Détection des défauts et inpainting vidéo pour la restauration de films

Arthur Renaudeau

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (Toulouse INP)*

Présentée et soutenue le *15/01/2021* **par :**

Arthur RENAUDEAU

Détection des défauts et *inpainting* vidéo pour la restauration de films

---

JURY

Jean-François AUJOL	IMB, Bordeaux	Co-encadrant
Francesca BOZZANO	Cinémathèque de Toulouse	Invité
Axel CARLIER	IRIT, Toulouse	Invité
Vincent CHARVILLAT	IRIT, Toulouse	Examinateur
Julie DELON	MAP5, Paris	Rapporteur
Jean-Denis DUROU	IRIT, Toulouse	Directeur
François LAUZE	DIKU, Copenhague	Co-encadrant
Joseph MOURE	Univ. Paris 1 Panthéon-Sorbonne	Examinateur
Fabien PIERRE	LORIA, Nancy	Co-encadrant
David TSCHUMPERLÉ	GREYC, Caen	Rapporteur

École doctorale et spécialité :
MITT : Image, Information, Hypermédia
Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (UMR 5505)
Directeur(s) de Thèse :
Jean-Denis DUROU
Rapporteurs :
Julie DELON et *David TSCHUMPERLÉ*

Titre : Détection des défauts et *inpainting* vidéo pour la restauration de films

Résumé :

Le cinéma a été inventé il y a maintenant plus d'un siècle. De nombreux films ont été tournés, suivant les évolutions technologiques de chaque époque, depuis le cinématographe des frères Lumière jusqu'à la 3D. Le support de ces films a lui aussi évolué avec le temps, passant des sels argentiques sur pellicule au disque dur pour le numérique. Avec le temps, les supports des films argentiques ayant subi des dégradations chimiques, il devient essentiel de les restaurer afin de pouvoir profiter de leur contenu dans de bonnes conditions. Il s'agit là d'un enjeu majeur de conservation du patrimoine, mais qui soulève des questions éthiques, quant à savoir jusqu'où doit aller la restauration d'un film. Dans ce contexte, et en collaboration avec la Cinémathèque de Toulouse, nous avons développé un processus de restauration automatique de films anciens qui comporte deux étapes.

La première étape consiste à détecter les défauts présents dans une image. La seconde consiste à remplir les zones ainsi détectées en procédant par *inpainting* vidéo. La prise en compte de l'information temporelle contenue dans les images adjacentes à l'image dégradée constitue un aspect essentiel de ces deux étapes. La détection des défauts de type taches, poussières ou autres, est opérée par apprentissage automatique sur réseaux de neurones profonds. En particulier, un réseau U-Net recevant trois images successives en entrée peut repérer les incohérences temporelles caractéristiques des défauts. La sortie de ce réseau est comparée à un masque de défauts créé à partir de la version restaurée de l'image centrale obtenue avec un logiciel spécialisé manipulé par un expert de la Cinémathèque de Toulouse. Enfin, le remplissage des zones endommagées est mené en alternant la reconstruction de la structure et la reconstruction de la texture de l'image à restaurer, qui toutes deux effectuent la recherche d'un optimum par des méthodes variationnelles.

Title : Defect Detection and Video Inpainting for Movie Restauration

Abstract :

Cinema was invented more than a century ago. Many films have been shot, following the technological evolutions of each era, from the Lumière brothers' cinematograph to 3D. The medium of these films has also evolved over time, from silver salts on film to hard disk for digital. With time, the supports of silver films having undergone chemical degradations, it becomes essential to restore them in order to be able to enjoy their content in good conditions. This is a major heritage conservation issue, but raises ethical questions about how far a film should be restored. In this context, and in collaboration with the Cinémathèque de Toulouse, we have developed an automatic restoration process for old movies that consists of two steps.

The first step is to detect the defects present in an image. The second step consists in filling the areas thus detected by video inpainting. Taking into account the temporal information contained in the images adjacent to the degraded image is an essential aspect of both these steps. The detection of defects such as spots, dust or other defects is performed by automatic learning on deep neural networks. In particular, a U-Net receiving three successive images as input can detect temporal inconsistencies characteristic of defects. The output of this network is compared to a defect mask created from the restored version of the central image obtained with specialized software manipulated by an expert from the Cinémathèque de Toulouse. Finally, the filling of the damaged areas is carried out by alternating the reconstruction of the structure and the reconstruction of the texture of the image to be restored, both of which carry out the search for an optimum using variational methods.

Remerciements :

Je remercie mes deux rapporteurs Julie et David pour leurs retours enrichissants malgré le peu de temps qu'ils ont eu. Je remercie Francesca et Joseph pour leur patte artistique dans ce monde de mathématiques et d'informatique. Je remercie Vincent qui, durant mon passage dans l'équipe, n'a jamais été avare d'anecdotes sur ses réunions et qui a su rester zen quand la situation pouvait s'analyser avec des données critiques, et qui tient la barre jusqu'au somment de l'INP désormais. Je remercie le contingent des bordelais, à savoir Jean-François et Fabien, avec qui l'aventure a commencé en Master pour ma reprise d'études et tout au long de ces dernières années, de la tempête à Berlin jusqu'au *road trip* à Hofgeismar. Je remercie l'expatrié danois François (qui est cylciste par définition) pour les différents séjours sur place, dont on a bien fait de profiter par rapport à la période actuelle. Je remercie aussi le jeune papa Axel qui creuse profondément en apprentissage, et qui arrive aussi à nouveau à dormir profondément avec sa petite Livia. Je vais conclure avec Jean-Denis, mon directeur de thèse, et accessoirement responsable de stage en amont de celle-ci, qui nous a fait découvrir le Lot au travers de séminaires mémorables. Ce ~~globe~~ [sud de France - Italie] trotter n'a, heureusement pour moi, pas eu trop besoin d'utiliser son stylo rouge pour corriger mes fautes.

Permanence oblige, je voudrais remercier Géraldine et Sylvie, nos Catherine et Liliane de l'équipe, qui ont toujours quelque chose à (se) raconter. Viennent ensuite les anciens doctorants de l'équipe avec qui j'ai pu partager les évènements suivants :

- La projection de film dans une grange de château réaménagée,
- Finir chez des pirates à la Victoire,
- La presque perte de lunettes près d'un éléphant rose,
- Une soirée raclette où il manque un câble d'alimentation,
- Les vocalises et jeux de mots plus ou moins sentis jusqu'à Auzeville-Tolosane,
- Des soirées *streaming* de Super Mario sur Twitch (entre autres),
- L'engloutissement d'un Mighty Mighty Burger,
- Des parties de baby-foot jusqu'au bout de la nuit,
- Une prestation de air saxophone pendant l'Eurovision,
- Des mots fleuris lors des matchs au stade (avant la relégation),
- Les ravages de la surconsommation de jus de pomme.

Et puis il y a les petits nouveaux qui n'ont pas encore eu l'occasion d'avoir une ligne de cette liste dans leur palmarès mais on espère que la situation s'améliorera pour remédier à cela.

Enfin je tiens à remercier ma famille qui m'a soutenu tout le long de ce processus de reconversion, même de loin, avec toutes les péripéties d'appartement associées. S'ajoutent à cette déjà longue liste mes amis depuis le lycée avec qui on a réussi à conserver un bon noyau dur, et ceux qui se sont rajoutés durant les années qui ont suivi (médecins, choristes, ...). Pour conclure, si finalement on ne pourra pas fêter ça dignement avant un moment en mangeant du jambon, je rappellerai juste que Bayonne est aussi un cochon tirelire dans les films *Toy Story*.

Table des matières

Chapitre 1

Introduction

1.1 Une petite histoire de cinéma

« Viva la cinema ! » a déclaré Quentin Tarantino, le poing levé, lorsqu'il a reçu la Palme d'Or du Festival de Cannes en 1994 pour son film Pulp Fiction. Avant d'en arriver là, la création du cinéma tel qu'on le connaît est passée par de nombreuses étapes, dont les origines remontent essentiellement à la fin du XIXème siècle.

1.1.1 Les débuts du cinéma

Avant même l'invention de la caméra, la première étape à l'origine du cinéma a été, du point de vue chronologique, l'invention du support souple pour les films en 1888, créé à l'origine afin de remplacer le support en verre pour les photos. En effet, l'Américain John Carbutt a inventé un support souple et transparent, en nitrate de cellulose, le celluloïd, en bandes de 70 *mm* de large commercialisées par l'industriel George Eastman. Ce sont ces bandes que l'on a retrouvées dans un premier temps dans la nouvelle génération d'appareils photo de l'époque, le Kodak, également commercialisé par Eastman via la Eastman Kodak Company.

Dès 1891, l'Américain Thomas Edison, secondé par William Kennedy Laurie Dickson, a expérimenté le format 19 *mm* à défilement horizontal et photogrammes circulaires, dont le premier résultat réussi est considéré à ce jour comme le premier film de l'histoire. Les deux hommes ont mis au point le kinétographe d'une part, un appareil pour les prises de vues animées, et le kinétoscope (Figure 1.1) d'autre part, un appareil de visionnement individuel des films obtenus avec le kinétographe. Ce dernier a alors permis au grand public de regarder les premiers films de cinéma.

Figure 1.1 – Kinétoscope d'Edison.

En 1892, le Français Émile Reynaud a conçu le premier dessin animé, qu'il a tracé et colorié aux encres à l'aniline directement sur une bande souple perforée de 70 *mm* de large, d'une grande longueur, faite de multiples carrés de gélatine, protégée par de la gomme-laque, renforcée par des cadres cartonnés et de fins ressorts, au défilement horizontal. À l'aide d'une machine de sa conception, dite « théâtre optique » (FIGURE 1.2), il s'est lancé par la suite dans les premières projections publiques payantes d'images en mouvement sur grand écran, et ce trois ans avant les projections des frères Lumière.

FIGURE 1.2 – Théâtre optique de Reynaud.

C'est également en 1892 que le français Léon Bouly a déposé le brevet d'un appareil réversible de photographie et d'optique pour l'analyse et la synthèse des mouvements, dit « Cinématographe Léon Bouly » (orthographié cynématographe lors du dépôt de brevet, puis corrigé l'année suivante par son créateur). Cet appareil n'a cependant laissé aucune trace de bon fonctionnement, et c'est d'ailleurs pour cette raison que Bouly a accepté d'en vendre l'appellation aux frères Lumière.

FIGURE 1.3 – Cinématographe des frères Lumière.

En 1894, Louis Lumière a mis au point un mécanisme ingénieux qui se différencie de ceux du kinétographe et du kinétoscope. Inspiré par le mécanisme de la machine à coudre de sa mère, où l'entraînement du tissu était assuré à l'aide d'un patin actionné par une came excentrique, Louis a dessiné une came originale qui actionnait un jeu de griffes dont les dents s'engageaient dans les perforations, déplaçaient la pellicule d'un pas tandis que, à l'instar du kinétographe, un obturateur rotatif empêchait la lumière d'atteindre la couche photosensible en déplacement. Puis les griffes se retiraient, laissant la pellicule immobile, que la réouverture de l'obturateur permettait d'impressionner d'un photogramme, et revenaient à leur point de départ pour entraîner la pellicule et impressionner un nouveau photogramme.

En 1895, les frères Lumière ont alors synthétisé les découvertes de leurs prédécesseurs pour concevoir à Lyon le cinématographe (FIGURE 1.3), un appareil capable d'enregistrer des images photographiques en mouvement sur un film Eastman de 35 *mm* de large (version de 70 *mm* coupée en deux, choix fait par Edison pour saisir plus de détails que sur sa version d'origine à 19 *mm*), et de les restituer

à la projection. Après quelques projections privées, les deux frères ont organisé à Paris les premières projections publiques payantes d'images photographiques en mouvement sur grand écran, ou du moins celles qui ont provoqué le plus grand retentissement mondial, car avant elles, d'autres projections du même type avaient eu lieu, à Berlin (Max Skladanowsky et son frère Eugen, avec leur caméra Bioskop) et à New York (Woodville Latham avec son panoptikon), mais sans le même succès. Le succès des projections dans le sous-sol du Grand Café n'était qu'un début. Dès 1896, les frères Lumière ont entrepris une gigantesque opération de tournage à travers le monde. Grâce à eux, des opérateurs ont parcouru les continents, apportant ce spectacle nouveau et étonnant qu'était un photographe actionnant consciencieusement une manivelle pour entraîner sa machine.

Mais le succès foudroyant des frères Lumière est aussi ce qui va les faire se retirer de la production de vues animées dès 1902. Comme ils n'ont pas saisi l'importance de la découverte du cinéma, dont ils sont indubitablement deux des principaux créateurs, ils géraient leur production de vues animées comme ils géraient leurs autres activités. C'étaient des patrons exigeants et durs, leur personnel le savait, et les fortes têtes ne faisaient pas carrière chez eux. Les femmes représentaient la masse fortement exploitée de leur usine, où régnaient les petits chefs chargés de mener droit les mains fragiles qui travaillaient sans protection les produits chimiques et les plaques de verre, cassantes et coupantes, qui étaient à la base de la fortune de la famille. Pour eux, seuls comptaient les classes aisées, celles qui avaient les moyens de fréquenter les beaux quartiers et le Grand Café, celles qui voulaient éduquer leurs enfants en leur montrant des images du monde.

