and more often in court. In order to answer this question, often an expert is tasked to analyze the picture of interest. To do so, the expert has a wide range of approaches, based on a variety of image characteristics. This thesis proposes a Bayesian approach to evaluate the result of the formentioned analysis by establishing Bayesian networks for several different situations.

Keywords : Interpretation, source camera, Bayesian networks

Des remerciements sont adressés à tout le monde qui m'a soutenu pendant ce travail, en particulier

Mon superviseur Prof. Franco Taroni pour le suivi tout le long du travail

Le Prof. Eoghan Casey pour son engagement à établir des contacts avec des experts

Le Prof. Zeno Geradts du *Netherlands Forensic Institute (NFI)* qui a pris le temps de répondre à mes questions

Coraline Henri et Jonathan Bourquenoud pour la relecture du travail

Et très spécialement, mes parents et Jorina Marti, qui m'ont soutenu à tout moment et qui m'ont permis de démêler plusieurs des problématiques rencontrées.

Table des matières

1	Introduction	1
2	Caractéristiques d'une image numérique	4
2.1	Terminologie	4
2.2	Traces du système optique	5
2.2.1	Poussières	5
2.2.2	Distorsion des lentilles	5
2.3	Traces du capteur	5
2.3.1	Photo Response Non-Uniformity (PRNU)	5
2.3.2	Pixels morts et pixels chauds	6
2.4	Balises EXIF	7
3	L'évaluation Bayésienne	9
3.1	Notation	9
3.2	Le théorème de Bayes	9
3.3	Utilisation du théorème de Bayes dans le cadre légal	10
3.4	Les réseaux Bayésiens	11
4	Observation d'une caractéristique unique	14
4.1	Structure du réseau Bayésien	14
4.2	Le nœud P : les propositions	14
4.3	Le nœud E : l'observation faite par l'expert	15
4.3.1	Observation d'une caractéristique discrète sans variabilité intra-source	16
4.3.2	Observations des caractéristiques avec potentiel de variabilité intra-source	18
4.3.3	Observation d'une caractéristique continue	19
5	Observation simultanée de plusieurs caractéristiques	22
5.1	Structure du réseau Bayésien	22

5.2	Le nœud E_2 : la deuxième observation	23
5.2.1	E_2 discrète	23
5.2.2	E_2 continue	25
5.3	Ordre de traitement des caractéristiques	26
5.4	Plus que deux caractéristiques	27
6	Prise en compte de la possibilité de modifications	28
6.1	Structure du réseau	28
6.2	Le nœud M : la modification de la caractéristique	29
6.3	Le nœud OE : la valeur observée	30
7	Réseau pour plusieurs observations avec modification	34
7.1	Structure du réseau	34
7.2	Le nœud M_2 : la modification de la deuxième caractéristique	35
7.3	Le nœud OE_2 : la valeur observée de la deuxième caractéristique	36
8	Conclusion	39
9	Bibliographie	41

1 Introduction

Avec la baisse des prix des caméras digitales, leur intégration dans des appareils mobiles et la facilité de distribution des images par internet, les dernières années ont vu une hausse importante d'images numériques produites dans de nombreuses situations. Par conséquent, de plus en plus d'images numériques sont utilisées en tant que moyen de preuve dans le cadre légal [KOT et CAO 2013]. Dans cette perspective se pose fréquemment la question de l'origine des images : soit pour corroborer l'histoire d'un témoin qui a pris l'image d'une situation faisant objet d'un procès pénal, soit pour résoudre la question de droit d'auteur dans une affaire civile. Dans des affaires d'abus sur des enfants, la question de la provenance est d'intérêt pour déterminer si l'inculpé a seulement retrouvé des images sur internet ou s'il les a produites lui-même, et dans de nombreuses affaires où l'auteur a pris des images de son acte, ajoute un élément de preuve très puissant à l'enquête.

Il n'est donc pas surprenant que les dernières années aient donné naissance à de nombreux travaux proposant des techniques d'identification de la caméra source, c'est-à-dire la caméra qui a été utilisée pour prendre l'image en question. Aujourd'hui, un expert face à un cas comme décrit ci-dessus peut choisir parmi une multitude de techniques différentes déclinées en plusieurs variantes en fonction de la situation. Au final, l'expert se retrouve avec des observations et des résultats, lui donnant une idée de la similitude entre les images incriminées et un nombre d'images de comparaison issus de la caméra d'intérêt. C'est avec ces résultats que l'expert se présente au tribunal où lui est posé la question centrale : «Est-ce que cette caméra a été utilisée pour faire l'image en question ?»

La question paraît simple. Suite à l'analyse, l'expert sait que l'image partage un certain nombre de similarités avec les images faites par la caméra d'intérêt. Peut-être, il sait de son expérience que cette similarité est plus haute que ce que l'on s'attend avec des caméras comme celle-ci ou il a fait des expérimentations pour tester ses résultats. Mais, en fin de compte, est-ce que les informations sont suffisantes pour dire que c'était cette caméra qui a fait l'image ? Peut-il vraiment en être sûr ?

D'un point de vue scientifique, le seul moyen pour l'expert de dire que, effectivement, la caméra d'intérêt a été utilisée pour prendre une certaine image, est de décider à un moment donné qu'un taux de similitudes suffisant est atteint. [BIEDERMANN, BOZZA et TARONI 2008] Mais que veut dire, «suffisant»? Est-ce que ce ne serait pas plutôt une question qui devrait être répondue par le tribunal et non pas par l'expert? Si oui, comment le tribunal est-il censé connaître les similitudes attendues entre des images issues des mêmes caméras et entre celles issues de caméras différentes?

Ce dont l'expert a besoin est un moyen d'exprimer au tribunal la signification de ses résultats sur la base de ses expériences, connaissances et expérimentations, sans que la décision finale ne soit prise par lui. Il a besoin d'un outil qui lui permet d'intégrer les niveaux de confiance et d'incertitude qu'il a dans son résultat. Dans l'interprétation Bayésienne, l'expert peut trouver un tel outil.

L'approche Bayésienne est déjà bien établie dans les sciences forensiques classiques. Pour la presque totalité des domaines en science forensique, des solutions Bayésiennes pour l'évaluation existent ainsi que de nombreuses recherches qui fournissent des valeurs de probabilité formant une base pour l'évaluation. De plus en plus des réseaux Bayésiens, des modélisations graphiques des dépendances entre toutes les variables en jeu, sont utilisés pour traiter des problématiques plus complexes. [TARONI et al. 2014]

Le but de ce travail est de proposer un cadre théorique pour effectuer une évaluation Bayésienne des résultats obtenus dans le cas d'une expertise d'identification de caméra sur la base d'une image. Plus précisément, des réseaux Bayésiens permettant de trouver une réponse à la question, «Est-ce que cette image a été prise par cette caméra?» sont proposés. Tout cela sera fait en travaillant par analogie aux traces physiques, en reprenant des modèles établis dans les sciences forensiques classiques et en les adaptant à la problématique posée.

Ce travail ne donne pas des valeurs concrètes nécessaires pour l'évaluation : il se pose les questions pertinentes pour planifier des recherches visant à la collection des données fournissant à l'expert une base pour l'évaluation. Il n'est pas non plus un manuel pour faire une expertise dans le domaine des images numériques et ne s'occupe pas des aspects techniques des méthodes discutées pour autant qu'ils n'ont pas d'influence sur le processus d'évaluation.

Le chapitre 2 énumère les différentes caractéristiques d'une image numérique. Le chapitre 3 explique les bases de l'évaluation Bayésienne et établit le cadre dans lequel une telle évaluation serait intégrée. Les chapitres ultérieurs décrivent la mise au point des réseaux Bayésiens permettant une évaluation de différents scénarios en commençant par

l'évaluation d'une caractéristique singulière dans le chapitre 4. Ce modèle est élargi afin de couvrir deux caractéristiques simultanément dans le chapitre 5. La possibilité des modifications des caractéristiques observées est traitée dans le chapitre 6 et la possibilité des modifications chez plusieurs caractéristiques est finalement traitée dans le chapitre 7.

2 Caractéristiques d'une image numérique

Une image peut être décrite de différentes manières. Tous ces aspects permettent par la suite de comparer une image spécifique avec les images provenant de la caméra d'intérêt. Ce chapitre traite des différentes traces donnant une information sur la provenance d'une image.

2.1 Terminologie

Le *Dictionnaire des techniques audiovisuelles et multimédias* définit une image comme *la représentation en deux dimensions d'une scène analysée par un capteur photosensible à travers un système optique plus ou moins élaboré.* [MARGUILLARD 2006] Le même ouvrage définit une caméra comme *Appareil complexe destiné à capter et à mémoriser la vue d'une scène vivante pour en assurer la reproduction.*[MARGUILLARD 2006] Cette définition englobe une énorme variété d'appareils imaginables : des appareils photographiques ainsi que des appareils vidéo-graphiques et, en faisant abstraction du système optique, même un scanner. Ces définitions peuvent facilement être restreintes sur des images et caméras numériques en spécifiant l'image en tant que *représentation numérique.*

L'image est donc le résultat d'un processus d'abstraction effectué par la caméra, qui permettra par reconstitution d'avoir un aperçu de la scène originale. Ce processus étant susceptible de laisser des traces, l'image numérique sera considérée, dans le cadre de ce travail, comme objet portant des traces laissées par la caméra qui a été utilisée pour le prendre. La caméra sera considérée comme l'outil utilisé pour produire l'image et les traces issues de cette activité seront donc considérées comme des traces d'outil.

2.2 Traces du système optique

Le système optique de chaque caméra est construit afin de permettre la prise de l'image dans des conditions optimales. Dans la réalité, des défauts, imperfections et dommages à ce système vont laisser des traces sur l'image.

2.2.1 Poussières

De manière générale, le système optique est produit afin d'éviter que de la poussière ne puisse rentrer. Par contre, beaucoup de caméras possèdent des objectifs échangeables ou pouvant être ouverts afin de remplacer des lentilles endommagées. Pendant cet intervalle, des poussières peuvent se placer sur des surfaces dans le chemin de la lumière et ainsi l'obstruer. Etant assez bien protégées de toute influence extérieure, elles peuvent y rester. Des images de ces poussières peuvent être retrouvées sur toutes les images produites par la suite et composent ainsi un bon moyen d'identification pour un appareil. [DIRIK 2013]

2.2.2 Distorsion des lentilles

Les lentilles et les miroirs constituant le système optique n'étant pas parfaits, des aberrations sont introduites dans l'image. Un type d'aberrations dont l'usage dans l'identification a été proposé par la littérature est l'aberration radiale : des lignes droites n'apparaissent plus droites sur l'image mais deviennent courbées. L'importance de la déviation est dépendante de la qualité de la lentille et de son processus de production. Le degré de cette distorsion peut être utilisé pour aider à identifier la lentille utilisée. [KAI SAN, LAM et WONG 2006]

2.3 Traces du capteur

Le cœur de toute caméra numérique est son capteur. Il transforme la lumière incidente en signal électrique qui peut par la suite être interprété comme une image. Ce processus n'étant pas parfait, des traces résultant des imperfections du capteur s'introduisent dans l'image finale. Ces traces et leur exploitation sont traités dans cette section.