Le 4 mai 1897, un événement tragique est survenu à Paris, frappant le public des beaux quartiers : l'incendie du Bazar de la Charité (Figure 1.4). Le feu a démarré au niveau de la cabine de projection de films, dont la lanterne fonctionnait à l'éther, faisant 129 victimes, dans leur majorité des femmes et des enfants de bonnes familles. En France et à l'étranger, l'émotion a été vive et les projections provisoirement interdites avant de reprendre. Le public des beaux quartiers commença alors à boycotter le cinéma alors que, dans le même temps, ce dernier a poursuivi sa conquête des foires, ne cessant de gagner des adeptes chez les bateleurs, et du public dans les classes populaires. Dès lors, les industriels ont fourni aux forains des films et des appareils de projection en concession, puis à l'achat, dégageant ainsi toute responsabilité en cas de sinistre.


Le Petit Journal
SUPPLÉMENT ILLUSTRÉ
Huit pages : CINQ centimes
DIMANCHE 16 MAI 1897
INCENDIE DU BAZAR DE LA CHARITÉ
LE SINISTRE

Figure 1.4 – Une du Petit Journal sur l'incendie du Bazar de la Charité.

En 1896, un prestidigitateur, Georges Méliès, s'est lancé dans la production de bobineaux qui étaient d'abord de simples répliques des vues photographiques animées des Lumière, puis il a redécouvert le trucage consistant à arrêter de faire tourner la caméra, procédé déjà employé par les Américains William Heise et Alfred Clark un an plus tôt. Ce que les cinéastes avaient utilisé une seule fois, Georges Méliès en a fait une sorte de marque personnelle, faisant apparaître, disparaître, ou se transformer des personnages ou des objets. Ces « tours de magie », plus économiques que les artifices de la scène, ont été un triomphe. Les frères Lumière eux-mêmes ont confié à des opérateurs le soin de « faire du Méliès ». Cependant, ils ont vite compris qu'ils n'étaient pas des hommes de scène, s'arrêtant par là même définitivement de produire en 1902. C'est une femme, Alice Guy, la première réalisatrice de films de l'histoire du cinéma, travaillant comme sténotypiste pour l'industriel Léon Gaumont, qui a été chargée de développer le secteur images animées de l'entreprise. En 1898, elle a inauguré un autre genre de films, en tournant une transposition fidèle des tableaux que l'on peut admirer dans les églises, introduisant de fait le concept de péplum. Pour le reste, tout le monde s'est contenté d'imiter laborieusement les films de Méliès ou ceux des cinéastes britanniques, mais à l'époque, le plagiat était la coutume au cinéma. Si les industriels avaient du mal à faire face aux saltimbanques, qui, eux, formés par le spectacle vivant, connaissaient les réactions du public et savaient les anticiper dans leurs films, un certain Charles Pathé a réussi à percer dans leur branche. Si sa première affaire, où il s'est associé avec son frère Émile, a failli le ruiner, Charles Pathé n'a pas abandonné l'exploitation du phonographe d'Edison (du moins la vente de phonographes contrefaits). Suite à une grosse commande, il a alors eu les moyens de créer la société Pathé Frères, qui est devenue au début des années 1900 la plus importante société de production de films du monde, plus puissante encore que l'Edison Manufacturing Company ou l'American Mutoscope and Biograph Company.

FIGURE 1.5 – Léon Gaumont à gauche, Alice Guy au centre, et Charles Pathé à droite.

Dans les années précédant la guerre de 1914-1918, le cinéma français s'est affranchi de sa double origine : le théâtre et le music-hall. En 1908, Victorin Jasset redécouvrait une vieille recette de la littérature : le roman-feuilleton, dans lequel un personnage récurrent entraîne le public dans les méandres

d'aventures toujours renouvelées. Le succès est au rendez-vous, démontrant bien l'attrait du public pour les fictions comprenant des personnages auxquels il pouvait s'identifier et surtout pour lesquels il pouvait trembler lors de confrontations avec des adversaires. La variété et la véracité des décors, intérieurs de studio et extérieurs naturels, ont aussi concouru à rendre plus crédibles les récits. La Première Guerre mondiale a mis la production cinématographique en pause. Le cinéma français, auparavant un des principaux producteurs de films mondiaux, s'en relèvera difficilement. Dès lors, les films américains ont envahi les écrans français. Dès le début de la guerre, la mobilisation générale a vidé les studios de tous les hommes valides. La plupart des films produits sont des films de propagande patriotique et, en 1915, la section cinéma de l'armée a été créée. Il s'agit du plus ancien service audiovisuel français.

L'arrivée du cinéma parlant à la fin des années 1920 est un tremblement de terre qui a redonné de l'allant à la fréquentation. Vingt salles sonorisées étaient recensées en France en 1929 ; elles étaient déjà 1000 en 1931 et jusqu'à 4250 en 1937. À cette époque, une belle génération de réalisateurs et une foule d'acteurs talentueux, venant le plus souvent du théâtre, ont permis la production de plusieurs chefs-d'œuvre du cinéma français. Si le public a suivi en termes de fréquentation, les chiffres restaient très loin des chiffres anglais, typiques d'une civilisation urbaine, tandis que la moitié de la population française était encore rurale. Mais l'embellie s'est arrêtée en 1939, car une grève paralyse pendant plusieurs mois le monde cinématographique français. Entre-temps, la première société a été créée par Marcel Pagnol, sous l'influence du modèle commercial de développement cinématographique américain et de ses grosses sociétés de production, après que ce dernier a quitté le théâtre en 1934. Il possédait alors ses propres studios et son propre matériel technique, qui ont été rachetés par la Gaumont quelque temps plus tard. Parallèlement, l'État marque son intervention par le biais du sous-secrétaire d'État aux Beaux-Arts, qui a créé en 1931 le Conseil supérieur de la cinématographie afin de combattre l'immoralité et la concurrence excessive d'Hollywood. Pendant la Seconde Guerre mondiale, le cinéma français n'a jamais retrouvé son niveau d'avant-guerre. En dépit de la baisse des revenus financiers et du manque de moyen, la qualité cinématographique était tout de même souvent remarquable. En 1940, le Comité d'Organisation de l'Industrie Cinématographique a été créé, pour devenir le Centre National du Cinéma après la Libération.

1.1.2 Vers l'âge d'or hollywoodien

Au début du XX^{e} siècle, tandis que l'industrie cinématographique n'en était pas encore une, la plupart de ceux qui se sont lancés dans le cinéma aux États-Unis étaient des immigrants pauvres, et plus particulièrement des immigrants juifs, venus de Russie, d'Allemagne ou d'Autriche-Hongrie. Ils ont débarqué à New York avec une volonté qui les distinguait de leurs contemporains européens, car pour eux le cinéma était plus qu'un simple divertissement, c'était la seule forme d'art qu'ils connaissaient. C'est ainsi que des sociétés ont vu le jour pour s'illustrer dans la production de chefs-d'œuvre du cinéma : William Fox a fondé la Fox Film Corporation, qui est devenue par la suite la 20th Century Fox, Samuel Goldfish a changé son nom en Samuel Goldwyn et a fondé la Goldwyn Picture Corporation, qui est devenue la Metro Goldwyn Mayer (ou MGM), et les frères Warner ont fondé les Warner Bros

Studios. Ils sont entrés en contact avec des écrivains renommés pour écrire les scénarios de leurs films ou ont acheté les droits des pièces de théâtre de Broadway, avec l'idée qu'un film adapté d'un livre ayant déjà eu du succès avait plus de chances de faire de bonnes recettes. Cette démarche n'a d'ailleurs jamais été autant d'actualité qu'aujourd'hui, puisque chacun des 19 films ayant fait le plus d'entrées en France en 2019 est une suite, un remake (film dont l'histoire a déjà été portée à l'écran) ou un *spin-off* (film se focalisant sur un ou plusieurs personnages secondaires d'un précédent film).

FIGURE 1.6 – Les premiers studios de production américains et leurs logos d'origine : 20th Century Fox à gauche, Metro Goldwyn Mayer au centre, Warner Bros Studios à droite.

Dans les années 1900, les premiers producteurs indépendants de la côte Est tournaient des petites bandes de 7 à 10 minutes dans la campagne du New Jersey autour de New York et l'hiver en Floride autour de Jacksonville, qui est devenue la première ville du cinéma à la fin des années 1900. Ils avaient recours à l'autofinancement pour leurs films. Dès les débuts du cinéma aux États-Unis, les producteurs ont eu un rôle qui allait au-delà de l'aspect purement financier car ils participaient également à la création des films au même titre que les réalisateurs et les scénaristes. Au cours des années 1910, Hollywood est devenu le principal centre de production de la nouvelle industrie cinématographique. En effet, une partie des producteurs a été attirée par la Californie, de par son climat ensoleillé, la diversité de ses décors naturels, ses terrains à bas prix, l'absence de syndicats, une main d'œuvre cosmopolite. Les lieux de tournage d'alors ressemblaient plus à des campements qu'à de véritables studios. Le milieu des années 1910 a révélé les premières grandes stars du cinéma américain, acteurs qui étaient jusque-là anonymes. Ces stars se sont alors associées en 1919 pour former la United Artists, destinée à l'exploitation de leurs films. Dès 1917, avec le développement des budgets et des stars pour fidéliser le public, les studios ont eu recours au financement extérieur via des particuliers et des banques dont le prêt était gagé sur les actifs qu'étaient devenus les stars, les scénarios et les réalisateurs.

Grâce à l'arrivée du cinéma parlant en 1927, l'industrie cinématographique américaine a passé sans encombre le début de la Grande Dépression. Cependant, dès 1931, la fréquentation des salles a baissé et la crise a provoqué un chômage massif parmi les artistes et les acteurs des années 1930, entraînant des grèves et la création de syndicats d'acteurs et de scénaristes, la Screen Actors Guild et la Writers Guild of America. À la suite du Code de production de 1930 et des opérations de restructuration et de fusion qui ont marqué les années 1929-1931, les Big Five (les cinq grandes « majors » qu'étaient la Fox,

Paramount, Metro Goldwyn Mayer, R.K.O. et Warner Bros) dominaient le nouveau marché. Pendant les années 1930, du fait de la baisse durable de la fréquentation des salles (notamment avec la hausse du prix d'entrée, les salles devant s'équiper pour le sonore) et de la hausse du coût de production d'un film, les bénéfices ont chuté, provoquant une situation financière catastrophique des compagnies de production (la R.K.O. a fait faillite et a été rachetée en 1931). Malgré cette situation, Hollywood produisait encore plus de 5000 films par an, alors que les majors imposaient le système du *block booking* (technique de marketing obligeant les exploitants de salles de cinéma à louer les films par lots). Les grands financiers prêtaient les fonds nécessaires aux studios et par là même se renforçaient dans les conseils d'administration des compagnies du cinéma hollywoodien. Pendant les années de guerre, les studios ont profité du boom économique du pays et de la standardisation de la production. Jusqu'en 1946, l'industrie cinématographie a alors connu une forte prospérité économique.

Les films des années 1930 se sont intéressés aux problèmes sociaux et au sort des plus démunis. Certains films de Charlie Chaplin dénonçaient la montée du fascisme dans *Le Dictateur* (FIGURE 1.7) et les conditions de travail des ouvriers dans *Les temps modernes*. Les comédies de Frank Capra, quant à elles, bien que légères, critiquaient les excès du capitalisme sauvage. Si l'industrie du cinéma était dominée par les hommes, c'était souvent les actrices qui imposaient leur présence indélébile et façonnaient le mythe hollywoodien à cette époque.

FIGURE 1.7 – Affiche du film *Le Dictateur* de Charlie Chaplin.

À la fin des années 1930 et pendant la guerre, le Technicolor a donné ses premières couleurs au cinéma américain, alors que le film noir, inspiré des romans de Hammett ou de Chandler, a connu un fort développement, succédant aux films de gangsters. Le Technicolor a aussi permis la réalisation du premier long-métrage d'animation marquant de l'histoire, *Blanche-Neige et les Sept Nains*, réalisé par les studios Disney, permettant à Walt Disney de devenir une figure majeure du cinéma américain avec d'autres films comme *Pinocchio* ou *Fantasia*. Lié à l'effort de guerre, Hollywood a également travaillé avec l'Office d'information de guerre, produisant des films de propagande et des films d'espionnage.

1.1.3 Le cinéma en France depuis l'après-guerre

L'accord Blum-Byrnes entre la France et les États-Unis, signé en 1946, a liquidé une partie de la dette française envers les États-Unis après la Seconde Guerre mondiale, en plus d'offrir un nouveau prêt à des conditions de remboursement considérées comme exceptionnelles. Une des contreparties de cet accord a été la fin du régime des quotas, imposé aux films américains en 1936 et resté en place à la Libération. Le compromis trouvé a été, d'une part, un abandon du quota de films américains et, d'autre part, une exclusivité accordée aux films français quatre semaines sur treize, soit une diminution

de moitié de la diffusion de films français par rapport aux années 1941-1942, entraînant un nouveau raz-de-marée de films américains dans les salles françaises. En réaction, les autorités françaises ont créé en octobre 1946 le Centre National de la Cinématographie (CNC) avec pour mission de protéger la création cinématographique française, placé sous l'autorité du Ministère de la Culture. Il a instauré par exemple une taxe sur les billets de cinéma destinée au redressement de l'industrie cinématographique du pays. Dans le même temps, le Festival de Cannes, créé en 1939 mais dont le lancement a été repoussé par la guerre, s'est affirmé très rapidement comme le plus prestigieux des festivals cinématographiques.