2.3.1 Photo Response Non-Uniformity (PRNU)

Le processus de fabrication des capteurs d'appareils photo ne permet jusqu'à présent pas de produire des capteurs dont toutes les diodes sont de sensibilité égale. Ces défauts

de capteurs sont appelés PRNU¹. Même si les fabricants de caméras font des efforts pour réduire les résidus de cette non-uniformité, des traces se trouvent encore dans les images finales. Comme l'arrangement des diodes sur le capteur est aléatoire, cette caractéristique est d'un grand intérêt pour l'identification d'une caméra et s'est établi en tant que méthode d'identification principale.

L'approche généralement choisie consiste à isoler ces résidus à l'aide d'un filtre de suppression de bruit depuis l'image en question. La même chose est faite avec des images provenant de la caméra d'intérêt, ce qui permet la constitution d'une configuration «type», caractéristique pour cette caméra. La configuration de l'image en question est comparée à la configuration «type», donnant comme résultat la valeur par mesure de corrélation. La plupart des auteurs proposent l'usage d'un seuil de décision à partir duquel l'image est considérée comme provenant de cette caméra. [LUKAS, FRIDRICH et GOLJAN 2006, FRIDRICH 2013] A cause des raisons discutées dans la section 3.3, cette dernière étape ne sera pas suivie dans le cadre de ce travail. Le facteur de corrélation sera donc le résultat final de l'analyse des PRNU.

NORDGAARD et HÖGLUND ainsi que HOUTEN, ALBERINK et GERADTS traitent l'évaluation des PRNU par une approche Bayésienne, avec HOUTEN, ALBERINK et GERADTS choisissant des propositions légèrement différentes de celles choisies dans ce travail. [HOUTEN, ALBERINK et GERADTS 2011, NORDGAARD et HÖGLUND 2011] Leurs résultats seront adaptés dans les sections 4.3.3 et 5.2.2.

2.3.2 Pixels morts et pixels chauds

Les variations de sensibilité peuvent être si extraordinaires qu'aucune information sur la lumière incidente ne peut être introduite dans l'image. Il existe deux cas de figure : soit le pixel est complètement insensible et n'émet quasiment pas de signal (ce qui est appelé un "pixel mort"), soit il émet un signal en tout temps (ceci est appelé un "pixel chaud"). Ces pixels apparaissent donc en tant que point noir / blanc² sur l'image et seraient facilement récupérables en isolant tous les pixels ayant une valeur supérieure ou inférieure à un certain seuil³.

Le pouvoir d'identification des pixels défectueux est considéré comme élevé, vu le nombre de configurations possibles. GERADTS et al. et KUROSAWA, KUROKI et SAI-

1. de l'anglais *Photo Response Non-Uniformity*, traduit librement «réponse photométrique non-uniforme».

2. Pour une image en couleur, seul un canal de couleur sera touché.

3. Pour un image en couleur, une étape de restitution du capteur original doit être faite dû au fonctionnement du processus d'établissement des images couleurs.

TOH proposent des calculs pour l'évaluation des pixels morts, respectivement des pixels chauds, trouvés. [GERADTS et al. 2001, KUROSAWA, KUROIKI et SAITOH 2002] GERADTS et al. intègrent ce calcul dans un rapport de vraisemblance, alors que KUROSAWA, KUROIKI et SAITOH l'utilisent pour soutenir l'identification proposée. Les deux groupes d'auteurs ne couvrent qu'une partie de ce qu'ils indiquent calculer. GERADTS et al. calculent la probabilité qu'une image ait le même nombre de pixel défectueux que celle d'intérêt, ne prenant pas en compte la configuration de ces pixels. KUROSAWA, KUROIKI et SAITOH calculent l'inverse et ne considèrent que la configuration. Les deux groupes d'auteurs supposent une indépendance entre les pixels, ce qui est soutenu par GERADTS et al. sur les dix caméras qu'ils ont testées. Vu la taille du capteur, cette assumption n'est pas vraiment supportée. Afin de pouvoir faire des évaluations sur la base de cette caractéristique, il serait nécessaire de conduire une étude à grande échelle afin de vérifier l'indépendance et trouver les valeurs respectives pour les fréquences d'apparition d'un pixel mort ou d'un pixel chaud.

2.4 Balises EXIF

EXIF est un standard définissant l'inclusion des informations sur l'image dans l'image elle-même. Ces «données décrivant les données», généralement appelées méta-données, contiennent des informations sur l'image elle-même (nom, taille, dimensions..), sur les circonstances de la prise de l'image (date / heure, géo-localisation, réglage de la caméra), sur la marque et le modèle de la caméra utilisée, et même, pour certains appareils haut de gamme, un identifiant unique lié à l'appareil. Ainsi, ils donnent des informations décrivant l'appareil utilisé, dont l'intérêt paraît évident pour l'identification. [KEE, JOHNSON et FARID 2011, SANDOVAL OROZCO et al. 2015]

Malheureusement, il existe une multitude de raisons pour lesquelles une partie ou l'entièreté des méta-données identifiantes sont absentes. De base, des caméras différentes enregistrent une quantité différente de méta-données. Certaines applications, comme par exemple Facebook ou WhatsApp, suppriment la presque totalité des méta-données de toute image passant par eux. Et finalement, de plus en plus de personnes suppriment les méta-données intentionnellement pour cacher leur identité.

Plusieurs auteurs ont exprimé des réserves concernant la fiabilité des méta-données. Notamment, la simplicité de modifier ces données a été citée comme raison pour ne pas pouvoir les utiliser devant un tribunal. [LUKAS, FRIDRICH et GOLJAN 2006, KOT et CAO 2013] Il est vrai qu'une personne motivée serait capable de trouver en ligne

les moyens pour modifier ces balises en leur assignant la valeur souhaitée. Par contre, cela ne signifie pas forcément que les méta-données seraient à exclure systématiquement comme moyen de preuve. Déjà, une telle modification ne se fait pas par accident : si une certaine marque et un certain modèle sont indiqués comme étant à l'origine, cela veut dire soit que l'image a été prise par une caméra de cette marque et modèle, soit que quelqu'un a intentionnellement changé ces champs afin qu'ils affichent cette marque et ce modèle. Pour que la dernière des options soit possible, une personne devrait avoir l'opportunité et l'intérêt de faire la modification. La prochaine difficulté sera la consistance des données. Toutes les caméras n'inscrivent pas les mêmes méta-données dans une image et deux caméras différentes ne permettent pas nécessairement les mêmes réglages. Si quelqu'un modifie les méta-données d'une image provenant de la camera d'un Samsung afin qu'elles affichent la provenance d'un Iphone, il y a de fortes chances qu'au moins une des valeurs affichées dans les méta-données ne puisse se retrouver dans une image fait avec un Iphone. L'une des seules possibilités pour falsifier des méta-données sans que les traces de la modification ne soient retrouvées nécessite un accès aux méta-données d'une image qui a effectivement été créée par la camera voulue, puis de les copier dans les méta-données de l'image en question. Comme il a été mentionné auparavant, cela nécessite que quelqu'un ait accès à ces informations et ait une raison de le faire. Si c'est vraisemblable, cela découle uniquement des informations circonstanciées du cas. Il est donc du domaine du tribunal de décider si la proposition de la modification des méta-données est vraisemblable ou pas et non pas du domaine de l'expert.⁴

La question de l'évaluation des éventuelles modifications sera traitée de manière plus étendue dans le chapitre 6.

4. L'expert doit bien évidemment informer le tribunal sur les aspects discutés dans ce paragraphe si les méta-données sont utilisées comme moyen de preuve.

3 L'évaluation Bayésienne

L'évaluation proposée se base sur les réseaux Bayésiens. Ce chapitre pose les bases mathématiques et théoriques de ces derniers et élabore le cadre dans lequel une telle évaluation s'applique.

3.1 Notation

Le présent travail tentera au mieux de faire abstraction des formules mathématiques. Pour les quelques formules restantes, les notations suivantes seront utilisées :

A Événement A

\bar{A} Événement complémentaire à A ; "non A "

$Pr(A)$ Probabilité que l'événement A a lieu / a eu lieu

$Pr(A|B)$ Probabilité que l'événement A a lieu / a eu lieu, sachant que l'événement B a eu lieu.

Dans le cadre de ce travail, les probabilités représentent la mesure de l'incertitude d'un événement. Une probabilité de 1 signifie que l'événement a lieu / a eu lieu place avec certitude. Une probabilité de 0 signifie qu'avec certitude l'événement n'a pas lieu / n'a pas eu lieu.

3.2 Le théorème de Bayes

Le théorème de Bayes est à la base du calcul des probabilités conditionnelles. Dans sa forme originale, il donne la possibilité de calculer la probabilité d'un événement A , sachant que l'événement B a eu lieu.

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} \quad (3.1)$$

Ceci signifie que la probabilité de l'événement A sachant l'événement B est égale à la probabilité de l'événement B sachant A multiplié par la probabilité de l'événement A divisé par la probabilité de l'événement B .

La version la plus connue du théorème de Bayes est appelée *odds form*. Cette forme permet l'évaluation de deux événements en compétition en vue d'un événement dont on sait qu'il a eu lieu. Par exemple, il serait possible d'évaluer laquelle de deux propositions alternatives P_1 et P_2 est plus probable en vue d'une observation E , faite en lien avec ces propositions. La *odds form* du théorème de Bayes est atteinte en divisant la formule pour la probabilité de P_1 sachant E par la probabilité de P_2 sachant E et se présente comme suit :

$$\frac{Pr(P_1|E, I)}{Pr(P_2|E, I)} = \frac{Pr(E|P_1, I)}{Pr(E|P_2, I)} \cdot \frac{Pr(P_1|I)}{Pr(P_2|I)} \quad (3.2)$$

Le premier terme est appelé chances *a posteriori*⁵ : il représente les croyances après la prise de connaissance de E . Le deuxième terme est appelé rapport de vraisemblance⁶ et représente le poids de l'observation E en vue des deux propositions. Le dernier terme est appelé chances *a priori*⁷ et représente les croyances avant avoir pris en compte E . Le I représente la totalité des informations circonstancielles à propos de l'affaire ainsi que tout les connaissances de l'expert. En tant que ça, toute probabilité est conditionnée par I . Dans ce travail, elles ne seront pas explicitées mais considérées comme implicites. [LEE 2016, TARONI et al. 2014]

3.3 Utilisation du théorème de Bayes dans le cadre légal

L'usage du théorème de Bayes dans le cadre légal était initialement proposée par FINKELSTEIN et FAIRLEY. L'idée est que devant un tribunal, les propositions des deux parties sont mises en opposition. A la fin du processus, après la présentation de tous les éléments de preuve, le juge (ou les jurées) tranchera sur la véracité des deux propositions. Le processus, partant des croyances initiales du juge au sujet des propositions jusqu'aux croyances finales sur la base desquelles la décision est prise, est décrit par le

5. *posterior odds* en anglais

6. *likelihood ratio* en anglais, en raccourci LR

7. *prior odds* en anglais

théorème de Bayes. C'est un processus itératif : chaque élément de preuve est intégré en prenant les *posterior odds* de l'étape antérieure en tant que *prior odds* de l'étape suivante. [FINKELSTEIN et FAIRLEY 1970]