Dans les années 1950, les entrées en salles ont battu des records, avec une moyenne de 400 millions de spectateurs par an durant la décennie. Cet engouement populaire a profité aussi bien aux films américains qu'aux films français : c'est une période d'euphorie pour le cinéma hexagonal. Pour attirer un grand nombre de spectateurs et se démarquer du cinéma américain, les producteurs français se sont appuyés sur des stars d'avant-guerre comme Fernandel ou Jean Gabin, accompagnées de nouvelles têtes d'affiche comme Yves Montand, Jean Marais ou Brigitte Bardot. Dans l'immédiat après-guerre, le cinéma français a rendu hommage aux résistants. Le cinéma français était avant tout un cinéma de studio et de scénaristes, friand d'adaptations littéraires et de films en costumes. Cette qualité française est certes caractérisée par des films très bien scénarisés, mais dont la réalisation est souvent académique (peu de mouvements de caméra et de jeux de lumière, afin de respecter au mieux les exigences dramaturgiques du scénario). La fin des années 1950 a été marquée par la Nouvelle Vague, arrivée massive de nouveaux réalisateurs, cinéphiles venus de la critique (Cahiers du cinéma) et des rangs de la Cinémathèque française, opposés au système traditionnel de production français et à son classicisme devenu académique, amoureux du grand cinéma de genre hollywoodien incarné par Alfred Hitchcock, John Ford, ou encore Jean Renoir. La Nouvelle Vague a également vu l'émergence de nouveaux langages cinématographiques représentatifs d'une modernité qui va balayer tout le cinéma mondial à cette époque.

À partir de la fin des années 1970, le cinéma français est entré dans une période difficile. Le nombre de spectateurs, qui avait déjà fortement chuté, passant de 424 millions en 1947 à 184 millions en 1970, a continué de décliner à cause du développement de la télévision. Les foyers s'équipaient de plus en plus alors que la couleur arrivait et que le nombre de chaînes augmentait. L'arrivée de Canal+ en 1984, qui a fait du cinéma l'un des éléments essentiels de sa grille, était le symbole de cette époque. Face à cette situation, plusieurs mesures ont été prises, comme la mise en place du Compte de soutien financier à l'industrie des programmes audiovisuels par le CNC en 1986. D'autre part, en 1988, un décret interdit aux chaînes en clair de diffuser des œuvres cinématographiques les mercredis et vendredis soir (sauf exception), le samedi toute la journée et le dimanche avant 20h30, afin de protéger les salles de cinéma et d'attirer potentiellement un maximum de spectateurs. De concurrent, la télévision va devenir l'un des principaux financeurs du cinéma, les chaînes étant mises à contribution, en particulier Canal+. Le CNC a également aidé au financement de nombreux films avec l'avance sur recettes. Enfin, les petits cinémas de quartier laissent place aux multiplexes, gérés par les grands groupes comme Pathé Gaumont (Gaumont ayant été racheté par Pathé en 2001), UGC ou CGR.

Après avoir touché le fond en 1992 avec 116 millions d'entrées, le renouveau du cinéma français a débuté en 1993 avec la réussite des *Visiteurs*, son plus grand succès depuis 25 ans. Les années 2000 confirment la bonne santé du cinéma français, avec en point d'orgue l'énorme succès de *Bienvenue chez les Ch'tis*, qui a fait mieux que *La Grande Vadrouille* (1966) et frôlé le record de *Titanic* (1997) avec 20 millions d'entrées. Les années 2010 ont commencé sur le même rythme avec plus de 19 millions d'entrées pour *Intouchables* en 2011, qui est devenu par là même le troisième film le plus vu de l'histoire du cinéma français, détrônant également *La Grande Vadrouille*. Le cinéma français de la fin des années 2000 et des années 2010 est également marqué par le triomphe international de films français tel que *La Môme* (2008) retraçant la vie d'Édith Piaf, récompensé par deux Oscars et un Golden Globe, *The Artist* (2012), film hommage aux années 1920 et à la transition du muet vers le parlant, qui s'est vu récompensé de cinq Oscars, de trois Golden Globes et du prix d'interprétation masculine du Festival de Cannes, ainsi que le thriller *Elle* (2016), récompensé de deux Golden Globes et d'une nomination aux Oscars.

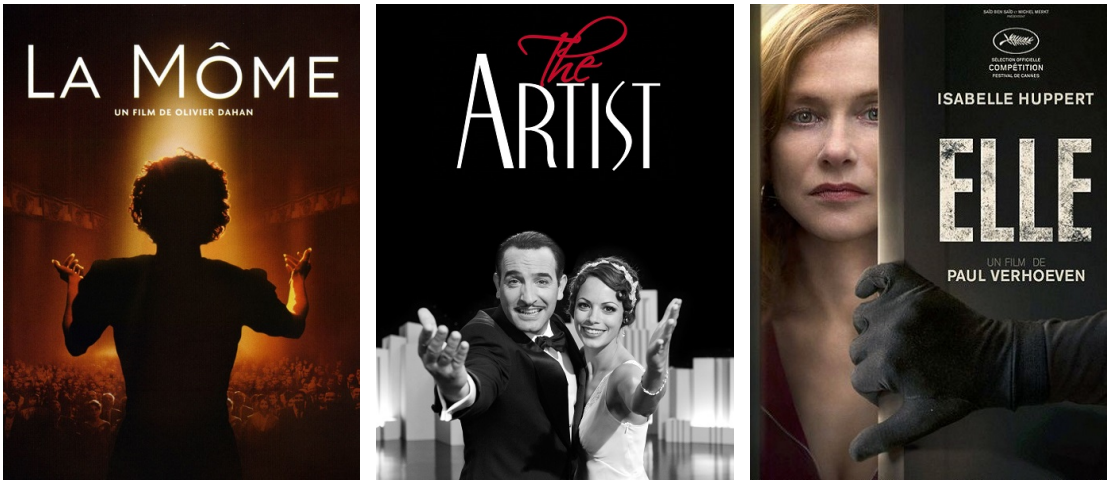


FIGURE 1.8 – Le renouveau du cinéma français à l'étranger avec de nombreuses récompenses ces dernières années.

1.1.4 Grands studios et grosses franchises américaines

Le cinéma américain des années 1950 était caractérisé par le triomphe des grandes productions hollywoodiennes. L'arrivée du CinemaScope a favorisé le succès des productions coûteuses, des films aux couleurs somptueuses, aux mises en scène spectaculaires, projetés sur de grands écrans panoramiques. Toutefois, sur le plan financier, cette stratégie s'est avérée peu rentable en raison du coût élevé des investissements, de la maintenance et du personnel, d'où une diminution du nombre des films produits, et ce malgré le maintien des petites compagnies indépendantes. L'exportation et la diffusion du cinéma américain à l'étranger, soutenu par le département d'État et la Motion Picture Export Association, a apporté une solution à ces difficultés et permis d'exporter le rêve américain, s'opposant à la propagation de l'idéologie communiste. La montée en puissance de l'Actors Studio entraîna le renouvellement

du jeu des acteurs qui composaient des personnages tourmentés à la forte charge physique et émotionnelle. Dans le même temps, le développement de la télévision et la migration des cadres vers la banlieue a entraîné le déclin des salles de centre-ville. Un phénomène typique de cette période a été le développement rapide du *drive-in* ou cinéma en plein air.

Durant les années 1970 et les années 1980, l'augmentation du nombre des films de science-fiction et de super-héros a contrebalancé la diminution des westerns. Nombre de ces films à gros budget ont eu des suites, des produits dérivés et ont fait l'objet d'un marketing intensif. S'ils étaient en partie modernes, par leurs sujets (rencontre avec l'autre, apparition extraterrestre, intelligence artificielle) et leurs effets spéciaux, le récit n'en restait pas moins linéaire, avec un traitement de l'histoire assez classique. Les producteurs visaient davantage une catégorie de la société que le grand public. Le film *Titanic* (1997) de James Cameron résume bien quant à lui la tendance des années 1990 : goût de la performance, énormes moyens financiers et techniques, numérisation des images, record d'audience, et bande originale à succès. Si les valeurs sûres étaient toujours très présentes, une nouvelle génération de réalisateurs commençait à s'affirmer avec force, jouant avec les codes et genres cinématographiques avec une grande virtuosité (Soderbergh, Tarantino ou Burton).

Les progrès informatiques dans les années 2000 ont permis l'explosion de nouvelles améliorations techniques, marquant une augmentation croissante des films d'animation entièrement réalisés en images de synthèse, en remplacement du traditionnel dessin animé. Sont également apparus à cette période des productions utilisant la capture de mouvement, ou *motion capture*, technique permettant d'enregistrer les positions et déplacements d'objets ou de membres d'êtres vivants, pour en contrôler une contrepartie virtuelle sur ordinateur. La technologie du cinéma 3D, jusqu'alors limité à des salles spécialisées, événements exceptionnels ou parcs à thèmes tels que le Futuroscope, ont également commencé à accompagner les productions destinées aux salles grand public habituelles. Les années 2000 sont également marquées par l'apparition de grosses productions de genre fantastique ou de science-fiction intégrant ces nouvelles technologies informatiques, comme la trilogie du *Seigneur des anneaux*, la prélogie de la saga *Star Wars*, l'octalogie *Harry Potter* (pour seulement sept livres...), ou *Avatar* (qui a largement contribué à la généralisation des salles en 3D). Les films de super-héros ont connu dès lors un fort regain de succès, comme les *X-Men*, *Spider-Man*, *Batman*, et les *Avengers* (avec le plus gros succès commercial au box-office de nos jours). Si les années 2010 ont suivi le rythme en termes de grosses franchises, les studios sont allés encore plus loin, au point d'exploiter les univers étendus (présents dans les livres) en plus d'avoir produit des films issus de la littérature, comme la trilogie du *Hobbit* ou la saga des *Animaux fantastiques*. Cette décennie a aussi été marquée par le rachat par Disney de Lucasfilm, entraînant la production d'une trilogie supplémentaire dans l'univers *Star Wars*, et encore d'autres projets en cours. Ce rachat n'était que le premier car il a été suivi par celui de Marvel Studios et de la 20th Century Fox par la firme aux grandes oreilles, ce qui fait de cette dernière le studio de cinéma le plus important dorénavant.

1.1.5 La concurrence de la télévision et de la VADA

L'arrivée de la télévision dans les foyers a commencé à se démocratiser dans les années 1950. La première baisse de fréquentation des cinémas vient avec le développement de la télévision dans les années 1960, conduisant les personnes à rester chez elles et à moins fréquenter les salles obscures. Cette chute de la fréquentation est compensée par la hausse du prix des billets, qui accroît encore l'abandon des salles par les classes populaires. Ainsi, entre 1960 et 1990, le prix d'une place a été multiplié par 17 alors que les prix ont été multipliés seulement par 7, en moyenne, sur cette même période. La fréquentation connaît une phase de stabilité dans les années 1970 et au début des années 1980, car l'influence de l'équipement des ménages en téléviseurs s'est amoindrie, ceux qui s'équipaient étant soit déjà des téléspectateurs ayant adapté leur consommation, soit des ménages moins consommateurs de films. La jeune génération est devenue une génération habituée aux films, notamment grâce à la télévision qui popularise le cinéma par un effet démultiplicateur de l'audience des œuvres cinématographiques. Néanmoins, la baisse se poursuivait avec la création et la multiplication des chaînes de télévision privées qui offrent une grande quantité de films en produit d'appel.

À partir de 1992, le développement des multiplexes, notamment dans les zones rurales, permet un nouveau mode de consommation du cinéma et un retournement de tendance. En créant des établissements modernes, plus attractifs et plus accessibles (la plupart étant situés dans des zones commerciales de périphérie), les exploitants de salles de cinéma ont inversé la tendance et la fréquentation est repartie à la hausse. Dès lors, les multiplexes ont permis de toucher une population plus large, finissant même par accueillir la majorité des spectateurs en France, aux dépens des petits cinémas indépendants. À la fin des années 1990, le lancement de formules d'abonnement illimité permet de stimuler la fréquentation des spectateurs assidus, avec un effet positif sur la fréquentation, par l'accroissement du taux d'occupation des fauteuils des cinémas concernés. Alors que les années 2000 voient l'émergence du piratage des films pour une diffusion illégale sur Internet, le développement progressif d'une offre à la demande pour la télévision, l'ordinateur, la tablette, et le *smartphone* font en sorte que le piratage a peu d'influence sur les entrées en France. En particulier, les offres de *streaming* et de téléchargement, légales ou non, sont plutôt à mettre en concurrence avec les achats physiques de films (DVD, Blu-ray). D'autre part, la Haute Autorité pour la Diffusion des Œuvres et la Protection des Droits sur Internet (HADOPI), qui doit lutter contre le piratage et la diffusion illégale des films, atteint clairement ses limites car son champ d'action est trop limité, et elle se révèle trop coûteuse (pour 1€ d'amende, 942€ sont dépensés pour assurer son fonctionnement). Finalement, le box-office français franchit même la barre des 200 millions d'entrées en 2009 et s'installe durablement au-dessus de ce seuil par la suite.

Pour un film sorti en salles, la chronologie des médias qui s'applique en France dicte les différentes durées d'attente aux autres diffuseurs pour pouvoir le proposer pour la télévision. En particulier, les plateformes de Vidéo À la Demande par Abonnement (VADA), plus connues sous l'anglicisme *SVOD* pour *Subscription Video On Demand*, sont les dernières servies pour pouvoir diffuser les films après leur sortie au cinéma. De fait, même si la part de films présents sur les plateformes représente plus

de la moitié du contenu proposé, elle n'est composée en très grande partie que de films sortis il y a au moins trois ans. Si le public s'abonne à ces plateformes, c'est avant tout pour les nouveautés en termes de séries télévisées, qui représentent les trois quarts du contenu visionné. Cependant, ce circuit de films qui passent par la case cinéma s'est vu quelque peu remis en question avec la sortie de certains films directement sur les plateformes, comme argument marketing supplémentaire d'exclusivité et de nouveauté. Loin de vouloir simplement ajouter des films à son catalogue, Netflix peut se targuer d'avoir gagné de nombreuses récompenses pour le film *Roma* d'Alfonso Cuarón (Oscar du meilleur réalisateur et meilleur film étranger en 2019) et de nombreuses nominations pour le film *The Irishmann* de Martin Scorsese en 2020. Des réalisateurs connus ont donc déjà passé le cap consistant à aller travailler ailleurs que pour les studios historiques.

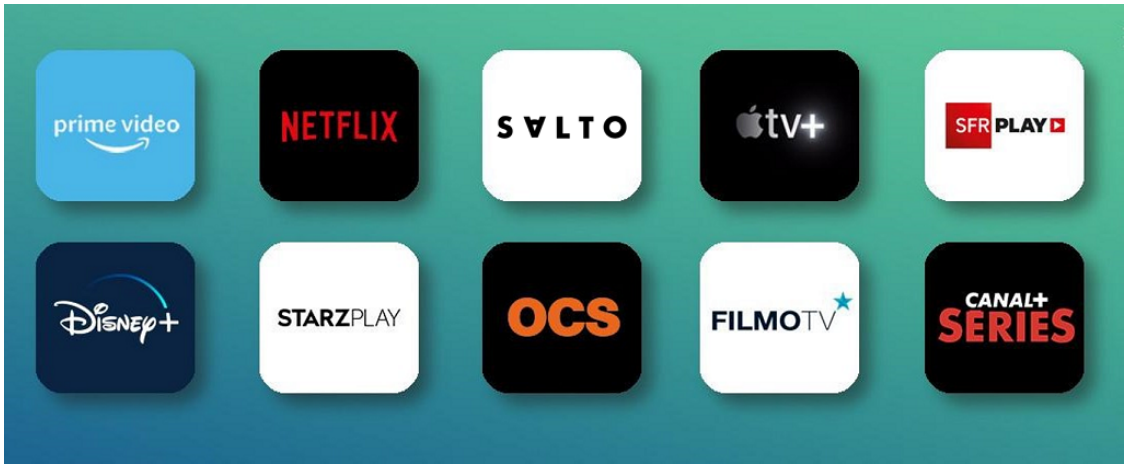


FIGURE 1.9 – Les différents acteurs de la vidéo par abonnement à la demande en France.

D'autre part, à cause de la pandémie de Covid19, certains studios ont décidé de sortir des films dont l'exploitation en salles était prévue de longue date, comme l'adaptation en prises de vues réelles de *Mulan* qui sortira directement sur la plateforme Disney+, au grand dam des exploitants de salles. Si pour le moment la grande majorité des sorties de films ont été reportées à 2021, on ne sait pas encore si les studios vont suivre l'exemple de Disney si la situation perdure, ou prendre exemple sur Warner avec le film *Tenet* de Christopher Nolan, finalement sorti en salles à la rentrée, mais dont le nombre d'entrées est pour le moment faible aux États-Unis, le film se rattrapant financièrement avec les entrées à l'étranger. La conséquence immédiate sur le nombre d'entrées en salles pour l'année 2020 en France est de -64%, le mois de janvier étant déjà sur la pente descendante par rapport à 2019. La fin d'année 2020 ne devrait pas pouvoir endiguer la chute avec les nombreux reports. Pour reprendre l'exemple de Disney, par le biais de ses filiales cette fois, 2020 sera la première année depuis 2009 sans aucun film Marvel, alors que le prochain film *Soul* du studio Pixar sortira également sur Disney+ pour les fêtes de Noël 2020. Le second confinement, décrété fin octobre 2020, sonne le glas du cinéma, qui aura connu une année 2020 catastrophique, malgré les aides de l'État pour compenser le manque à gagner.

1.2 La restauration de films

Après plus d'un siècle de cinéma, les films sur bobines sont légion, et se pose alors la question de la conservation de ces œuvres. La préservation et la restauration sont nécessaires pour la survie des grands classiques cinématographiques et des bandes sonores afin que leur état demeure le plus près possible de l'original. Inévitablement, avec le temps, les interventions humaines et l'usure, les films et rubans magnétiques se détériorent jusqu'à devenir irréparables et illisibles. Heureusement, des techniques de préservation et de restauration peuvent réduire les dommages que subissent ces supports physiques. La restauration des films, travail d'orfèvre qui permet aux classiques de retrouver leur éclat et de refaire leur apparition dans les salles de cinéma, est un marché en pleine expansion, comme le montre par exemple la projection de longs métrages rénovés au Festival Lumière.

1.2.1 Les différentes altérations du temps liées au support (pellicules)

Jusque dans les années 1950, les pellicules 35 *mm* utilisées pour tourner des longs métrages étaient composées de nitrate de cellulose. Les professionnels se sont rapidement rendu compte des désavantages de cette substance, qui subit une décomposition chimique continuelle rendant le nitrate extrêmement inflammable. Pour cette raison, seulement 10% des pellicules originales des films américains tournés avant 1929 ont pu être conservés et environ la moitié de ce qui a été filmé avant 1950 a survécu. Les pertes sont incalculables et inéluctables même si de plus en plus d'efforts sont fournis pour sauver ces chefs-d'œuvre. Malgré tous les efforts prodigués par les archivistes, il n'y a aucune façon d'endiguer de manière permanente la détérioration des supports d'information. Il en va de même pour les films et les rubans magnétiques contenant la trame sonore et les dialogues. En revanche, il existe des moyens permettant de minimiser les dommages causés par le temps, la poussière, les désastres naturels, les mauvaises manipulations, l'usure, les problèmes techniques d'équipement, les insectes, la rouille et les variations de température et d'humidité. La préservation fait toute la différence pour prolonger la durée de vie d'un film. Trois types de pellicules sont utilisés pour le tournage de longs métrages à partir des années 1950. Il s'agit du nitrate de cellulose, de l'acétate de cellulose et du polyester. Selon le type de pellicule choisi, le film peut développer, au cours de son existence, une dépolymérisation pour les supports en nitrate, ou le syndrome du vinaigre pour les supports en acétate.

Le nitrate de cellulose, utilisé jusque dans les années 1950, en se désagrégeant, cause la détérioration du film par un processus chimique lent causé par la nature même de la cellulose et par des mauvaises conditions d'entreposage (voir Figure 1.10). De plus, le nitrate de cellulose est une substance extrêmement inflammable et impossible à éteindre car, en se consumant, il produit son propre oxygène. Les gaz émis par sa dégradation provoquent également l'apparition de la rouille qui attaque le boîtier contenant le film si ce dernier est métallique. Pour remédier à cette propension à s'auto-détruire, il est nécessaire de conserver les pellicules en nitrate de cellulose dans un environnement à température et humidité contrôlées. Il est conseillé de rapidement dupliquer ces films sur polyester afin de ne pas perdre l'information consignée sur ces supports.

Le film composé d'acétate de cellulose, utilisé jusque dans les années 1980, subit le syndrome du vinaigre, processus irréversible de rétrécissement, de friabilité et de déformation de l'émulsion de gélatine. Son nom provient de l'odeur vinaigrée dégagée lors de la dégradation du support. La meilleure façon d'enrayer le problème, en plus d'un diagnostic rapide, est de tester les bobines avec les « A-D Strips », qui sont des languettes de papier équivalentes au papier pH. Elles servent à déterminer le degré de détérioration de ces films par le syndrome du vinaigre ou simplement à savoir si ces derniers sont attaqués. Dans le cas positif, il est alors possible d'effectuer un traitement approprié. Afin de sauvegarder les pellicules, celles-ci peuvent être dupliquées, mais elles doivent toutes être conservées dans un milieu de stockage frais et sec, soit à environ 5°C et 50% d'humidité relative. Contrôler l'humidité relative et la température dans les centres de conservation de films est très important car cela permet d'éviter des détériorations de l'émulsion, la prolifération des moisissures et des insectes, ou que le film devienne collant. En général, afin de préserver le plus longtemps possible les pellicules de films, celles-ci doivent être conservées à une température oscillant entre 18°C et 21°C et entre 40% et 50% d'humidité. Depuis les années 1980, c'est le polyester qui est utilisé puisque, constitué d'une matière qui est un résidu de pétrole, ce corps inerte n'est pas soumis aux dégradations chimiques comme ses prédécesseurs.

FIGURE 1.10 – Les différentes étapes de dégradation du nitrate de cellulose.

1.2.2 Les différents types de défauts dans les images

Dans un premier temps, il est utile d'apporter des précisions sur la « fabrication » des images argentiques de films. Ici le support de la pellicule est recouvert d'une émulsion sur laquelle sont couchés en suspension des cristaux d'halogénure d'argent (ou de bromure d'argent pour les émulsions plus récentes). Après exposition de la pellicule à la lumière, des photons sont adsorbés pour créer des électrons qui vont être captés par les ions d'argent, formant des atomes d'argent. Pour chaque cristal, selon l'intensité de la lumière reçue, entre zéro et une dizaine d'atomes se forment, ceux-ci ayant tendance à former un agrégat. Seuls les cristaux contenant un nombre minimal d'atomes d'argent agrégés pourront être entièrement réduits, lors du développement photographique, en particules noires visibles par l'œil humain. Dans le cas des pellicules en couleur, il y a plusieurs couches d'émulsion

avec des sensibilités différentes dues à des leuco-colorants ou coupleurs colorés. Le cas particulier du Technicolor doit aussi être rappelé. La captation de la couleur s'effectuait initialement par trois bobines différentes dans les caméras avec un prisme séparant la lumière en cyan, magenta et jaune. Ensuite, à travers un processus d'imbibition de colorant, une bobine de film de gélatine était pressée pour chaque couleur. La conséquence de la séparation des couleurs sur trois bobines a pu entraîner des problèmes de recalage au niveau des contours des objets sur les images. Parmi les différents défauts contenus dans les images, que l'on peut apercevoir à la projection d'un film, en excluant les dégradations globales des pellicules mentionnées précédemment, on peut distinguer les défauts locaux des défauts non locaux, d'un point de vue spatial (intra-image) et également d'un point de vue temporel (inter-images).

Le défaut le plus controversé est celui du grain argentique, créé justement par l'agglomération des atomes d'argent évoquée précédemment. L'assimilation de ce grain à un défaut fait débat, dans le sens où ce grain peut être assimilé à un bruit présent dans l'image et qui va « bouger » aléatoirement lors de la projection du film. Il faut réfléchir quelque peu avant de se lancer dans la restauration en enlevant purement et simplement le grain par filtrage pour obtenir un rendu plus lisse et donc potentiellement plus agréable à l'œil. En effet, on peut également considérer que le grain fait partie à part entière du style du film original et qu'il ne devrait pas être supprimé. La gestion du grain doit se faire au cas par cas en fonction de son importance relative dans l'image pour savoir s'il est nécessaire d'en atténuer ou non les effets. Le bruit numérique dans les vidéos vient quant à lui de conditions de tournage défavorables (principalement de nuit avec un manque de lumière) et il est beaucoup plus souhaitable de le supprimer au besoin. La tendance actuelle est même finalement à vouloir recréer numériquement le grain argentique, via des algorithmes appropriés, pour l'inclure dans des photos et vidéos numériques, comme dans [Newson *et al.*, 2017].

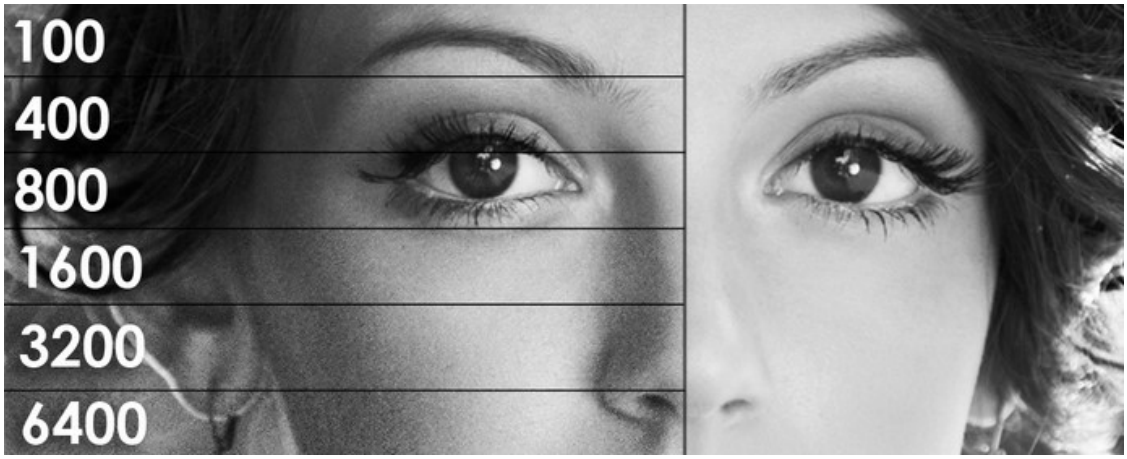


FIGURE 1.11 – Exemple de différents niveaux de grain sur une photographie.