Ce modèle laisse aussi une place bien définie à l'expert. De manière générale, il lui est demandé de se prononcer sur la valeur de la preuve qu'il a examiné et non pas sur l'affaire en générale. Ce qui est attendu de lui est donc de s'exprimer sur le rapport de vraisemblance de l'élément de preuve qui lui a été confié⁸. Par conséquent, ce n'est pas non plus le rôle de l'expert de fixer les propositions. Ce rôle est en principe celui des parties, éventuellement en entretien avec l'expert pour rester dans le cadre du possible. Aujourd'hui, l'évaluation Bayésienne a été acceptée de manière générale comme approche logique pour gérer les incertitudes liées aux résultats d'analyses. [ROBERTSON et VIGNAUX 1993, TARONI et al. 2014] Pour tout rapport évaluatif, l'évaluation des résultats à l'aide du théorème de Bayes s'est établie et est recommandée comme unique manière de représenter les résultats par le guide du réseau Européen des instituts de sciences forensique «ENFSI» [ENFSI 2014]. L'interprétation forensique, comme est appelé la domaine qui s'occupe de la problématique de l'évaluation des traces, offre ainsi une solution à un grand spectre des problématiques, recouvrant la quasi totalité des domaines forensiques.[TARONI et al. 2014]

3.4 Les réseaux Bayésiens

Des problèmes d'évaluation deviennent vite trop complexes pour que la formule simple des probabilités conditionnelles puisse être utilisées. Ceci est dû à la quantité des aspects et dépendances entre les variables qui doivent être considéré. Bien qu'il soit possible de résoudre ces problèmes à l'aide de formules mathématiques plus développées, il devient rapidement difficile de gérer la problématique de manière structurée. Une solution pour pallier ce problème sont les réseaux Bayésiens : ce sont des graphes dirigés et acycliques avec un nombre d'états finis, dont chaque nœud représente une variable et chaque flèche représente une dépendance probabiliste. Ces dépendances sont représentées par des tables de probabilités conditionnelles associées aux nœuds respectifs. [TARONI et al. 2014] La figure 3.1 montre un réseau Bayésien basique pour deux variables A et B . Dans cet exemple, B dépend de A , c'est-à-dire que les probabilités de l'état de B sont influencées par l'état de A . Les tables 3.1 et 3.2 sont les tables des probabilités associées

8. En réalité, l'expert est souvent amené à se prononcer sur des probabilités *a posteriori*. C'est généralement le résultat d'une incompréhension du tribunal sur les implications de cette demande. Dans ce cas, il est impératif que l'expert informe le tribunal de son rôle.

respectivement aux variables A et B . Les probabilités dans une même colonne doivent toujours donner la somme de 1. Si l'état de A est connu, les probabilités des états de B découlent directement de la table des probabilités de B . Si l'état de B est connu, les probabilités des états de A peuvent être déduites à l'aide du théorème de Bayes. Cette propriété permet aussi de déduire les probabilités des états de A si B est connu.

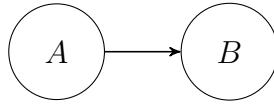


FIGURE 3.1 – Réseau Bayésien avec deux nœuds

A	$Pr(A)$
\bar{A}	$Pr(\bar{A})$

TABLE 3.1 – Table des probabilités de A

$A :$	A	\bar{A}
B	$Pr(B A)$	$Pr(B \bar{A})$
\bar{B}	$Pr(\bar{B} A)$	$Pr(\bar{B} \bar{A})$

TABLE 3.2 – Table des probabilités de B

Un autre très grand avantage des réseaux Bayésiens est qu'ils permettent de bien visualiser le problème. Par conséquent, ils permettent de mieux comprendre les problématiques en jeu et facilitent la discussion avec des personnes qui ne sont pas spécialisées dans le domaine. Un aspect important est l'influence que certaines variables ont sur le résultat final. Il se peut qu'il ne soit pas toujours simple de fixer les valeurs nécessaires à la problématique posée. Dans ces cas, il peut être pertinent de faire varier la valeur afin de voir l'influence sur le rapport de vraisemblance final. S'il est possible de montrer que même avec des valeurs considérées comme extrêmes, le résultat final reste dans le même ordre de grandeur, il sera plus simple de défendre le choix fait devant le tribunal.

Dans des environnements permettant de modéliser des réseaux Bayésiens, la structure du réseau doit être définie et les tables de probabilités spécifiées. Une fois initialisés, les états des variables connues peuvent être initiés sur leurs états respectifs. Les autres nœuds se mettent par la suite à jour par des calculs de probabilité conditionnel pour afficher les probabilités en fonction des informations introduites. Dans le réseau 3.1 par

exemple, il est possible d'initialiser le nœud B sur son état connu et de sortir par la suite les probabilités mises à jour du nœud A ⁹.

9. Pour un réseau de cette taille, le gain d'efficacité est faible par rapport au calcul manuel à l'aide du théorème de Bayes. Par contre, si la problématique devient plus compliquée, il devient vite plus facile d'utiliser des réseaux Bayésiens.

4 Observation d'une caractéristique unique

Supposons que dans l'analyse, l'expert n'a observé qu'une unique caractéristique de l'image. De plus, il est supposé que l'expert peut exclure de manière catégorique toute manipulation des images, problématique qui sera traitée dans des chapitres ultérieurs. L'expert procède ainsi à l'évaluation de la caractéristique observée.

Sur la base des discussions avec les parties, l'expert pose les propositions suivantes :

P_1 : La caméra inculpée a été utilisée pour prendre l'image d'intérêt.

P_2 : Une caméra inconnue a été utilisée pour prendre l'image d'intérêt.

4.1 Structure du réseau Bayésien

Ce cas peut être modélisé par un réseau Bayésien tout simple : deux nœuds avec une connexion. Le premier nœud, P , représente les propositions en jeu dans la présente affaire. Le deuxième nœud, E représente le résultat de l'observation faite par l'expert. La figure 4.1 montre cette structure.

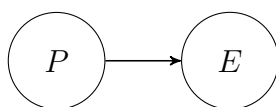


FIGURE 4.1 – Réseau Bayésien pour une seule caractéristique.

4.2 Le nœud P : les propositions

La variable P représente les propositions en jeu et a, par conséquent, deux états possibles : P_1 et P_2 pour la première proposition et la deuxième proposition respectivement. Puisque ce nœud n'a pas de dépendances à d'autres variables, les probabilités associées

ne dépendent que des informations générales (I) et son tableau de probabilité est très simple. Il est montré dans la table 4.1.

P_1	$Pr(P_1)$
P_2	$Pr(P_2)$

TABLE 4.1 – Table des probabilités de P .

Les cellules contiennent les probabilités *à priori* des propositions en jeu. Ces propositions doivent forcément être exhaustive et leur somme doit donc toujours être égale à 1. Comme indiqué dans le chapitre 3, ces probabilités ne devraient pas être fixées par l'expert. Il en découle deux possibilités : soit l'expert remplit ce tableau de probabilités conjointement avec le tribunal, soit il met des valeurs équiprobables. La première option est souvent rendue difficile par le fonctionnement du système légal et donc rarement réalisable. Au contraire, la deuxième variante, la variante des *priors* équiprobables, a plusieurs avantages. D'un point de vue conceptuel, les *à priori* neutres peuvent être interprétés comme ignorance *à priori* de l'expert sur la réponse à la question posée. Pour l'expert, avant d'effectuer ses analyses, la première version proposée peut être aussi bien vraie que la deuxième version. Bien que ceci ne soit que rarement la réalité, c'est en général ce qui est attendu par un expert considéré comme impartial. D'un point de vue pratique, cette approche permet à l'expert de directement trouver le rapport de vraisemblance pour le réseau complet sur ce nœud. Cela est dû au fait que la valeur numérique des probabilités *a posteriori* est égale au rapport de vraisemblance si la valeur des probabilités *a priori* est égale à 1. Pour ces raisons ce tableau sera rempli avec des *priors* équiprobables par la suite.

4.3 Le nœud E : l'observation faite par l'expert

En principe, il serait possible de définir les états de ce nœud de manière que tout résultat possible soit considéré individuellement. Vu la grande variabilité des appareils permettant de produire des images numériques et de la variance résultant des caractéristiques de ces images, il s'impose, à titre illustratif, de réduire ce nœud à deux états possibles : l'observation qui a été faite (E) et toute autre observation (\bar{E}).

De part la dépendance au nœud P , le tableau des probabilités est représenté selon la table 4.2.

$P :$	P_1	P_2
E	$Pr(E P_1)$	$Pr(E P_2)$
\bar{E}	$Pr(\bar{E} P_1)$	$Pr(\bar{E} P_2)$

TABLE 4.2 – Table des probabilités de E .

4.3.1 Observation d'une caractéristique discrète sans variabilité intra-source

Supposons que l'observation faite est une caractéristique discrète, comme par exemple la marque de l'appareil indiquée dans les méta-données. A titre d'exemple, prenons une caméra Canon et une image où figure dans les méta-données «Canon» comme marque de l'appareil. Rappelons que pour le moment il est considéré que l'image n'a subi aucune modification. Les valeurs du tableau des probabilités sont déterminées comme suit :

$Pr(E|P_1)$: Il s'agit ici de la probabilité d'observer «Canon» dans les méta-données sachant que l'image a effectivement été prise par la caméra en question. Comme toute image faite par cette caméra (et dont les méta-données n'ont pas été modifiées) contiendra la valeur «Canon», cette probabilité est égale à 1.

$Pr(\bar{E}|P_1)$: La probabilité d'observer une mention différente de «Canon» si l'image provient effectivement de la caméra d'intérêt est égale à 0. Il n'existe aucun cas, sachant que les méta-données n'ont pas été modifiées, dans lequel une image de notre camera «Canon» aura une valeur différente dans ses méta-données. Ce résultat intuitif remplit aussi la nécessité de compléter la probabilité précédente à 1.

$Pr(E|P_2)$: La valeur recherchée ici est la probabilité d'observer «Canon» dans les méta-données de l'image si celle-ci a été faite par une caméra inconnue, c'est-à-dire une caméra différente de la caméra d'intérêt. Afin d'obtenir cette valeur, la question de définir quelles autres caméras pourraient avoir été utilisées pour prendre cette image doit être abordée. Ceci va fortement dépendre du cas en question¹⁰. Si l'image d'intérêt est une image témoin prise spontanément, il y a de fortes chances que l'image ait été prise par la camera d'un smartphone. Si le contenu de l'image laisse sous-entendre qu'elle était prise à longue distance avec un téléobjectif, la caméra recherchée sera plutôt de type haut de gamme permettant l'usage d'un tel objectif. Si l'image a été pris autour d'une école, la

10. Ceci est l'influence de la lettre I de la formule 3.2 qui est traitée comme implicite dans ce travail.

population va être différente que si elle avait été prise dans les murs d'un domicile de retraités et ainsi de suite.

Dans une étape consécutive, il est question de quantifier l'occurrence de la caractéristique observée. Dans cet exemple, la question posée est de savoir combien d'images prises par les caméras d'intérêt ont la valeur «Canon» inscrite dans les méta-données comme valeur «marque». En survolant les situations mentionnées préalablement, il semble intuitivement clair que cette rareté va fortement varier en fonction du cas. De manière générale, cette valeur est dénommée γ et elle est égale à la probabilité recherchée¹¹.

Pour obtenir la valeur de γ , il est quasiment indispensable d'avoir accès à des bases de données actuelles contenant des cas avec des circonstances similaires. Dans le cas idéal, cette base de données permet de trouver la fréquence des images prises dans des circonstances similaires portant cette caractéristique, dans notre exemple, la marque «Canon». Si cela n'est pas possible, la fréquence des caméras dont les images montrent cette caractéristique pourrait aussi donner des résultats utiles, comme la caractéristique se traduit directement de la caméra sur l'image. Si la caractéristique est trop rare pour apparaître dans la base de données, une probabilité peut être choisie telle que la véritable probabilité soit en inférieure à la probabilité choisie¹². Pour cela, il est possible de s'inspirer des valeurs des éléments plus rares dans la base de données.