Le problème de la stabilisation, qui survient dans une séquence d'images, est lié à un léger mouvement de la caméra qui fait que la scène filmée n'est pas stable. Ce phénomène peut survenir lors de l'enregistrement d'une scène opéré de façon manuelle, ce qui entraîne une secousse au niveau de la caméra, ou plus simplement si le caméraman se déplace avec la caméra pour filmer, ce qui entraîne également des secousses liées au déplacement. Pour pallier ce problème, il existe des stabilisateurs à effet gyroscopique pouvant être attachés à la caméra qui compensent les petits mouvements brusques du caméraman. Il est notable que les derniers *smartphones* et caméras disposent de stabilisateurs intégrés au logiciel de capture vidéo. Le problème de la stabilisation a notamment été traité dans [Kokaram *et al.*, 2003].

La décoloration des films des années 1950 est causée par des changements chimiques modifiant les couleurs, en particulier le cyan et le jaune. Ce problème dépend essentiellement de la qualité de la pellicule et des colorants utilisés à l'origine, ainsi que du traitement du support prodigué ultérieurement. Les films plus récents sont moins touchés par ce problème, car les nouveaux colorants sont plus stables. Afin de prévenir cette décoloration, il est important de conserver les films dans un environnement frais, voire même froid, et relativement sec. Nos contributions sur l'*inpainting* sont en particulier inspirées des travaux sur la colorisation de films de [Pierre *et al.*, 2017].

L'effet de « pompage », appelé aussi effet de scintillement, modifie la clarté d'une image au milieu d'une séquence, ce qui perturbe la continuité entre images successives. Il convient de différencier cet effet entre films argentiques et films numériques. Dans le cas argentique, certaines images peuvent paraître plus claires ou plus foncées, du fait d'un temps d'exposition de la pellicule variable, à cause par exemple d'une vitesse de défilement irrégulière de la bobine lors de l'enregistrement. Dans le cas du numérique, l'effet de scintillement se produit quand la vitesse d'obturation de la caméra n'est pas synchrone avec la fréquence du courant électrique des lampes ou avec la fréquence de balayage de certains écrans. En particulier, il faut privilégier la fréquence de 25 images par seconde pour un courant à $50Hz$ en Europe, alors que la norme est de 30 images par seconde pour le courant à $60Hz$ en Amérique et en Asie. Par rapport aux défauts précédents, cet effet peut donc être plutôt localisé dans l'image (tout en étant persistant sur plusieurs images). L'atténuation du pompage a notamment été traitée dans [Delon et Desolneux, 2010].

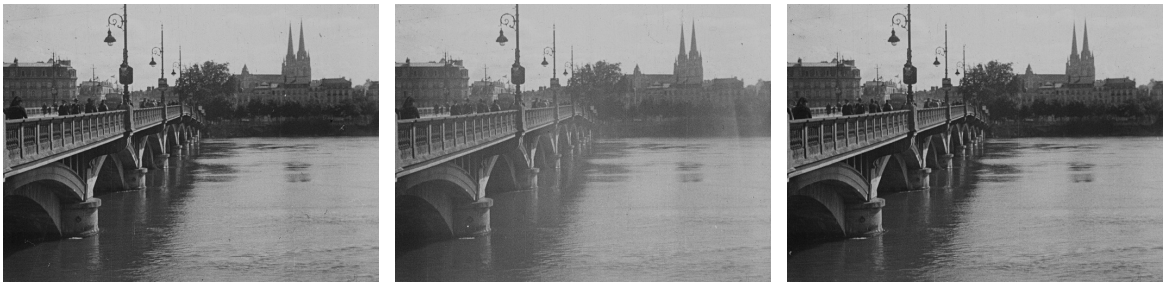


FIGURE 1.12 – Trois images consécutives où l'image centrale a été surexposée, ce qui provoque un effet de « pompage » à la projection du film.

Les rayures verticales sont des artefacts facilement visibles sous la forme de lignes d'intensité claire ou sombre, sur une grande partie de l'image. Elles peuvent être causées par le frottement vertical d'une particule sur le matériau du film dans le projecteur ou par l'abrasion du film lorsqu'il frotte sur une particule prise dans le mécanisme. La rayure est alors présente sur plusieurs images consécutives jusqu'à ce que la particule soit évacuée. Si ces défauts sont fins horizontalement, ils n'en restent pas moins « profonds » verticalement et temporellement.

FIGURE 1.13 – Trois images consécutives contenant une rayure sur le côté gauche, proche du cadre.

Les « collures » sont les soudures effectuées à l'aide d'une colleuse et permettant de joindre, dans l'ordre choisi, deux plans de montage d'une pellicule cinématographique ou d'une bande magnétique. Le terme de collure désigne par là même l'endroit de la pellicule où est faite cette opération. La presse à souder les films est constituée de deux parties extérieures qui servent à aligner les deux plans que l'on veut assembler après les avoir coupés aux ciseaux. Deux volets tiennent la pellicule bien à plat. Au centre se trouvent des ergots sur lesquels sont accrochées les perforations de chaque plan. Un grattage de la gélatine est exécuté, le dissolvant est déposé à l'aide d'un petit pinceau, le presseur de la partie centrale est ensuite rabattu et bloqué aussitôt, le temps de la soudure. Au niveau des images collées (anciennement en bord de pellicule), des traces de produit ou de manipulation peuvent subsister après l'opération. Les presses à souder ont maintenant pratiquement disparu quand ont été lancées les presses à scotch.

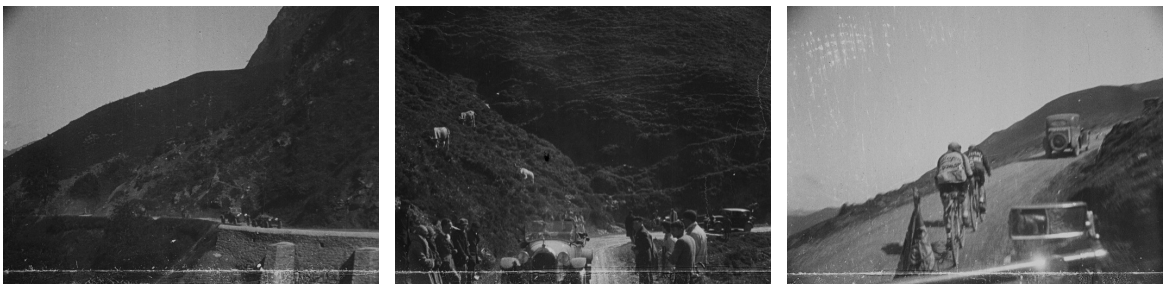


FIGURE 1.14 – Rayures horizontales au bas de la dernière image de différentes scènes, qui illustrent les dégradations liées au montage.

Les salissures, poussières, éraflures, déchirures et taches sont dues à l'environnement ou à de mauvaises manipulations du film. Présentes sur une seule image (pas de corrélation temporelle comme pour les rayures), l'emplacement, la forme et la localisation sont très variables.

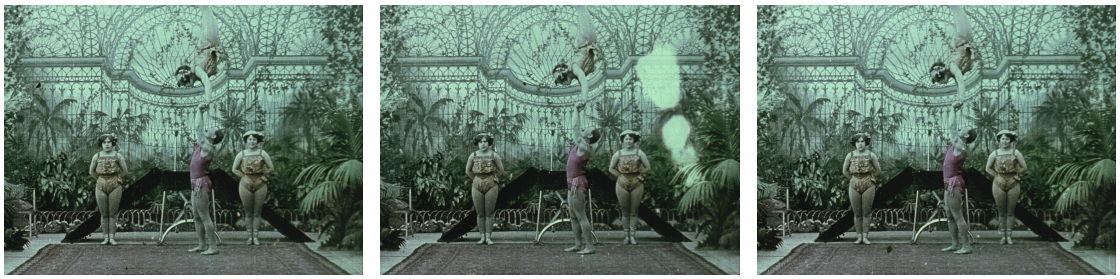


FIGURE 1.15 – Trois images successives comportant un défaut de type tache présent uniquement dans l'image centrale.

1.2.3 Les enjeux de la restauration

Pour un coût allant de 80000 à 200000 euros, la restauration d'un film peut être financée jusqu'à 90% par le Centre National du Cinéma et de l'image animée (CNC), à l'aide du « Plan de numérisation pour la restauration des films ». Mais avant que ces films ne reviennent à l'affiche, des questions de déontologie agitent aussi les laboratoires chargés de leur rénovation.

Chaque étape de la restauration n'a en théorie qu'un seul objectif, celui de retrouver l'esthétique première de l'œuvre. Ceci s'avère parfois utopique pour les films très anciens ou dont le réalisateur n'est plus de ce monde. Si la bonne éthique veut que les restaurateurs travaillent à effacer les usures du temps, et non à corriger les défauts d'origine, la pratique diffère parfois : un réalisateur encore vivant peut attacher de l'importance à « nettoyer » des imperfections, tandis qu'un autre peut considérer qu'elles font partie de l'histoire du film. À l'instar d'un Jacques Tati, qui a remixé toute son œuvre avant sa mort, ou d'un George Lucas avec la trilogie *Star Wars*, certains metteurs en scène profitent d'une restauration pour corriger à l'écran un oubli, un regret ou ajouter du contenu grâce aux avancées technologiques. Jean-Marie Poiré, lui, a voulu gommer son équipe technique, oubliée à l'arrière-plan d'une scène des *Visiteurs*. Parfois, les restaurateurs se heurtent à de petites subtilités, comme dans *Week-end* de Jean-Luc Godard, où un décadrage de l'image était en fait un choix artistique.

Pour en revenir à l'exemple de *Star Wars* en 1993, en vue de l'anniversaire des 20 ans de la saga, Lucas a souhaité ressortir la trilogie sur grand écran. Cependant, les pellicules ont été retrouvées dans un état de dégradation avancé en dépit de leur entreposage dans une voûte souterraine à température contrôlée. En effet, malgré toutes ces précautions, les couleurs étaient devenues fades à cause des changements chimiques des colorants et de la mauvaise qualité des pellicules. Afin de redonner une seconde jeunesse aux films, un processus long de trois ans a été nécessaire à la reconstitution des trucages optiques originaux et à la rénovation des couleurs. Certains effets spéciaux ont également

été refaits dans le but d'obtenir une meilleure qualité technique et visuelle, en donnant un aspect plus naturel à certaines scènes. L'ensemble des opérations de restauration de la trilogie a coûté la somme de 10 millions de dollars, soit le même montant que le budget du premier film. Toutefois, les interventions de restauration n'ont pas simplement consisté en un rajeunissement cinématographique visant à redonner une apparence neuve à la trilogie, elles ont été un moyen pour George Lucas de parachever son œuvre.

1.2.4 Conservation et restauration des films par le CNC

Le Centre National du Cinéma et de l'image animée (CNC), appelé Centre National de la Cinématographie jusqu'en 2009, est un établissement public à caractère administratif français placé sous l'autorité du Ministère de la Culture. Il a été créé en 1946, suite à la concertation entre les pouvoirs publics et les professionnels du cinéma afin de réglementer ce dernier en France. Le CNC a également pour missions de soutenir économiquement et de promouvoir le cinéma auprès de tous les publics, ainsi que de veiller à la protection et à la diffusion du patrimoine cinématographique. Le cinéma de patrimoine désigne ici, selon la définition du CNC, un film dont la première date de sortie en salles est antérieure à dix ans.

À ce titre, la Direction du Patrimoine Cinématographique (DPC) est chargée de la gestion directe et indirecte de l'activité patrimoniale cinématographique française. Elle gère la conservation, la sauvegarde, la restauration et le catalogage des films sur tous supports, confiés au CNC dans le cadre de dons, de dépôts volontaires et du dépôt légal des œuvres cinématographiques, en vigueur à partir de 1977 et pris en charge par le CNC en 1992. Elle met en œuvre la politique du patrimoine cinématographique et a également pour mission la coordination des grandes institutions patrimoniales françaises consacrées au cinéma : la Cinémathèque française, la Cinémathèque de Toulouse, la Cinémathèque de Grenoble, l'Institut Jean Vigo de Perpignan, l'Institut Lumière de Lyon... C'est à Bois d'Arcy dans les Yvelines que le Service conservation et logistique des collections du DPC conserve quelque 110000 films qui couvrent plus de cent ans de cinéma. Afin de préserver ce qui constitue la plus grande collection de films en Europe et de permettre à ces œuvres du passé de déjouer les aléas du temps, les équipes du CNC déploient tout leur savoir-faire pour donner une nouvelle jeunesse aux films de patrimoine.

Dans un premier temps, le chargé de restauration cherche à rassembler tout le matériel d'origine : négatifs, copies, mais aussi l'ensemble des informations qui pourront aider les équipes à reconstituer le film (dossier de production, journal de tournage, script annoté, contexte historique, équipements utilisés...) afin de retenir les éléments qui permettront la meilleure restauration. En effet, en France, la restauration d'un film se fait en général à partir de l'original et non d'une copie. Le choix des films à restaurer se fait selon des critères physiques (état de la pellicule, longueur ou durée, état des couleurs) et techniques (nature de l'élément). Ce travail de reconstitution nécessite souvent de pister un élément à travers différentes collections dans le monde, chez des particuliers par exemple, de rechercher des bobines à travers des réseaux spécifiques comme les laboratoires, les services d'archives, ou de faire

appel aux autres cinémathèques en France comme à l'étranger. Cette collaboration, facilitée par la Fédération Internationale des Archives du Film (FIAF), permet de retrouver les éléments parfois manquants et de sélectionner les meilleurs éléments pour effectuer la restauration.