Un problème récurrent des bases de données est leur actualité. Si les caméras introduites sont issues des cas, la base de données ne sera jamais représentative de l'état actuel de la distribution dans la population au moment des faits. Cela est dû à deux phénomènes. D'un côté, l'introduction de nouveaux appareils ne peut pas être prise en compte. Au moment de la première apparition d'un appareil dans un cas, il n'apparaîtra pas dans la base de données. Cela ne veut cependant pas dire qu'il est rare. Notamment chez les smartphones, généralement équipés d'une caméra, une distribution rapide peut être supposée dû à la popularité de ces appareils. De l'autre côté, avec le vieillissement de la base de données, des appareils qui ont pratiquement disparus de la population dominent la base par rapport aux appareils actuels, puisqu'elle les a documentés pendant une plus grande période de temps. Par ces deux effets, des appareils plus anciens seront toujours sur-représentés tandis que des appareils plus récents seront sous-représentés. Pour contrer cela, il est opportun de prendre en compte l'évolution historique de l'ap-

11. Si les circonstances du cas génèrent uniquement une population de caméras très restrictives en tant que source de l'image d'intérêt, il se peut que cette approximation ne soit pas valable. Comme de tels cas sont difficilement généralisables, ils ne seront pas traités dans le cadre de ce travail.

12. Ainsi, l'évaluation sera favorable à l'accusée, un critère généralement imposé aux experts par le système juridique.

partition des caméras dans des affaires et d'adapter les probabilités par la suite.

Une autre possibilité serait d'effectuer une étude de marché. Pour cela, la probabilité γ est estimée sur la base des chiffres de vente des appareils en question. Le grand désavantage de cette approche est, qu'elle présuppose une distribution homogène des appareils dans la population. Ceci n'est par contre pas forcément le cas¹³.

$Pr(\bar{E}|P_2)$: Par complémentarité, cette probabilité est égale à $1 - Pr(E|P_2)$ et donc égale à $1 - \gamma$.

En combinant toutes ces probabilités, cela donne le tableau 4.3.

$P :$	P_1	P_2
E	1	γ
\bar{E}	0	$1 - \gamma$

TABLE 4.3 – Table des probabilités de E remplie.

4.3.2 Observations des caractéristiques avec potentiel de variabilité intra-source

Une variante doit être considérée si la caractéristique de l'image ne se traduit pas de manière directe de la caméra à l'image. Ceci est, par exemple, le cas pour l'ouverture relative de l'objectif utilisé¹⁴. Dans ce cas de figure, différentes valeurs sont possibles pour une caméra donnée, toutes les valeurs ne sont pas possibles pour chaque caméra. Supposons par exemple que la valeur «f/8.0» est observée sur l'image et que la question d'intérêt est de savoir si elle a été prise par une caméra donnée de modèle «Canon EOS 450D». Cette fois, les valeurs du tableau seront attribuées comme suit :

$Pr(E|P_1)$: Les «Canon EOS 450D» permettent 17 valeurs différentes allant de f/4.5 à f/29 pour l'ouverture relative. Une variante simple serait de dire qu'il y a donc une chance sur 17 d'observer cette valeur précise. Toutefois, comme tout photographe pourra le dire, cela ne représente pas vraiment la réalité, car cette valeur est largement dictée

13. Par exemple, des différences de prix vont avoir une forte influence sur les distributions des caméras dans différentes parties de la population.

14. A la connaissance de l'auteur, il n'existe pas d'études sur les capacités de l'ouverture relative pour identifier une caméra. Cette caractéristique est donc plutôt choisie dans une perspective d'explication que dans l'idée d'une application réelle.

par les circonstances de la prise de l'image. Dans des conditions similaires de luminosité et de distance de l'objet de l'image, la même caméra aura toujours des valeurs similaires. Une possibilité serait d'essayer de créer un modèle pour prédire la valeur de l'ouverture relative. Un tel modèle devrait bien évidemment être vérifié par des données expérimentales et ne serait pas forcément applicable à tout type de caméra. Il est donc plus simple de prendre un grand nombre d'images avec la caméra d'intérêt dans des circonstances similaires et de voir combien d'images portent la même valeur d'ouverture relative de l'objectif. On désigne cette valeur par γ_1 comme étant la probabilité sous la proposition P_1 .

$Pr(E|P_2)$: La détermination de cette probabilité se passe de manière analogue à la probabilité pour la valeur du champ «marque» élaborée dans la section 4.3.1. Par contre, dans ce cas, il est difficile de montrer depuis une base de données des caméras la rareté de l'ouverture relative spécifique, pour les mêmes raisons que discuté pour P_1 . Si seule une telle base de données est à disposition, il sera nécessaire de prendre des images dans des circonstances semblables et des configurations adaptées avec un choix de caméras représentatif de celles sorties de la base de données. La probabilité est à nouveau déterminée par la fréquence des images qui portent cette valeur. Elle est nommée γ_2 comme étant la probabilité sous la proposition P_2 .

Les probabilités des autres observations sont attribuées par complémentarité. Le tableau final est présenté dans la table 4.4.

$P :$	P_1	P_2
E	γ_1	γ_2
\bar{E}	$1 - \gamma_1$	$1 - \gamma_2$

TABLE 4.4 – Table des probabilités de E remplie.

Le cas précédent d'une caractéristique unique sans variance possible est donc un cas spécial du tableau ci-dessus, où la probabilité γ_1 est égale à 1.

4.3.3 Observation d'une caractéristique continue

Si la caractéristique observée est continue, comme par exemple la corrélation entre les configurations des PNUR des images de la caméra d'intérêt et de l'image en question, la détermination des valeurs de probabilité ne se fait pas de manière complètement égale.

La plus grande différence vient du fait que le nœud E n'est plus associé à un tableau de probabilités mais à des distributions de probabilités, dont les paramètres varient suivant les états des nœuds parents. Par convention, des nœuds à double cercles sont utilisés dans les réseaux Bayésiens pour symboliser des variables continues.

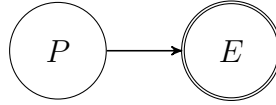


FIGURE 4.2 – Réseau Bayésien pour une observation continue.

Afin de modéliser ces distributions, il est indispensable de passer par une étape d'expérimentation qui va fournir les données nécessaires. Comme E dépend de P , il est nécessaire de trouver des valeurs permettant de modéliser la valeur de E si P_1 est vraie ainsi que si P_2 est vraie. L'examen de la configuration des PNUR par corrélation sera considéré à titre d'exemple par la suite. Cette approche a été décrite dans l'étude de HOUTEN, ALBERINK et GERADTS. [HOUTEN, ALBERINK et GERADTS 2011]

Valeurs sachant P_1 : Si P_1 est vraie, l'image en question provient de la caméra d'intérêt. Il est donc nécessaire d'obtenir des valeurs de corrélation pour des images dont on sait avec certitude qu'elles proviennent de la caméra d'intérêt. Celles-ci peuvent être facilement obtenues en faisant exactement les mêmes examens avec des images dont la provenance de la caméra d'intérêt est assurée. Pour que les valeurs obtenues soient effectivement représentatives de ce à quoi on peut s'attendre lors des tels examens, les images utilisées en tant qu'images de comparaison ne doivent pas être utilisées pour faire la comparaison, car cela risque d'augmenter artificiellement les valeurs de corrélation. Si un nombre insuffisant d'images est à disposition, ou si leur provenance ne peut être garantie, il est recommandé de prendre soi-même des images avec la caméra d'intérêt. La quantité d'images nécessaire à la modélisation doit être déterminée par une étude structurée.

Valeurs sachant P_2 : Si P_2 est vraie, l'image d'intérêt provient d'une caméra différente de celle d'intérêt. Une distribution représentative des valeurs de corrélation obtenues sous cette proposition doit donc être basée sur des valeurs obtenues par des images provenant de caméras différentes. Ces caméras doivent être représentatives de toutes les caméras qui auraient potentiellement pu être utilisées pour prendre l'image en question. Ce choix est fait de manière analogue au processus décrit auparavant pour les caractéristiques discrètes (cf. 4.3.1). Un certain nombre d'images est choisi pour chaque caméra

afin d'être comparé à la caméra d'intérêt. A nouveau, une étude déterminant le nombre d'images nécessaires doit être faite. Sur toutes les images choisies, la même procédure utilisée pour l'image de comparaison doit être appliquée pour obtenir les valeurs.

Modélisation A partir des données obtenues, l'expert peut modéliser une fonction représentative de la distribution de probabilités pour la caractéristique en question. Le type de distribution doit donc être connu soit de la littérature, soit des expérimentations structurées suivies de tests statistiques. Le type de distribution ne doit pas forcément être le même sous les deux propositions. La méthode proposée par HOUTEN, ALBERINK et GERADTS est l'usage de l'outil *dfitool* de Matlab sur les données expérimentales. [HOUTEN, ALBERINK et GERADTS 2011] Cette approche va donner pour toute distribution le type de modèle statistique qui est le plus adapté aux données. Par-contre, la distribution choisie va fortement dépendre de petites variations si un petit nombre d'échantillons est sélectionné. Il est donc recommandé d'utiliser un nombre d'images suffisamment grand pour assurer que la distribution trouvée soit le plus représentative possible de la distribution réelle.

Une fois que le type de distribution est connu, les paramètres des fonctions pour les deux propositions sont déterminés à partir des valeurs mesurées et les nœuds sont configurés de cette manière. Le tableau suivant montre la configuration du nœud E s'il est admis que les deux caractéristiques suivent une loi normale où μ est la moyenne et σ^2 la variance des mesures faites sous les propositions respectives.

$P :$	P_1	P_2
modèle statistique	normale	normale
moyenne	μ_1	μ_2
variance	σ_1^2	σ_2^2

TABLE 4.5 – Paramètres des fonctions de densité de E .

5 Observation simultanée de plusieurs caractéristiques

Dans le chapitre précédent, plusieurs cas d'observation d'une unique caractéristique ont été décrits. Il est évident qu'il n'y a aucune raison de se focaliser uniquement sur une caractéristique et d'ignorer toutes les autres observations possibles. Dans ce chapitre, un réseau Bayésien modélisant l'observation de deux caractéristiques simultanées est proposé. Dans la dernière section, la méthode pour utiliser plus que deux caractéristiques est discutée.

5.1 Structure du réseau Bayésien

Le réseau Bayésien utilisé ici est une adaptation du réseau proposé dans le chapitre précédent et peut aussi être retrouvé dans la littérature forensique [TARONI et al. 2014]. De nouveau, le nœud P représente les propositions en question. Le nœud E_1 de ce réseau correspond au nœud E du réseau précédant mais a été renommé pour des raisons de consistance. La différence principale consiste en l'introduction d'un nouveau nœud E_2 qui représente l'observation d'une deuxième caractéristique et qui dépend des deux autres nœuds dans le réseau. Comme les tableaux de probabilités associées à P et E_1 se remplissent de la même manière que discutée dans le chapitre précédent, ils ne seront

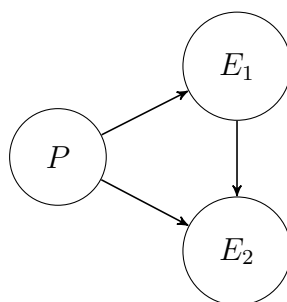


FIGURE 5.1 – Réseau Bayésien pour deux observations.

pas traités ici. La figure 5.1 montre la structure de ce réseau. A titre d'exemple, il est considéré que E_1 désigne l'observation de la valeur «Canon» dans le champ «marque» des méta-données de l'image et que la caméra d'intérêt sous P_1 est du modèle «Canon EOS 450D».