Si l'état de la pellicule d'origine est jugé suffisamment bon, les équipes procèdent à une restauration photochimique. Les bobines sont étudiées dans le moindre détail afin de déterminer l'état physique de chaque élément. C'est un travail méticuleux et laborieux, qui nécessite de réparer des perforations, des déchirures ou des collures qui se sont défaites avec le temps. Vient ensuite l'étape de « l'essuyeuse » : les plus petits défauts du film, comme les poussières ou les moisissures, sont nettoyés pour préparer la pellicule à l'étape suivante du processus argentique : le tirage. Il arrive que certains films, trop abîmés pour être traités, nécessitent une restauration manuelle. La pellicule, qui, avec le temps, a pu se décomposer, est devenue cassante ou poisseuse. Il convient de l'assécher ou de l'assouplir et de l'humidifier au moyen de différents produits. De même, les perforations de la pellicule peuvent être endommagées, en raison d'un passage trop fréquent en tireuse d'époque. Les techniciens remplacent alors les perforations manquantes nécessaires au défilement de la pellicule dans le projecteur. Le support, désormais réparé et prêt à être manipulé dans de bonnes conditions, passe au tirage. Cette opération concerne essentiellement les vieux films sur support en nitrate de cellulose, un type de pellicule utilisé jusqu'au début des années 50, dont la particularité est d'être très facilement inflammable.

Si l'on corrige numériquement tout ce qui est venu détériorer l'image (poinçons, rayures, décoloration...), certains « artefacts » sont néanmoins conservés. En effet, le processus de restauration cinématographique implique de préserver une version la plus proche possible de l'œuvre originale et impose donc de respecter les techniques de l'époque, comme l'exigent le plan de sauvegarde et de restauration des films anciens mis en place par le Ministère de la Culture en 1990 ou le plan de restauration et de numérisation des films du patrimoine. Il s'ensuit une duplication de la pellicule sur un nouveau support pour en avoir une seconde sauvegarde, par mesure de sécurité. L'objectif est de toujours revenir à la pellicule argentique après la restauration car il s'agit d'un support bien plus fiable et pérenne que le numérique. En effet, les pellicules fabriquées aujourd'hui peuvent être conservées au moins trois cents ans. En revanche, si un élément est trop fragile pour passer au tirage, on procède alors à une restauration numérique.

Lorsqu'un film nécessite une restauration numérique, la bobine originale est alors scannée en 2K, en 4K voire en 8K en format non compressé. Le film apparaît non pas au format vidéo, mais sous la forme d'une suite de fichiers images. Divers logiciels sont utilisés pour « nettoyer » numériquement l'image (stabilisation, étalonnage, réparation des déformations, harmonisation des plans pour assurer la cohérence visuelle du film). Ce travail de restauration de longue haleine est supervisé par un chargé de restauration qui intervient, muni d'une palette graphique, pour venir à bout des défauts les plus importants, comme la reconstitution de parties d'images manquantes ou du son, la suppression des rayures... L'œuvre restaurée passe ensuite à l'imageur, une machine qui reporte les fichiers numériques sur la pellicule négative. Ce négatif sert de sauvegarde au film restauré, à partir duquel des copies

argentiques peuvent être tirées. D'ailleurs, si les copies argentiques sont numérisées dans un but d'accessibilité ou de restauration, les films actuels tournés en numérique doivent quand même être copiés sur un support argentique. La copie doit être déposée au dépôt légal dans le même objectif de conservation et de sauvegarde.

Autre organisme important de conservation du patrimoine en France créé par la réforme de l'audiovisuel menée en 1974, l'Institut National de l'Audiovisuel (INA) est notamment chargé d'archiver les productions audiovisuelles, de produire, d'éditer, de céder des contenus audiovisuels et multimédia à destination de tous les publics, professionnels ou particuliers. En 1995 est créée l'Inathèque, chargée de la conservation et de la mise à disposition des archives aux chercheurs et aux étudiants. La numérisation à grande échelle des archives de l'INA a été décidée en 1999, sous la forme d'un plan de sauvegarde et de numérisation massif et systématique. À la suite de ce travail de numérisation de ses archives, l'INA a lancé en 2006 le site ina.fr, ouvert au grand public, qui accueillait à son lancement 100000 archives représentant 10000 heures de programmes.

1.2.5 Partenariat avec la Cinémathèque de Toulouse

Dans le cadre de l'appel à projet « Services Numériques Innovants » lancé par le Ministère de la Culture en 2019, et qui a pour objectif de permettre la réalisation de la preuve de concept d'une solution numérique innovante à destination d'un acteur culturel, un partenariat a été mis en place avec la Cinémathèque de Toulouse, concrétisé par le projet Restauration Automatique de Films par Intelligence Artificielle (RAFIA). Ce projet a pour but la détection et la restauration automatique de défauts dans les films, via des algorithmes basés sur l'apprentissage profond et les méthodes variationnelles. La Cinémathèque de Toulouse dispose d'une collection très importante de vieux films argentiques, dont certains ont été numérisés, et dont les défauts ont déjà été corrigés manuellement par un expert. Ces données numérisées de la Cinémathèque de Toulouse nous ont servi de base d'apprentissage. Une prise de parole lors du festival Histoires de cinéma, organisé par la Cinémathèque de Toulouse en 2021 (initialement prévu en novembre 2020 mais reporté suite aux mesures sanitaires liées à la Covid19), permettra de montrer nos résultats et d'échanger avec des professionnels du secteur.

FIGURE 1.16 – Affiche de la 4ème édition du festival Histoires de cinéma (reporté à 2021 suite aux mesures sanitaires mises en place fin 2020).

1.3 Contributions à la détection et la correction des défauts

Cette thèse s'articule autour de deux axes pour procéder à la restauration de films : la détection des défauts dans les images, puis le remplissage des zones endommagées ayant été détectées. Dans les deux cas, la prise en compte de l'information temporelle contenue dans les images adjacentes à l'image traitée est un aspect essentiel.

La détection des défauts de type taches, poussières ou autres, qui sont présents sur une seule image, est opérée en utilisant l'apprentissage automatique par réseaux de neurones profonds. En particulier, un réseau U-Net prend trois images en entrée pour repérer les incohérences temporelles caractéristiques des défauts. La sortie du réseau est comparée avec la version restaurée du film par un expert de la Cinémathèque de Toulouse, restauration effectuée à l'aide d'un logiciel spécialisé.

— **Learning Defects in Old Movies from Manually Assisted Restoration**, Arthur Renaudeau, Travis Seng, Axel Carlier, Fabien Pierre, François Lauze, Jean-François Aujol, Jean-Denis Durou, *International Conference on Pattern Recognition (ICPR), 2020.*

— **L'intelligence artificielle au service de la restauration cinématographique : une collaboration en cours entre la Cinémathèque de Toulouse et l'IRIT**, Arthur Renaudeau, Travis Seng, Axel Carlier, Fabien Pierre, François Lauze, Jean-François Aujol, Jean-Denis Durou, *Festival Histoires de cinéma, 2020.*

Le remplissage des zones endommagées est effectué en alternant une approche par pixels pour la diffusion de la structure avec une approche par *patchs* pour récupérer la texture. Ces deux approches utilisent les images voisines temporellement dans leur recherche d'un optimum, via des méthodes variationnelles.

— **Alternate Structural-Textural Video Inpainting for Spot Defects Correction in Movies**, Arthur Renaudeau, François Lauze, Fabien Pierre, Jean-François Aujol, Jean-Denis Durou, *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), 2019.*

— **Inpainting vidéo pour la restauration de films par reconstructions alternées de la structure et de la texture**, Arthur Renaudeau, François Lauze, Fabien Pierre, Jean-François Aujol, Jean-Denis Durou, *Journées ORASIS, 2019.*

Le chapitre 2 présente un état de l'art autour de la détection et de la restauration des défauts. Le chapitre 3 s'intéresse à notre méthode de détection des défauts dans les films, qui utilise l'apprentissage profond. Le chapitre 4 présente le modèle de restauration par méthodes variationnelles pour la correction desdits défauts. Le chapitre 5 fait office de synthèse en validant sur un film de la Cinémathèque de Toulouse notre pipeline complet de détection et de restauration.

Chapitre 2

État de l'art

Cet état de l'art se concentre sur la détection des défauts, ainsi que sur les méthodes d'*inpainting* dans les vidéos. Ainsi, le déroulement des sections de ce chapitre suit une logique applicative plutôt qu'une simple logique chronologique. En plus de séparer la partie détection du chapitre 3 de la partie *inpainting* du chapitre 4, cette décomposition met également en lumière le changement d'approche des méthodes de résolution de l'*inpainting* en fonction de l'application visée. D'autre part, les différents articles ci-dessous peuvent faire référence à la restauration sans pour autant mentionner le nom même de la technique d'*inpainting*. Ainsi, après regroupement selon les applications qu'elles visent, les références bibliographiques suivent le code couleur suivant pour plus de clarté :

- pour les références se consacrant exclusivement à la détection de défauts,
- pour les références se concentrant sur l'*inpainting* pour la restauration, avec potentiellement la détection préalable des défauts,
- pour les références se concentrant sur l'*inpainting* visant d'autres applications, en particulier la réalité diminuée,
- pour les références se concentrant sur les calculs de mouvements entre images, ou plus généralement sur les correspondances entre images.

2.1 Détection des défauts

Les premiers détecteurs de défauts dans les vidéos ont été utilisés à l'origine pour détecter les bruits d'impulsion, comme celui mis en œuvre par [Storey, 1985] pour la BBC, qui utilisait un seuil sur la valeur absolue des différences entre images successives pour détecter les défauts. Cependant, celui-ci ne prenait pas en compte le mouvement apparent entre les images. C'est pourquoi [Kokaram et Rayner, 1992] a introduit l'indice de détection des pics (ou SDI pour *Spike Detection Index*), avec quelques variantes SDIa ou SDIx dans [Kokaram, 2004], où les différences absolues étaient calculées sur des images après compensation du mouvement. Au-delà du bruit, il s'est ensuivi différentes propositions de détection de défauts, en particulier des rayures et des taches, présentées ci-après.

2.1.1 Détection des rayures

Le premier détecteur de rayures a été proposé par [Kokaram, 1996]. Chaque ligne verticale y est modélisée comme une sinusoïde amortie contenue dans l'image. Un sous-échantillonnage et un filtrage vertical sont d'abord appliqués à l'image pour éliminer le bruit. Une image binaire est ensuite obtenue en seuillant la différence entre l'image sous-échantillonnée d'une part, et cette même image après application d'un filtre médian d'autre part. Par la suite, la transformation de Hough du champ binaire résultant du seuillage donne des lignes candidates. Il s'ensuit un raffinage bayésien pour permettre de conserver seulement les lignes correspondant au modèle. Le modèle de décomposition de [Kokaram, 1996] a été généralisé dans [Bruni et Vitulano, 2004] aux films en niveaux de gris. On y retrouve l'ajout d'un coefficient γ pour obtenir une pondération entre l'image bruitée et les lignes verticales. Ce dernier tient compte du rapport entre l'amplitude moyenne du signal et l'amplitude de la rayure. La sélection s'opère suivant la largeur et la hauteur des lignes verticales, puis suivant la loi de Weber-Fechner pour la perception de stimulus par l'œil humain. La généralisation aux films en couleur a ensuite été opérée dans [Bruni *et al.*, 2008]. Une autre amélioration a vu le jour dans [Xu *et al.*, 2007]. Dans un premier temps, une décomposition en ondelettes de l'image contenant les rayures est effectuée sur trois échelles. Les coefficients d'ondelettes horizontaux sont ensuite utilisés pour représenter les contours des lignes. Les rayures réelles parmi toutes les positions potentielles des artefacts sont extraites en appliquant une contrainte sur la hauteur et la largeur. La transformée de Hough a également été utilisée dans [Chishima et Arakawa, 2009] après application d'un filtre de Canny pour obtenir une image de contours.

Dans [Joyeux *et al.*, 1999], le détecteur de rayures a pris la forme d'une différence entre les projections sur l'axe horizontal de l'image d'origine et cette même image après application d'une fermeture morphologique. Une fois tous les candidats potentiels détectés, la seconde étape consiste à opérer un suivi des rayures le long des trames suivantes en utilisant un filtre de Kalman, et ainsi à éliminer les fausses détections présentes à l'étape précédente. Cependant, la projection d'intensité ne fonctionne pas très bien en présence de variations d'environnement le long de la colonne de pixels. C'est pourquoi cette idée de fermeture morphologique a été reprise par [Shih *et al.*, 2006], mais après avoir découpé l'image en plusieurs bandes horizontales pour séparer les différents plans. De plus, chaque bande est subdivisée en fenêtres pour pouvoir chercher un minimum local dans chacune d'elles. Les lignes présentes au même endroit sur plusieurs bandes (fenêtres alignées verticalement) sont ensuite fusionnées, ce qui permet également de trouver des rayures qui ne sont pas présentes sur toute la hauteur de l'image.