5.2 Le nœud E_2 : la deuxième observation

Le nœud E_2 caractérise l'observation faite sur la deuxième caractéristique. Comme pour la première caractéristique, elle peut être soit discrète, soit continue. Les deux options seront discutées dans cette section.

5.2.1 E_2 discrète

Si E_2 est une caractéristique discrète, par exemple l'ouverture relative de l'objective, le nœud est associé à un tableau de probabilités. Par exemple, la valeur observée est «f/8.0». Comme E_2 est dépendant de P et E_1 simultanément, la table de probabilités associées à E_2 se présente comme dans le tableau 5.1.

$P :$	P_1		P_2	
$E_1 :$	E_1	$\overline{E_1}$	E_1	$\overline{E_1}$
E_2	$Pr(E_2 E_1, P_1)$	$Pr(E_2 \overline{E_1}, P_1)$	$Pr(E_2 E_1, P_2)$	$Pr(E_2 \overline{E_1}, P_2)$
$\overline{E_2}$	$Pr(\overline{E_2} E_1, P_1)$	$Pr(\overline{E_2} \overline{E_1}, P_1)$	$Pr(\overline{E_2} E_1, P_2)$	$Pr(\overline{E_2} \overline{E_1}, P_2)$

TABLE 5.1 – Table des probabilités de E_2 .

$Pr(E_2|E_1, P_1)$: Il s'agit ici de la probabilité d'observer «f/8.0» comme valeur d'ouverture relative si la marque est «Canon» et l'image provient effectivement de la caméra d'intérêt. Comme toute image provenant de la caméra d'intérêt va avoir «Canon» comme marque, la probabilité se détermine de la manière décrite dans la section 4.3.2. On dénote cette fraction γ_{21} afin de pouvoir la distinguer des probabilités de la variable E_1 et de celles qui seront observées sous la proposition alternative.

$Pr(E_2|\overline{E_1}, P_1)$: La probabilité recherchée suppose qu'une autre marque a été observée ($\overline{E_1}$) bien que P_1 soit vraie. Ceci est logiquement impossible et les probabilités mises

dans cette colonne n'ont donc aucune importance. Afin de représenter cette impossibilité, $Pr(E_2|\overline{E_1}, P_1)$ est mise à zéro.¹⁵

$Pr(E_2|E_1, P_2)$: Il s'agit ici de la probabilité d'observer la valeur de l'ouverture relative de l'objectif, sachant que la marque «Canon» était observée mais que l'image ne provient pas de la caméra d'intérêt. Cette probabilité se détermine de manière analogue à la section 4.3.2, mais cette fois, seul les caméras «Canon» doivent être prises en compte pour former la population pertinente. La valeur trouvée est notée γ_{22} .

$Pr(E_2|\overline{E_1}, P_2)$: Ceci est la probabilité d'observer l'ouverture relative de l'objectif à «f/8.0», sachant qu'une autre marque que «Canon» a été observée et que la caméra provient d'une caméra différente de celle d'intérêt. Pour cette probabilité, le même procédé que décrit dans le paragraphe précédent sera fait, cette fois-ci en prenant toutes les caméras qui ne montrent pas «Canon» comme marque. L'occurrence trouvée est notée γ'_{22} . Si le réseau Bayésien est utilisé uniquement afin de déterminer la valeur du rapport de vraisemblance dans un cas précis, il ne vaut pas la peine de passer du temps à déterminer cette valeur. Elle n'aura aucune influence sur le résultat final, car le nœud E_1 sera fixé à la valeur E_1 et qu'il sera donc catégoriquement exclu d'observer $\overline{E_1}$.

Toutes les autres probabilités sont déterminées par complémentarité, ce qui nous donne le tableau 5.2. Il est à noter que si γ_{21} et γ_{22} sont identique, les noeuds E_1 et E_2 peuvent être considérés indépendants et la flèche entre eux n'est pas nécessaire.

$P :$	P_1		P_2	
$E_1 :$	E_1	$\overline{E_1}$	E_1	$\overline{E_1}$
E_2	γ_{21}	0	γ_{22}	γ'_{22}
$\overline{E_2}$	$1 - \gamma_{21}$	1	$1 - \gamma_{22}$	$1 - \gamma'_{22}$

TABLE 5.2 – Table des probabilités de E_2 remplie

Un cas spécial de ce tableau peut être observé si par exemple E_1 est considéré comme le champ «marque» et E_2 comme le champ «modèle» des méta-données de l'image. Dans ce cas, E_2 va toujours être observé si P_1 est vraie et il sera impossible d'observer ce modèle chez une autre marque. La valeur de γ_{22} deviendra donc l'occurrence des

15. Si $\overline{E_1}$ n'est pas exclu sous P_1 , les probabilités se déterminent en cherchant des images de la caméra d'intérêt qui montrent une autre caractéristique que E_1 et en cherchant la fraction de celles-ci qui montrent la caractéristique E_2

caméras étant de ce modèle parmi toutes les caméras de cette marque qui auraient pu être utilisées pour prendre l'image en question. La table des probabilités ainsi obtenue est représentée dans la table 5.3.

$P :$	P_1		P_2	
$E_1 :$	E_1	$\overline{E_1}$	E_1	$\overline{E_1}$
E_2	1	0	γ_{22}	0
$\overline{E_2}$	0	1	$1 - \gamma_{22}$	1

TABLE 5.3 – Table des probabilités de E_2 pour les champs «marque» (E_1) et «modèle» (E_2)

5.2.2 E_2 continue

Si la caractéristique E_2 est une caractéristique continue, comme la corrélation des PNUR, l'approche à suivre sera similaire à celle décrite dans la section 4.3.3. A nouveau, la représentation du réseau est légèrement différente puisque le nœud E_2 devient un nœud continu. La figure 5.2 montre la structure du réseau.

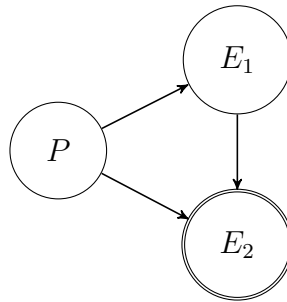


FIGURE 5.2 – Réseau Bayésien avec une observation discrète E_1 et une observation continue E_2 .

La table 5.4 montre les paramètres des fonctions de E_2 si toutes les distributions sont considérées de suivre une loi normale.

$P :$	P_1		P_2	
$E_1 :$	E_1	$\overline{E_1}$	E_1	$\overline{E_1}$
μ	μ_{21}	μ'_{21}	μ_{22}	μ'_{22}
σ^2	σ_{21}^2	$\sigma'_{21}{}^2$	σ_{22}^2	$\sigma'_{22}{}^2$

TABLE 5.4 – Table des paramètres de E_2 .

$P :$	P_1	P_2
$E_1 : E_1$	Caméra d'intérêt caractéristique E_1	Autre caméra caractéristique E_1
$\overline{E_1}$	Caméra d'intérêt pas caractéristique E_1	Autre caméra pas caractéristique E_1

TABLE 5.5 – Restrictions demandées pour les échantillons selon les états de P et de E_1 .

Comme pour une unique caractéristique continue, il est question de trouver une distribution qui représente la densité des probabilités pour la valeur observée sur la base des expérimentations. Ce qui change sont uniquement les images qui sont utilisées afin de générer les données. La table 5.5 montre les restrictions demandées pour chaque distribution.

Comme il a déjà été discuté pour les caractéristiques discrètes, il ne vaut pas la peine de faire l'expérimentation pour déterminer les distributions si E_1 est $\overline{E_1}$. Comme le réseau est utilisé afin de déterminer le rapport de vraisemblance pour E_1 et E_2 , ces distributions n'auraient aucune influence sur le résultat, car, une fois initialisée, l'option $\overline{E_1}$ est catégoriquement exclue.

5.3 Ordre de traitement des caractéristiques

En principe, l'ordre dans lequel des observations sont traitées n'a aucune influence sur le résultat final. Par contre, une approche du général au particulier s'impose par sa logique intuitive, généralement déjà suivi dans l'étape d'analyse. Suivre la même logique permet ainsi de mieux structurer le raisonnement et évite d'avoir à modifier l'ordre d'approche choisi. De plus, il est recommandé d'évaluer d'abord les caractéristiques discrètes avant les caractéristiques continues, car il est ainsi plus simple de trouver des données pour remplir les tableaux de probabilités. Imaginons par exemple, qu'on aimerait évaluer l'observation du modèle d'une image dans les méta-données ainsi que la corrélation des PNUR. Si les PNUR sont évalués avant les méta-données, il sera par la suite question de trouver la fréquence des caméras de ce modèle parmi toutes les caméras dont une telle valeur de corrélation aurait pu être trouvée. Il sera évidemment difficile de trouver une base de données permettant de retrouver cette information.

5.4 Plus que deux caractéristiques

Le concept présenté ici se généralise facilement lorsque plus de deux caractéristiques sont prises en compte. Pour toute observation supplémentaire, un nouveau nœud E_n est ajouté au réseau. Ce nœud est dépendant de P et de tout autre nœud E précédemment ajouté au réseau, sauf si une indépendance peut être démontrée entre la variable ajoutée et une des variables précédentes. Le tableau des probabilités se remplit par analogie au tableau de ce chapitre avec la différence qu'au lieu de considérer comme connu l'état de E_1 , on considère les états de toutes les autres variables E comme connus. La figure 5.3 montre un exemple pour un réseau Bayésien avec quatre observations, dont une continue. Dans cet exemple, il est supposé que des expérimentations structurées ont permis de déterminer une indépendance entre les caractéristiques E_1 et E_2 et la caractéristique E_3 . Les tables de probabilités se remplissent de manière analogue à celles décrites dans les chapitres précédents.

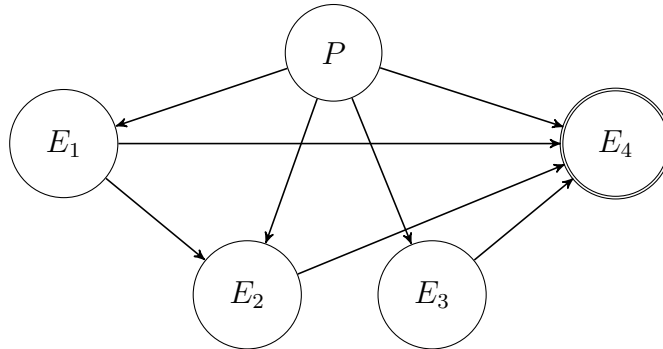


FIGURE 5.3 – Réseau Bayésien avec trois observations discrètes E_1 à E_3 et une observation continue E_4 . Dans cet exemple, la caractéristique E_3 est indépendante des caractéristiques E_1 et E_2 .

6 Prise en compte de la possibilité de modifications

Jusqu'à présent, il a été considéré que toutes les caractéristiques observées n'ont pas subis de modifications. Cette affirmation a notamment permis de considérer les méta-données des images comme des caractéristiques des caméras. Comme mentionné dans le chapitre 2, cette considération n'est pas forcément valide dans tous les cas. Effectivement, la modification des méta-données nécessite peu de connaissances afin de définir n'importe quelle valeur permettant ainsi de faussement inculper quelqu'un. Ce chapitre propose une solution pour la considération des observations potentiellement modifiées. La modification des autres traces comme par exemple de la corrélation PRNU [FRIDRICH 2013] ne sont pas traitée dans ce travail.