Les rayures étant caractérisées par des changements d'intensité conséquents dans la direction de l'axe horizontal, cette caractéristique a été traitée dans [Khriji *et al.*, 2005]. Une dérivée horizontale avec seuillage est d'abord appliquée à l'image pour détecter les changements d'intensité suivant cet axe. Ensuite, la moyenne de chaque colonne est calculée et seuillée à nouveau pour ne conserver que les rayures présentes sur une plus grande partie d'une colonne. Pour leur part, [Newson *et al.*, 2012] ont ajouté une étape après la détection des rayures de [Kokaram, 1996]. Les pixels à gauche et à droite de la rayure sont comparés et, s'ils sont trop différents, la ligne est retirée de la sélection car elle

correspond simplement à un contour vertical. Ensuite, les différents pixels ainsi détectés sont regroupés pour ne former que des lignes verticales par une méthode a contrario. Une autre idée pour limiter les fausses alertes a été introduite dans [Newson *et al.*, 2013]. Elle consiste à réaligner les différentes images de masque en suivant le mouvement apparent et à éliminer celles qui conservent la verticalité, car les rayures ne suivent pas le mouvement apparent. Ces deux dernières méthodes ont ensuite été synthétisées dans [Newson *et al.*, 2014b].

2.1.2 Détection des taches

Dans [Kokaram *et al.*, 1995], deux méthodes de détection de taches ont été développées : l'une avec un modèle de champs aléatoires de Markov, l'autre avec un modèle autorégressif 3D. Le premier d'entre eux a ensuite été utilisé dans [Wang et Mirmehdi, 2012], suivi de deux étapes de raffinage pour l'élimination des fausses alarmes. Les deux contraintes consistent à imposer, d'une part, une continuité spatiale avec un nouveau champ de Markov et, d'autre part, une corrélation temporelle avec un traceur pyramidal de type Lucas-Kanade.

Le détecteur de différences ordonnées par rang (ou détecteur ROD pour *Rank Ordered Differences*) a été introduit par [Nadenau et Mitra, 1997]. Il consiste à prendre les pixels des images situés au voisinage temporel du pixel courant, en particulier d'une colonne de trois pixels voisins dans l'image arrière (de même dans l'image avant). Une fois les six pixels ordonnés, si le pixel courant a une valeur supérieure à la moyenne des deux pixels centraux parmi les six, trois différences absolues entre le pixel courant et les trois plus grands niveaux de gris ordonnés de ces pixels sont calculées (inversement si le pixel courant a une valeur inférieure à la moyenne). Si parmi ces trois distances, au moins une est supérieure au seuil de détection choisi, le pixel courant est détecté comme défaut. Une version simplifiée du détecteur ROD (appelée SROD) a été introduite dans [Biemond *et al.*, 1999], ne tenant compte que des valeurs minimales et maximales parmi les pixels voisins temporellement. Le détecteur SROD a également été utilisé dans [Tilie *et al.*, 2007]. Il y est combiné à un détecteur spatial basé sur la croissance et la fermeture morphologique. Une autre application du détecteur SROD a été introduite dans [Gullu *et al.*, 2008]. Il y est appliqué en deux étapes : la première, classique, avec les images avant et arrière, suivie par une version utilisant les images avec compensation du mouvement, afin d'éliminer des faux positifs. Une variante du SDI de [Kokaram et Rayner, 1992] a également été introduite dans [Buades *et al.*, 2010]. Cette version adaptative (ASDI) est utilisée après l'estimation de mouvement entre images bloc par bloc. La version adaptative vient du fait que le détecteur utilise un seuil variable qui est calculé localement entre les différences d'intensité, pour éviter de détecter du bruit.

Dans [Ren et Vlachos, 2007], la détection a été effectuée comme une segmentation de l'image par croissance de régions. Les images sont parcourues pour identifier tout pixel qui n'est pas encore affecté à une région. Ce pixel est ensuite utilisé comme source pour la croissance de la région. Tous les pixels précédemment fusionnés pour former une région étant marqués, ne sont pas revisités. La croissance de la région est contrôlée en examinant la similarité des pixels adjacents. Une nouvelle mesure de confiance

fondée sur les différences temporelles de l'image y est également introduite. D'autres méthodes ont nécessité plusieurs étapes pour détecter les taches, comme par exemple dans [Xu *et al.*, 2015]. La première étape consiste à trouver, après suppression du bruit et du scintillement, des candidats potentiels pour les taches en fonction de leurs caractéristiques spatiales (taille, intensité, contours...). Ensuite, après compensation du mouvement, les taches réelles sont détectées au travers de la discontinuité temporelle de l'intensité parmi les candidats restants. Une extraction de la région candidate de la tache a, quant à elle, été effectuée dans [Yous et Serir, 2016]. Un filtrage médian dans le domaine temporel est utilisé pour détecter des changements soudains dans une région. Ensuite, chaque région est classée comme une tache ou non, en trouvant la similarité de la région candidate dans les images adjacentes dans l'espace de gradient, avec un histogramme de gradients orientés. Les mêmes auteurs dans [Yous et Serir, 2017] ont allié leur filtre médian temporel à un détecteur SDIx intégrant une stratégie de vote par consensus sur les régions candidates afin de séparer les taches des autres régions candidates.

2.1.3 Détection conjointe et/ou par apprentissage profond

Pour détecter les taches mais aussi les rayures, [Li *et al.*, 2013] a expérimenté une méthode de détection fondée sur la décomposition en *cartoon* + texture dans le domaine spatial, associée à la séparation entre défauts et contenu sain dans le domaine temporel. La distinction entre défauts et zone saine est traitée dans le temps, en utilisant une décomposition matricielle en une matrice de rang faible en plus d'une matrice creuse représentant les défauts.

La première application de l'apprentissage profond dans ce contexte a été consacrée à la détection des rayures dans [Kim et Kim, 2007]. Les images sont d'abord décomposées en forme et texture. Ensuite, la détection de la forme est effectuée par filtrage, tandis que la texture est classée par un réseau de neurones avec les images de contours en entrée. Un détecteur de taches en trois étapes a été proposé dans [Sizyakin *et al.*, 2017], qui consistent en une compensation du mouvement, une détection SROD [Biemond *et al.*, 1999], et enfin une classification de tous les pixels ayant des valeurs anormales en utilisant un réseau de neurones convolutif. Les mêmes auteurs ont proposé une autre approche en trois étapes dans [Sizyakin *et al.*, 2019]. La première étape consiste à créer un descripteur contenant l'intensité des trois images consécutives, l'intensité des images avec compensation de mouvement, l'amplitude du flux optique de Lucas-Kanade, et enfin les motifs binaires locaux (ou LBP pour *Local Binary Patterns*). Ensuite, une détection SDI [Kokaram et Rayner, 1992] est effectuée. Son résultat et le descripteur constituent alors l'entrée d'un CNN. Pour la détection des taches et des rayures, [Yous *et al.*, 2019] a appliqué une classification avec une architecture d'encodeur-décodeur CNN, comportant la concaténation de couches dans la partie encodeur. Ensuite, en utilisant la sortie du réseau avant la dernière convolution pour l'image courante et l'image précédente, une mise en commun de la moyenne spatiale est effectuée. Un seuil sur la distance euclidienne entre les deux résultats permet alors de détecter les taches. Les rayures, quant à elles, sont détectées après la fermeture morphologique de la sortie du réseau et en tenant compte de l'analyse de la forme : les défauts dont la hauteur est beaucoup plus grande que la largeur sont alors conservés.

2.2 Notion d'*inpainting* appliqué aux images

Une fois la détection des défauts effectuée, la correction de ces défauts (en particulier, rayures et taches) fait partie d'une sous-branche d'un domaine d'application que l'on appelle *inpainting*, nom donné à la technique de remplissage des zones endommagées ou manquantes d'une image. Le terme *inpainting* n'est utilisé qu'à partir de 2000 dans [Bertalmio *et al.*, 2000], par analogie avec le procédé de restauration utilisé dans le domaine de l'art, après celui de désoccultation introduit dans [Masnou et Morel, 1998] en 1998. Les premières applications de l'*inpainting* sont issues des modèles de diffusion pour le débruitage, qui remontent au début des années 1990. Ce domaine de recherche a été très actif ces dernières années, stimulé par de nombreuses applications : suppression de rayures ou de texte superposé à une image, restauration d'une image altérée suite à une transmission, élimination d'objets dans un contexte d'édition pour la réalité diminuée.

Le remplissage de zones à restaurer est un problème inverse mal posé car il n'y a pas de solution unique bien définie. Il est donc nécessaire d'introduire des connaissances a priori dans le modèle. Toutes les méthodes existantes sont guidées par l'hypothèse que les pixels situés dans les parties connues et manquantes de l'image partagent les mêmes propriétés statistiques ou structures géométriques. Cette hypothèse se traduit par différentes hypothèses a priori locales ou globales, afin d'obtenir une image restaurée visuellement plausible. Dans le cas de l'*inpainting* par diffusion, on veut propager les informations contenues dans les pixels depuis le bord de la zone endommagée jusqu'à l'intérieur de cette zone. En particulier, la variation totale pour l'*inpainting* a été introduite dans [Chan, 2001] pour bloquer la diffusion aux bords des objets et récupérer des données constantes par morceaux. La diffusion a été effectuée en utilisant des tenseurs dans [Tschumperlé, 2006], afin de régulariser les images tout en tenant compte des courbures. D'autres méthodes basées sur la diffusion peuvent être trouvées et détaillées dans [Guillemot et Le Meur, 2013].

Cependant, ces modèles sont limités car ils sont dans l'incapacité de gérer les textures. C'est pourquoi des modèles se fondant sur la recopie complète ou partielle de *patchs* ont été développés (voir plus de détails dans [Buyssens *et al.*, 2015]) pour conserver les détails correspondant à des fréquences élevées, en commençant par la synthèse de texture dans [Efros et Leung, 1999]. Un autre modèle d'*inpainting* local par *patchs* (les *patchs* sont uniquement recherchés dans un voisinage de la zone du défaut) a été présenté dans [Criminisi *et al.*, 2004]. Il inclut une priorité de remplissage dépendant de l'intensité du gradient spatial aux bords de la zone à remplir. Afin de se rapprocher des méthodes par diffusion, [Aujol *et al.*, 2010] a considéré un point de vue variationnel pour l'*inpainting* par *patchs*. Dans le même ordre d'idée, [Arias *et al.*, 2011] a utilisé un cadre variationnel pour calculer un mélange de *patchs* suite à une recherche spatiale non locale de ces mêmes *patchs*.

Comme même ces méthodes peuvent éprouver quelques difficultés à récupérer des structures régulières, l'idée de mélanger les deux approches a été développée dans [Bugeau et Bertalmio, 2009]. L'image y est décomposée en *cartoon* + texture avant d'être remplie séparément, avec une approche

de diffusion pour la structure, et une approche par *patchs* pour la texture. On retrouve également une autre combinaison entre structure et texture dans [Cao *et al.*, 2011], où la texture est récupérée en étant guidée par les lignes de niveau.

Plus récemment, l'apprentissage profond a introduit un nouveau type d'*inpainting* sémantique, comme dans [Pathak *et al.*, 2016]. Les encodeurs contextuels sont formés avec une fonction de perte de reconstruction en plus d'une fonction de perte antagoniste. Ces encodeurs contextuels ont pour but de comprendre le contenu de l'image entière et de produire une hypothèse plausible pour les parties manquantes. Ce modèle a été amélioré par [Iizuka *et al.*, 2017] en introduisant un discriminateur supplémentaire pour garantir la cohérence de l'image locale. Le discriminateur global évalue si l'image complète est cohérente dans son ensemble, tandis que le discriminateur local se concentre sur une petite zone centrée sur la région générée afin de renforcer la cohérence locale. En outre, l'utilisation des convolutions dilatées sert à élargir le champ de réception, alors que le mélange de Poisson est présent pour affiner l'image. Cependant, les réseaux de neurones convolutifs se révèlent inefficaces pour récupérer explicitement des informations provenant de lieux distants, ce qui est le cas des techniques de *patchs*. Pour pallier cela, [Yu *et al.*, 2018] a proposé une approche utilisant un modèle génératif profond. Ce dernier réussit à synthétiser de nouvelles structures d'images mais aussi à utiliser explicitement les caractéristiques des images environnantes comme références, pendant la formation du réseau, pour faire de meilleures prédictions. Le modèle comporte deux réseaux génératifs successifs : un réseau grossier pour générer une première image, puis un autre utilisant le module d'attention contextuelle pour raffiner cette image intermédiaire et ainsi produire le résultat final de l'*inpainting*. Le modèle de [Iizuka *et al.*, 2017] a récemment été amélioré par [Jiang *et al.*, 2020] en ajoutant une *skip-connection* dans le générateur pour améliorer la puissance de prédiction du modèle et la fonction de perte de GAN de Wasserstein, afin d'assurer la stabilité du processus d'entraînement.

2.3 *Inpainting* vidéo pour la restauration

Là où l'*inpainting* appliqué aux images est considéré comme un problème d'extrapolation car on ne connaît pas les informations manquantes, le passage à la vidéo s'apparente plus, quant à lui, à un problème d'interpolation. En effet, les zones que l'on cherche à reconstruire existent déjà sous une autre forme (déformation, mouvement) dans les images voisines de celle où le défaut est présent.

2.3.1 *Inpainting* vidéo suite à la détection des défauts

Dans [Kokaram, 1996], l'interpolation des rayures a été effectuée en partant de l'hypothèse que les données dans l'image étaient générées à partir d'un système autorégressif 2D. Plutôt que d'estimer les valeurs manquantes en utilisant les moindres carrés, une approche probabiliste est adoptée, évitant ainsi les problèmes d'excitation tendant vers zéro dans les grandes régions à remplir. Elle génère des interpolations en échantillonnant la densité de probabilité a posteriori pour les données inconnues, étant donné l'hypothèse que l'excitation provient d'une distribution normale.