6.1 Structure du réseau

La structure proposée s'inspire des réseaux Bayesiens développés pour considérer d'éventuelles erreurs de laboratoire dans l'évaluation du résultat. [TARONI et al. 2014] Le concept de base consiste en la séparation de la valeur réelle d'une caractéristique (E) et de la valeur qui est observée par l'expert (OE). Appliquée à la problématique des valeurs modifiées, cette signification est légèrement altérée en «valeur définie par la caméra» (E) et «valeur observée par l'expert» (OE). Cette dernière va dépendre du nouveau nœud «modification» (M), qui représente le fait que la caractéristique a été modifiée ou non.

Les tableaux de probabilités associées à P et E sont les mêmes que dans le chapitre 4. Ils ne sont donc pas spécifiés dans ce chapitre.

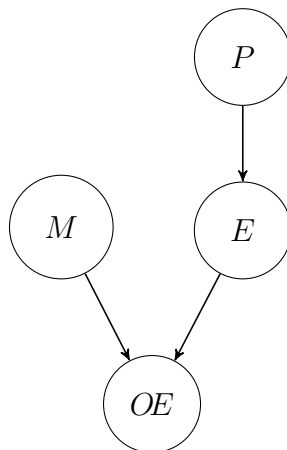


FIGURE 6.1 – Réseau Bayésien prenant en compte la possibilité d’une modification.

6.2 Le nœud M : la modification de la caractéristique

Le nœud M a deux états : soit une modification a été apportée à la caractéristique observée (M), soit il n’y a pas eu modification de la caractéristique observée (\bar{M}). On considère une modification uniquement s’il y a eu un changement de la valeur. Une modification dont le résultat final est égal au résultat initial ne constitue pas un véritable changement et n’est donc pas prise en compte. Le tableau de probabilité est similaire à celui de P . Il est montré dans le tableau 6.1.

M	$Pr(M)$
\bar{M}	$Pr(\bar{M})$

TABLE 6.1 – Table des probabilités de M

La valeur $Pr(M)$ est appelée μ . S’il est supposé d’avoir accès à une base de données contenant un grand nombre de cas dans lesquels se pose la question de la caméra source, la valeur μ peut être inspirée par ces données. En cherchant tous les cas avec des faits similaires et dans lesquels l’image ne provenait pas de la caméra source, il est possible de calculer la fraction des cas dans lesquels la caractéristique observée a été modifiée.

Cette situation idéale ne se produit quasiment jamais dans la réalité. L’existence d’une telle base de donnée est peu probable, et, même si elle existait, elle ne pourrait empêcher un nombre important de difficultés. Premièrement, pour des cas réels, il n’est pas possible de savoir avec certitude si la prise de décision finale est la bonne. Il est possible, et à partir d’une certaine taille de la base de données, fortement probable, que

la base de données contienne des erreurs qui distordent ainsi la fréquence. Deuxièmement, les événements recherchés sont ceux pour lesquels l'image recherchée ne provient pas de la caméra d'intérêt et sont donc plutôt rares dans la base de données. C'est un effet du processus investigatif qui favorise des caméras étant soumises à observation qui sont effectivement à la source de l'image. Il pourrait donc être très difficile de trouver un nombre de cas suffisant pour avoir une idée raisonnable de la fréquence des modifications.

Il est donc fortement probable que l'expert n'ait pas accès à des données lui permettant de fixer la valeur de μ . Ce que l'expert peut alors proposer au tribunal différents rapports de vraisemblance en prenant des valeurs représentant des probabilités de modification faibles, moyennes et hautes. Le tribunal pourra par la suite se mettre d'accord sur l'option la plus adaptée au cas en question¹⁶. Dans la section 6.3, un exemple pour cela peut être trouvé dans les simulations du paramètre δ .

Les autres valeurs sont trouvées par complémentarité, ce qui nous donne le tableau 6.2.

M	μ
\overline{M}	$1 - \mu$

TABLE 6.2 – Table des probabilités de M remplie

6.3 Le nœud OE : la valeur observée

La variable OE a les états OE et \overline{OE} qui correspondent respectivement à «observer la valeur E» et «observer une autre valeur». Le tableau 6.3 montre le tableau de probabilités pour OE .

$E :$	E		\overline{E}	
$M :$	M	\overline{M}	M	\overline{M}
OE	$Pr(OE E, M)$	$Pr(OE E, \overline{M})$	$Pr(OE \overline{E}, M)$	$Pr(OE \overline{E}, \overline{M})$
\overline{OE}	$Pr(\overline{OE} E, M)$	$Pr(\overline{OE} E, \overline{M})$	$Pr(\overline{OE} \overline{E}, M)$	$Pr(\overline{OE} \overline{E}, \overline{M})$

TABLE 6.3 – Table des probabilités de OE

16. Effectivement, il est possible d'argumenter que l'attribution de cette probabilité n'est pas du domaine de l'expert, comme il lui manque les connaissances nécessaires pour l'attribuer.

$Pr(OE|E, M)$: Il s'agit de la probabilité d'observer la valeur E , si la vraie valeur est E et s'il y a eu une modification de cette valeur. Comme seuls des changements de valeur sont considérés comme étant des modifications, cela est impossible. La probabilité est donc égale à zéro.

$Pr(OE|E, \overline{M})$: Cette fois, c'est la probabilité que la valeur observée soit E , si la vraie valeur est E et s'il n'y a pas eu de modification. Ceci est une certitude et la probabilité associée est donc 1.

$Pr(OE|\overline{E}, M)$: La probabilité recherchée est celle d'observer E , si la vraie valeur n'est pas E et s'il y a eu une modification. Celle-ci est égale à la probabilité que la personne ou le logiciel qui a apporté la modification à l'image aurait choisi la valeur de E comme nouvelle valeur. Cette probabilité est désignée par δ . Afin de déterminer δ de manière empirique, il est nécessaire de chercher la proportion des cas dans lesquels la caractéristique observée a été modifiée et la nouvelle valeur choisie est celle qui a été observée. Il est évident qu'il est très difficile d'obtenir une valeur représentative, notamment parce que la caméra d'intérêt qui porte cette caractéristique a sans doute une influence sur cette probabilité. Une alternative est alors de prendre la fréquence des cas dans lesquels la caractéristique observée était modifiée pour avoir la même valeur que la caméra d'intérêt. De nouveau, c'est une valeur difficile à savoir, notamment car la vérité n'est pas connue dans des cas réels (cf. section 6.2)

Comme alternative, il est donc proposé d'étudier l'influence de la valeur de δ sur le rapport de vraisemblance obtenu. La situation considérée est l'observation d'une correspondance au niveau du champ «marque» dans les méta-données. Afin de pouvoir analyser l'influence de δ , les valeurs γ et μ doivent être fixées. Il est proposé d'observer quatre valeurs de γ : 0.5 représentant une haute fréquence dans la population pertinente, 0.1 pour une fréquence moyenne, 0.01 représentant une fréquence faible et 0.001 représentant une caractéristique très rare. Pour chaque valeur de γ , quatre valeurs de μ sont observées : une modification est beaucoup plus probable qu'une non-modification ($\mu = 0.9$), une équiprobabilité entre modification et pas de modification ($\mu = 0.5$), une probabilité de modification moyennement moins probable que l'absence de modification ($\mu = 0.1$) et une probabilité de modification nettement inférieur à la probabilité qu'aucune modification ait eu lieu ($\mu = 0.01$). Ces valeurs peuvent évidemment être choisies différemment en fonction des croyances personnelles. Pour chaque configuration de ces valeurs, le rapport de vraisemblance est évalué en fonction de δ pour tout valeur entre

zéro et un.

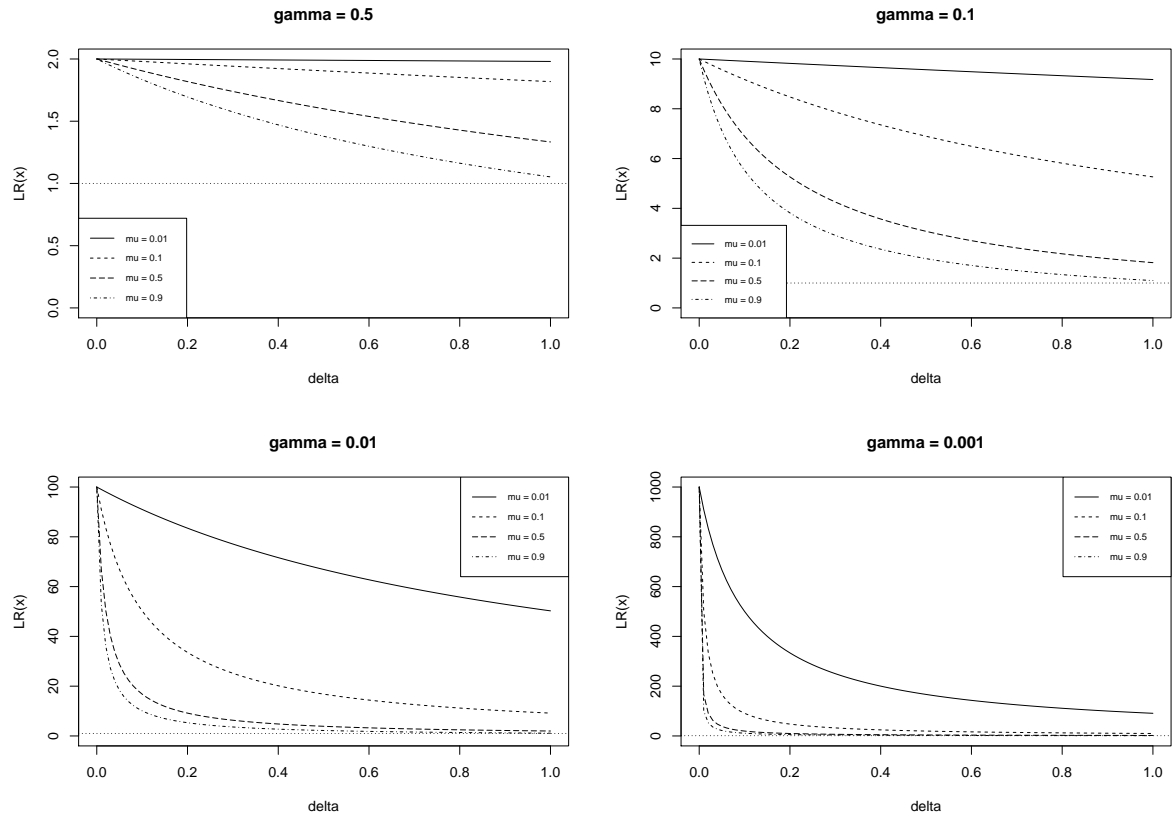


FIGURE 6.2 – Rapport de vraisemblance en fonction de δ pour différentes valeurs de γ et de μ

La figure 6.2 montre les résultats de la simulation. Il peut être observé que le rapport de vraisemblance est une fonction entre les deux extrêmes $\frac{1}{\gamma}$ si $\delta = 0$ et une valeur proche de 1, si δ s'approche de 1. L'influence de δ dépend fortement de la valeur de μ . S'il n'est que peu probable qu'une modification ait eu lieu, la probabilité que la modification se passait d'une certaine façon ne va que faiblement influencer le résultat. De même, il peut être observé que l'influence de δ est plus grande si γ est petit.

Pour un γ et un μ fixes, il est possible d'observer que le choix de δ influence le plus le résultat dans l'intervalle $[0; 0.2]$ et reste plutôt stable pour le reste. Cela signifie que si la valeur de δ est choisie entre 0.2 et 1, la valeur précise aura peu d'influence sur le résultat final. Si la proposition suivie est celle d'une modification délibérée afin d'inculper le propriétaire de la caméra, il semble cohérent de choisir une valeur dans cette gamme, comme l'entité qui a effectué la modification aura pour but de faire passer l'image comme provenant d'une autre caméra.

$Pr(OE|\bar{E}, \bar{M})$: Finalement, il s'agit de la probabilité d'observer E , alors que la vraie valeur n'est pas E et qu'il n'y a pas eu de modification. De nouveau, cela est impossible et la probabilité est donc zéro.

Toutes les autres probabilités sont trouvées par complémentarité. Le tableau 6.4 montre le tableau rempli.

$E :$	E		\bar{E}	
$M :$	M	\bar{M}	M	\bar{M}
OE	0	1	δ	0
\bar{OE}	1	0	$1 - \delta$	1

TABLE 6.4 – Table des probabilités de OE rempli

7 Réseau pour plusieurs observations avec modification

Comme discuté dans le chapitre 2, la cohérence entre différentes observations est une bonne base pour argumenter qu'il n'y a pas eu de modifications. Ce chapitre propose d'étendre le modèle du chapitre précédent afin de permettre l'évaluation de multiples caractéristiques qui auraient toutes pu être modifiées. Comme dans le chapitre 5, la solution discutée pour deux caractéristiques se généralise sans problème à un nombre de caractéristiques plus grand.

7.1 Structure du réseau

Le réseau proposé combine les réseaux créés dans les chapitres 5 et 6. Afin de simplifier le remplissage des tableaux de probabilités, une éventuelle modification est considérée comme nœud à part (nœuds M_1 et M_2) pour chaque caractéristique au lieu d'un seul nœud. Comme il est plus probable d'observer une modification dans une autre caractéristique si on sait qu'il y a eu une modification dans une autre caractéristique, M_2 est dépendant de M_1 . Comme les probabilités que la personne qui a effectué le changement choisisse des modifications cohérentes n'est pas égale à la probabilité d'une valeur non-cohérente, une dépendance entre OE_1 et OE_2 est introduite. La figure 7.1 montre cette structure.

Dans ce chapitre, seuls les nœuds qui n'ont pas été considérés dans les chapitres précédents vont être traités. Tous les autres nœuds restent identiques à ceux décrits préalablement. Une réserve doit toutefois être faite concernant le nœud E_2 : dans le chapitre 5, il a été mentionné qu'il n'est pas nécessaire de passer du temps sur la détermination des valeurs sachant $\overline{E_1}$, mais cela n'est plus le cas pour ce réseau adapté, puisque les valeurs qui seront initialisées sont les nœuds OE_1 et OE_2 . Par conséquent, il ne sera pas possible de faire des déclarations catégoriques concernant l'état du nœud E_1 . Dans ce chapitre, les variables du nœud OE_1 sont indiquées avec indice 1.

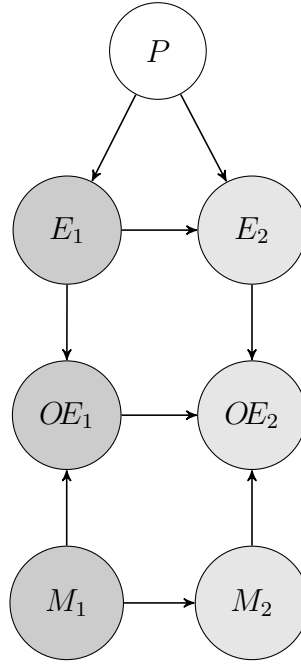


FIGURE 7.1 – Réseau Bayésien pour caractéristiques multiples avec possibilité de modification pour chacune d’entre elles. Les nœuds sont colorés afin de permettre un meilleure distinction entre les nœuds de la première caractéristique et ceux de la deuxième caractéristique.

7.2 Le nœud M_2 : la modification de la deuxième caractéristique

M_2 représente une éventuelle modification de la deuxième variable et est dépendant de M_1 . Cette dépendance représente le fait qu’une modification de la deuxième variable devient plus probable si la modification de la première caractéristique est avérée. De même, cette probabilité baisse si l’on sait que la première caractéristique n’a pas été modifiée. La table de probabilités de M_2 est donnée par le tableau 7.1.

$M_1 :$	M_1	$\overline{M_1}$
M_2	$Pr(M_2 M_1)$	$Pr(M_2 \overline{M_1})$
$\overline{M_2}$	$Pr(\overline{M_2} M_1)$	$Pr(\overline{M_2} \overline{M_1})$

TABLE 7.1 – Table des probabilités de M_2

Comme pour M_1 , il est très difficile d’attribuer ces probabilités. La probabilité $Pr(M_2|M_1)$ est dénotée μ_2 et la probabilité $Pr(M_2|\overline{M_1})$, μ'_2 . Il est estimé que μ_2 est supérieur à μ_1 , comme une modification va être plus probable sachant que quelqu’un a déjà modifié la

première caractéristique. De même, il est supposée que μ'_2 est inférieure à μ_1 comme la probabilité d'une modification va être diminuée s'il est certain que la première caractéristique n'a pas été modifiée. En ce qui concerne les valeurs concrètes de μ_2 et μ'_2 , il est considéré que leur attribution ne peut pas être uniquement du domaine de l'expert, puisqu'elles vont dépendre des circonstances du cas. Comme discuté pour μ_1 (cf. 6.2), il est recommandé d'explorer l'effet de plusieurs probabilités et de laisser le choix définitif au tribunal. Vu le nombre de variables affectées par cela, il serait judicieux de se mettre d'accord avec les parties au préalable si le système légal le permet.

Les probabilités pour \overline{M}_2 sont déterminées par complémentarité. La table des probabilités est donnée par la table 7.2.

$M_1 :$	M_1	\overline{M}_1
M_2	μ_2	μ'_2
\overline{M}_2	$1 - \mu_2$	$1 - \mu'_2$

TABLE 7.2 – Table des probabilités de M_2 remplie

7.3 Le nœud OE_2 : la valeur observée de la deuxième caractéristique

Le nœud OE_2 représente la valeur qui a été observée pour la deuxième caractéristique. Il a deux états, OE_2 et \overline{OE}_2 : soit la variable observée est la valeur de E_2 , soit une autre valeur est observée. Ce faisant, la table des probabilités est similaire à celle de OE_1 décrite dans la section 6.3.

La seule différence entre le tableau de OE_1 et celui de OE_2 est la dépendance de OE_1 par rapport à ce dernier, qui est influencée dans les cas où la vraie valeur n'est pas E_2 , mais que cette valeur a été modifiée. Il est considéré, que la probabilité qu'une personne qui aurait modifié la caractéristique E_2 l'ait fait avec plus grande probabilité de manière que les éléments restent cohérent. Par exemple, il est peu probable que quelqu'un modifie le champ «modèle» afin d'afficher la valeur «Iphone 4» si le champ marque est «Samsung». En cela, il n'est pas d'une grande importance si le champ «marque» a été modifié ou non. Par conséquent, la valeur de la probabilité que quelqu'un ait choisi la valeur E_2 en modifiant ce champ n'est pas égale en fonction de la première observation faite (\overline{OE}). Afin de pouvoir les distinguer, la probabilité de choisir la valeur E_2 si la valeur observée pour la première caractéristique est E_1 est notée δ_2 , et δ'_2 sinon.

Afin de trouver la valeur de δ_2 , les mêmes considérations que pour δ_1 s'appliquent (cf. section 6.3). L'effet de δ_2 sur le résultat final a été exploré pour les cas spécifiés dans la table 7.3. Il a été considéré que les caractéristiques observées sont la marque (E_1) et le modèle (E_2) de la caméra, inscrites dans les méta-données. La table de probabilité est donc égale à la table 5.3 discutée dans la section 5.2.1. Les valeurs des paramètres ont été choisies à titre d'exemple et ne sont pas basées sur des expériences. Les résultats sont donnés dans la figure 7.2.

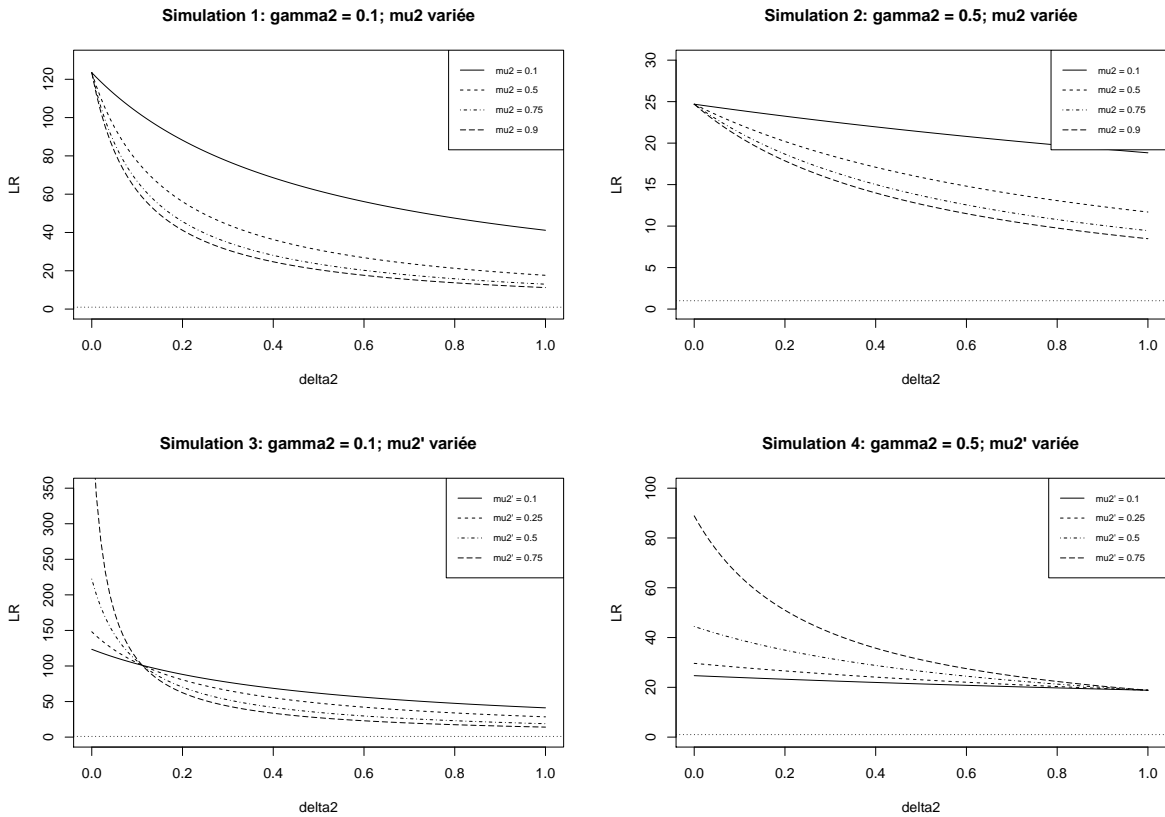


FIGURE 7.2 – Rapport de vraisemblance en fonction de δ pour différentes valeurs de rareté γ_2 , de μ_2 et de μ'_2 respectivement

A nouveau, il peut être observé que la valeur de δ_2 n'a que peu d'influence sur le résultat final, pour autant qu'elle est choisie à un seuil supérieur à 0.2. Comme déjà discuté pour le paramètre δ_1 , il paraît logique de choisir la valeur dans cette intervalle car une personne effectuant une modification a une chance élevée de choisir une telle valeur si elle tente d'inculper le propriétaire de la caméra.

En ce qui concerne δ'_2 , il ne vaut pas la peine de passer trop de temps sur la détermination de sa valeur. En effet, si le réseau est uniquement utilisé afin de déterminer

	Simulation 1	Simulation 2	Simulation 3	Simulation 4
γ_1	0.1	0.1	0.1	0.1
δ_1	0.9	0.9	0.9	0.9
μ_1	0.1	0.1	0.1	0.1
γ_2	0.1	0.5	0.1	0.5
μ_2	0.1;0.5;0.75;0.9	0.1;0.5;0.75;0.9	0.1	0.1
μ'_2	0.1	0.1	0.1; 0.25; 0.5; 0.75	0.1; 0.25; 0.5; 0.75

TABLE 7.3 – Valeurs des variables choisies pour chaque simulation du rapport de vraisemblance en fonction de δ_2 .

un rapport de vraisemblance, δ'_2 n'aura aucune influence sur le résultat final, car il n'influence que s'il est possible que l'état du nœud OE_1 soit $\overline{OE_1}$.

Le reste du tableau est rempli de manière égale à la table de E_1 , indépendamment de l'état de OE_1 . Si E_2 est la vraie valeur et s'il y a une modification sur cette caractéristique (M_2) il n'est pas possible d'observer E_2 . Ainsi $Pr(OE_2|E_2, M_2, OE_1)$ et $Pr(OE_2|E_2, M_2, \overline{OE_1})$ sont égales à zéro. Si E_2 est la vraie valeur qu'elle n'a pas été modifiée, E_2 sera dans tous les cas observé. Par conséquent, $Pr(OE_2|E_2, \overline{M_2}, OE_1)$ et $Pr(OE_2|E_2, \overline{M_2}, \overline{OE_1})$ sont égales à 1. Finalement, si la vraie valeur n'est pas E_2 et qu'il n'y a pas eu de modification de cette caractéristique, jamais E_2 ne sera observé et les probabilités $Pr(OE_2|\overline{E_2}, \overline{M_2}, OE_1)$ et $Pr(OE_2|\overline{E_2}, \overline{M_2}, \overline{OE_1})$ seront donc égales à zéro. Les probabilités pour $\overline{OE_2}$ se trouvent par complémentarité. Le tableau rempli est donné par la table 7.4.

E_2 :	E_2				$\overline{E_2}$			
M_2 :	M_2		$\overline{M_2}$		M_2		$\overline{M_2}$	
OE_1 :	OE_1	$\overline{OE_1}$	OE_1	$\overline{OE_1}$	OE_1	$\overline{OE_1}$	OE_1	$\overline{OE_1}$
OE_2	0	0	1	1	δ_2	δ'_2	0	0
$\overline{OE_2}$	1	1	0	0	$1 - \delta_2$	$1 - \delta'_2$	1	1

TABLE 7.4 – Table des probabilités de OE_2 remplie

8 Conclusion

Si la tendance actuelle persiste, les images numériques ainsi que la question de leur provenance prennent une importance croissante. Ce travail propose un premier pas dans la résolution de la problématique de l'évaluation des résultats forensiques liés à cette question. Il est évident, que les réseaux proposés ne répondent pas à toutes les questions. Toutefois, un des plus grands avantages des réseaux Bayesiens est leur souplesse : avec peu de changements, ils peuvent être adaptés pour traiter une nouvelle situation.

Modifications : Comme pour le réseau à deux caractéristiques, le réseau avec modifications s'adapte à plus que deux caractéristiques observées. De manière analogue à ce qui est mentionné dans la section 5.4, pour toute nouvelle caractéristique, trois nouveaux nœuds sont ajoutés : un nœud E_n pour l'état original de la caractéristique, un nœud M_n pour la possibilité de modification et un nœud OE_n pour la valeur observée. Il serait aussi possible de créer un réseau combinant des caractéristiques modifiables avec des caractéristiques dont une modification peut quasiment être exclue comme c'est le cas pour la distorsion radiale. Dans ce cas, un unique nœud E serait introduit, dépendant de tout autre nœud E déjà présent dans le réseau.

Dans son travail, FRIDRICH propose des techniques pour identifier des éventuelles falsifications des PRNU. [FRIDRICH 2013]. Cela pourrait être utilisé comme base pour l'évaluation d'une éventuelle altération des PRNU, en introduisant des nœuds pour des observations qui seront les indicateurs d'une altération.

Identification de la classe : Plusieurs techniques existantes visent, non pas l'identification d'un appareil précis, mais une classe d'appareils ayant un certain nombre de caractéristiques communes. En choisissant les propositions respectives, des réseaux Bayesiens axés sur la réponse à cette question sont envisageables. [KOT et CAO 2013]

Source inconnue : La situation peut se présenter où la question se pose de savoir si deux images proviennent de la même source, sans que l'on ait accès à celle-ci. Des réseaux

Bayesiens traitant ce problème existent déjà pour des traces classiques et s'adaptent aussi pour des images numériques. [TARONI et al. 2014]

Différences entre l'image et les caractéristiques : Pour l'entièreté de ce travail, il a été estimé que les caractéristiques observées sur l'image en question et sur celles provenant de la caméra d'intérêt correspondaient. Cela peut ne pas être le cas, mais n'aboutit pas directement à l'exclusion de la caméra d'intérêt. Notamment si une éventuelle modification est prise en compte, cette situation peut être d'intérêt. Afin de la résoudre, les mêmes réseaux Bayesiens pourraient être utilisés, en adaptant les états et les probabilités des nœuds à la nouvelle situation.

La mise en œuvre de la méthodologie proposée nécessitera encore du travail. Le comportement des réseaux proposées face à des valeurs extrêmes est encore à étudier. Des recherches quant aux quantités et cycles de vie des caméras afin de déterminer les fréquences de ces appareils, ainsi que des analyses approfondies sur les influences des différents paramètres des réseaux proposés sont nécessaires pour que cette méthodologie puisse être utilisable dans le cadre professionnel. Des logiciels gérant toutes les dépendances logiques pourraient permettre de gérer l'observation simultanée d'un très grand nombre de caractéristiques dans une même affaire.

9 Bibliographie

- BIEDERMANN, A., S. BOZZA et F. TARONI (2008). “Decision theoretic properties of forensic identification : Underlying logic and argumentative implications”. In : *Forensic Science International* 177.2–3, p. 120 –132. ISSN : 0379-0738.
- DIRIK, A. E. (2013). “Source Attribution Based on Physical Defects in Light Path”. In : *Digital Image Forensics*. Sous la dir. d’Husrev Taha SENCAR et Nasir MEMON. Springer, p. 219–236. ISBN : 978-1-4614-0757-7. DOI : 10.1007/978-1-4614-0757-7_7. URL : https://doi.org/10.1007/978-1-4614-0757-7_7.
- ENFSI (2014). *ENFSI Guideline for Evaluative Reporting in Forensic Science*. European Network of Forensic Science Institutes ENFSI.
- FINKELSTEIN, M. O. et W. B. FAIRLEY (1970). “A Bayesian approach to identification evidence”. In : *Harvard Law Review* 83.3, p. 489–517. ISSN : 0017811X. URL : <http://www.jstor.org/stable/1339656>.
- FRIDRICH, J. (2013). “Sensor defects in digital image forensic”. In : *Digital Image Forensics*. Sous la dir. d’Husrev Taha SENCAR et Nasir MEMON. Springer, p. 179–218. ISBN : 978-1-4614-0757-7. DOI : 10.1007/978-1-4614-0757-7_6. URL : https://doi.org/10.1007/978-1-4614-0757-7_6.
- GERADTS, Z. J., J. BIJHOLD, M. KIEFT, K. KUROSAWA, K. KUROKI et N. SAITOH (2001). “Methods for identification of images acquired with digital cameras”. In : *Proc. SPIE* 4232, p. 505–512. DOI : 10.1117/12.417569. URL : <http://dx.doi.org/10.1117/12.417569>.
- HOUTEN, W. van, I. ALBERINK et Z. GERADTS (2011). “Implementation of the likelihood ratio framework for camera identification based on sensor noise patterns”. In : *Law, Probability & Risk* 10, p. 149–159. DOI : 10.1093/lpr/mgr006.
- KAI SAN, Choi, Edmund Y. LAM et Kenneth K. Y. WONG (2006). “Source camera identification using footprints from lens aberration”. In : t. 6069, p. 6069 –6069 –8. DOI : 10.1117/12.649775. URL : <http://dx.doi.org/10.1117/12.649775>.

- KEE, E., M. K. JOHNSON et H. FARID (2011). “Digital image authentication from JPEG headers”. In : *IEEE transactions on information forensics and security* 6.3, p. 1066–1075. DOI : 10.1109/TIFS.2011.2128309.
- KOT, A. C. et H. CAO (2013). “Image and video source class identification”. In : *Digital Image Forensics*. Sous la dir. d’Husrev Taha SENCAR et Nasir MEMON. Springer, p. 157–178. ISBN : 978-1-4614-0757-7. DOI : 10.1007/978-1-4614-0757-7_5. URL : https://doi.org/10.1007/978-1-4614-0757-7_5.
- KUROSAWA, K., K. KUROKI et N. SAITOH (2002). “An approach to individual video camera identification”. In : *Journal of Forensic Science* 47.1, p. 97–102.
- LEE, P. M. (2016). *Bayesian Statistics : an introduction*. 4th. Chichester : John Wiley & Sons. ISBN : 9781118332573.
- LUKAS, J., J. FRIDRICH et M. GOLJAN (2006). “Digital camera identification from sensor pattern noise”. In : *IEEE Transactions on Information Forensics and Security* 1.2, p. 205–214. ISSN : 1556-6013. DOI : 10.1109/TIFS.2006.873602.
- MARGUILLARD, F. (2006). *Dictionnaire des techniques audiovisuelles et multimédias*. Audio-Photo-Vidéo. Dunod. ISBN : 9782100056668.
- NORDGAARD, A. et T. HÖGLUND (2011). “Assessment of Approximate Likelihood Ratios from Continuous Distributions : A Case Study of Digital Camera Identification”. In : *Journal of Forensic Science* 56.2, p. 390–402. DOI : 10.1111/j.1556-4029.2010.01665.x.
- ROBERTSON, B. et G.A. VIGNAUX (1993). “Probability—the logic of the law”. In : *Oxford Journal of Legal Studies* 13.4, p. 457–478. DOI : 10.1093/ojls/13.4.457. URL : <http://dx.doi.org/10.1093/ojls/13.4.457>.
- SANDOVAL OROZCO, A. L., D. M. ARENAS GONZÁLEZ, L. J. GARCÍA VILLALBA et J. HERNÁNDEZ-CASTRO (2015). “Analysis of errors in exif metadata on mobile devices”. In : *Multimedia Tools and Applications* 74.13, p. 4735–4763. ISSN : 1573-7721. DOI : 10.1007/s11042-013-1837-6. URL : <https://doi.org/10.1007/s11042-013-1837-6>.
- TARONI, F., A. BIEDERMANN, S. BOZZA, P. GARBOLINO et C.1 AITKEN (2014). *Bayesian Networks for Probabilistic Interference and Decision Analysis in Forensic Science*. 2nd. Chichester : Wiley. ISBN : 9780470979730.