Dans [Joyeux *et al.*, 1999], la technique de restauration utilise la somme d'une version contenant les basses fréquences et d'une version contenant les hautes fréquences de l'image. Pour les basses fréquences, un modèle polynomial d'ordre 3 est utilisé, avec une estimation des paramètres en moindres carrés sur un voisinage de la rayure. Pour les hautes fréquences, une fois les basses fréquences soustraites à l'image, le résidu est modelisé sous la forme d'une série de Fourier. Les coefficients sont déterminés l'un après l'autre à partir de celui d'indice 0, jusqu'à ce que la somme des moindres carrés à l'indice courant passe en dessous d'un certain seuil.

Faisant suite au modèle de détection proposé dans [Bruni et Vitulano, 2004], [Bruni *et al.*, 2004] a étendu le processus à la restauration. Tout d'abord, une transformée en ondelettes multi-échelles est calculée. Ensuite, le coefficient de pondération γ est calculé à chaque étage pour pouvoir éliminer les rayures. Dans [Xu *et al.*, 2007], les défauts ont également été éliminés par une méthode de lissage par ondelettes multi-échelles. Ensuite, une transformation inverse par ondelettes est effectuée pour obtenir l'image corrigée. Pour la restauration en couleur dans [Bruni *et al.*, 2008], la même démarche a été opérée avec les ondelettes. Dans un premier temps, le canal ayant subi le plus gros impact de la part de la rayure est sélectionné pour être restauré. Une fois la rayure restaurée dans ce canal, la nouvelle luminance est ensuite évaluée pour savoir si le canal suivant doit également être restauré.

Dans [Khriji *et al.*, 2005], la restauration des rayures a été effectuée à partir d'une moyenne pondérée des pixels situés au voisinage du pixel défectueux. La méthode de restauration de [Shih *et al.*, 2006] reprend celle de [Shih *et al.*, 2004]. Les différents éléments de l'image sont séparés en différentes couches suivant les couleurs. Ensuite, la partie *inpainting* est réalisée dans un cadre de multirésolution, où une moyenne spatiale ou temporelle est appliquée au cas par cas, en fonction du nombre de pixels dans chaque défaut. La méthode de détection de [Shih *et al.*, 2006] a également été utilisée dans [Kao *et al.*, 2007]. Les pixels de contour des rayures sont ensuite regroupés dans un vecteur B. Les solutions pour les pixels du contour intérieur sont alors obtenues par l'opération $A^{-1}B$, où A est une matrice tridiagonale de laplaciens. L'opération est répétée jusqu'à obtenir un remplissage complet de chaque couche de contour intérieur contenu dans la rayure.

Dans [Gullu *et al.*, 2008], la méthode proposée utilise une stratégie de correction utilisant les contours similaire à celle de [Criminisi *et al.*, 2004], mais avec un critère de priorité différent, fondé sur la différence entre le maximum et le minimum d'intensité entre les pixels du voisinage du bord du défaut. Ensuite, une stratégie de recherche du *patch* optimal est opérée dans l'image précédente et l'image suivante, en considérant que la tache n'est pas présente dans ces dernières.

La restauration des rayures dans [Chishima et Arakawa, 2009] est opérée par un filtre médian. La particularité est que, si une ligne non verticale est traversée par la rayure, l'application simple d'un filtre médian horizontal va créer une distorsion. Par conséquent, en présence de ce genre de ligne, le filtre médian est appliqué suivant l'orientation de la ligne traversée pour les pixels concernés, et horizontalement sinon.

La restauration dans [Wang et Mirmehdi, 2012] a été réalisée en plusieurs étapes. Une marche aléatoire part d'un pixel dégradé et s'arrête lorsqu'elle atteint un changement significatif dans toutes les caractéristiques des pixels (intensité, mouvement, texture). Après avoir construit la région des pixels candidats pour un pixel dégradé, une probabilité de remplacement est affectée en calculant d'abord la moyenne des probabilités de transition au cours de chaque marche aléatoire qui part du pixel dégradé et qui visite le pixel candidat. Ensuite, les moyennes des probabilités de ces marches aléatoires sont additionnées pour obtenir une mesure de la similarité entre le pixel candidat et les autres pixels de la région. Plus cette valeur est élevée, plus la similarité est grande. Ainsi, le pixel candidat ayant la plus grande probabilité est sélectionné pour remplacer le pixel dégradé de la cible.

Dans [Yous et Serir, 2017], la restauration des taches est effectuée en quatre étapes. Tout d'abord, le mouvement des pixels situés sur le contour extérieur de la tache est estimé. Ensuite, les différents pixels qui ont un mouvement similaire sont regroupés dans des *clusters* par la méthode de partitionnement des k-moyennes. Chaque pixel du contour intérieur se voit ensuite affecter le même mouvement que le *cluster* le plus proche spatialement. Enfin, ces pixels sont restaurés en appliquant une interpolation de l'image précédente par le mouvement qu'ils ont reçu. Ces étapes sont réitérées jusqu'au remplissage complet de la tache.

2.3.2 *Inpainting* vidéo par diffusion

Avant l'*inpainting* vidéo, la première extension à la vidéo des modèles par diffusion a été le débruitage vidéo dans [Kornprobst *et al.*, 1998]. Ensuite, l'extension de l'*inpainting* par diffusion à la vidéo a commencé avec [Cocquerez *et al.*, 2003] et [Lauze et Nielsen, 2004] (avec son extension à plusieurs variations de l'énergie dans [Lauze et Nielsen, 2018]). Le mouvement est simultanément estimé pour remplir la zone endommagée. Dans ce cas, la base des modèles pour estimer le mouvement est le modèle de flux optique bien connu de [Horn et Schunck, 1981] avec une régularisation lisse en norme L^2. À partir de celui-ci, [Aubert *et al.*, 1999] a transposé la régularisation à la norme L^1 afin de préserver les discontinuités des différents mouvements. L'estimation du mouvement en norme L^1 a ensuite été résolue à l'aide d'algorithmes proximaux dans [Zach *et al.*, 2007].

Parmi les différentes applications de l'*inpainting* vidéo, certaines ne concernent pas la restauration ou la diminution de la réalité en tant que telle, mais sont plutôt orientées vers l'amélioration des vidéos. Par exemple, dans [Matsushita *et al.*, 2006], l'*inpainting* vidéo a été utilisé pour stabiliser les vidéos en remplissant les côtés des images. Pour ce faire, des estimations de mouvement sont effectuées à la fois au niveau global [Anandan, 1989, Bergen *et al.*, 1992] et au niveau local [Lucas et Kanade, 1981]. La première estimation, qui donne un mouvement lisse, sert à déflouter l'image, mais aussi à initialiser l'estimation locale. Cette dernière, quant à elle, effectue l'*inpainting* du mouvement de manière plus précise. Les deux estimations sont opérées dans un schéma multirésolution. Ensuite, le meilleur pixel à remplir est choisi comme la médiane des meilleurs candidats le long du flux optique.

Une autre application particulière se trouve dans [Keller *et al.*, 2008], où la technique a été utilisée pour le désentrelacement des images. En particulier, les pixels manquants sont répartis en damier le long de l'axe temporel. Avec cette même répartition régulière des pixels manquants, [Keller *et al.*, 2011] a également appliqué la technique dans le cadre de la super-résolution vidéo.

2.3.3 Autres méthodes d'*inpainting* vidéo pour la restauration

La même idée de transfert du champ de mouvement dans [Matsushita *et al.*, 2006] a été reprise dans [Shiratori *et al.*, 2006]. Cependant, la minimisation s'effectue sur des distances entre *patchs* de mouvement. Ensuite, le processus d'*inpainting* est obtenu en calculant, pour chaque pixel, une moyenne pondérée suivant le mouvement.

Depuis 2017, des méthodes d'*inpainting* vidéo utilisant l'apprentissage profond ont vu le jour. La première application a été l'interpolation d'images supplémentaires entre deux images successives. La première méthode de [Niklaus *et al.*, 2017a] utilise un réseau neuronal convolutif avec deux trames et des noyaux convolutifs 2D. Une autre méthode, toujours par les mêmes auteurs dans [Niklaus *et al.*, 2017b], utilise quant à elle des noyaux convolutifs 1D. Ce faisant, le second réseau nécessite moins de paramètres que le premier.

2.4 *Inpainting* vidéo pour la réalité diminuée

La restauration n'est pas la seule application de l'*inpainting*. En effet, de nombreuses recherches récentes se sont orientées vers la réalité diminuée. Dans ce contexte, le but n'est plus de retrouver un élément manquant dans la vidéo, mais d'enlever un élément du premier plan clairement défini (personne, objet, ...) qui est récurrent temporellement dans les images. Sans doute plus à la mode pour son côté trucage et son lien avec la réalité augmentée utilisée par des applications sur les *smartphones*, ces techniques, largement inspirées des approches par *patchs* et par apprentissage profond, n'en restent pas moins adaptables, sous certaines conditions, à la restauration.

2.4.1 *Inpainting* vidéo par recopie de *patchs*

De même que pour l'image, l'approche par *patchs* de l'*inpainting* a également été étendue aux vidéos. En particulier [Wexler *et al.*, 2004] (et son extension [Wexler *et al.*, 2007]) a introduit la cohérence entre une image défectueuse et une image voisine comme une fonction objectif maximisant les mesures de similarité entre les *patchs* 3D spatio-temporels.

Dans [Patwardhan *et al.*, 2005], l'idée a été de séparer le premier plan de l'arrière-plan pour effectuer l'*inpainting*. Le remplissage de l'arrière-plan fixe est réalisé en utilisant une version temporelle de [Criminisi *et al.*, 2004] suivie d'un remplissage spatial pour les parties restantes. L'*inpainting* du premier plan, mobile quant à lui, est effectué en recherchant la partie mobile la mieux adaptée dans

les autres images pour la recopier. Si [Patwardhan *et al.*, 2005] ne se concentrait que sur des plans statiques, l'algorithme a été généralisé pour une caméra en mouvement dans [Patwardhan *et al.*, 2007]. La même décomposition d'image est présente dans [Cheung *et al.*, 2006]. L'*inpainting* temporel du fond est réalisé avec des filtres de Kalman, tandis que l'*inpainting* dynamique de l'objet est réalisé en choisissant les meilleurs candidats interpolés à partir des autres images.

Des *patchs* spatio-temporels ont aussi été recherchés dans [Newson *et al.*, 2014a]. Pour ce faire, leur méthode utilise une extension de l'algorithme PatchMatch de [Barnes *et al.*, 2009] à la dimension temporelle, afin d'obtenir une certaine similarité temporelle dans les comparaisons de *patchs*. La mesure de similarité inclut également des *patchs* de texture, composés de la valeur absolue des gradients de l'image. Ce modèle a par la suite été amélioré dans [Le *et al.*, 2017]. Un troisième terme dans la distance entre *patchs* a été ajouté, égal au module du vecteur de flux optique, pour une meilleure cohérence temporelle. La propagation guidée par le flux avec des *patchs* a, quant à elle, été introduite dans [Huang *et al.*, 2016]. Cette propagation permet de faire tourner les *patchs* en suivant les vecteurs de flux locaux, ce qui produit une prédiction précise de la position et de la transformation des *patchs* candidats. Cette approche améliore la cohérence temporelle par rapport aux approches spatio-temporelles par *patchs* de [Newson *et al.*, 2014a].

Si ces modèles donnent de meilleurs résultats que la diffusion en matière de récupération de la texture, ils sont toutefois fortement dépendants du remplissage initial de la zone. Si ces algorithmes restent bloqués dans un minimum local vis-à-vis des différences entre *patchs* lors de la recherche de minimum, ils échouent généralement à reconstruire des structures régulières. C'est pourquoi de nouvelles méthodes ont été mises en œuvre pour améliorer le remplissage.

2.4.2 Autres méthodes d'*inpainting* vidéo pour la réalité diminuée

Dans le cadre de l'application qu'est la réalité diminuée, la combinaison entre diffusion et *patchs* a aussi vu le jour. En particulier, la méthode d'*inpainting* de [Criminisi *et al.*, 2004] a été adaptée à la vidéo par [Daisy *et al.*, 2015]. L'extension à la vidéo amène à l'utilisation de tenseurs spatio-temporels pour mélanger les *patchs* afin de réduire les artefacts typiques des effets de bloc.

Avec l'idée qu'il est plus facile de compléter un flux optique qu'une image, [Xu *et al.*, 2019] a développé un réseau profond de complétion de flux optique (DFC-net pour *Deep Flow Completion*). Le réseau en lui-même est composé de trois sous-réseaux pour effectuer une estimation de plus en plus précise du mouvement. Comme la majorité du flux est lisse, la fonction de perte du réseau contient également un terme de régularisation afin de donner un poids plus important aux zones de contour contenues dans le flux, zones qui sont les plus difficiles à estimer correctement. Ensuite, le flux est utilisé pour guider la propagation des pixels dans les images. Si certains pixels restent encore inconnus après cette étape, ils sont remplis à partir d'une méthode d'*inpainting* pour l'image.

Le terme d'*inpainting* vidéo profond a été introduit dans [Kim *et al.*, 2019a] avec le réseau portant

le nom de VINet (pour *Video Inpainting Network*), prenant la forme d'un encodeur-décodeur. La partie encodeur est subdivisée en deux entreés : d'une part, l'image à remplir, et d'autre part, quatre images voisines ainsi que le résultat de l'*inpainting* sur l'image précédente. Pour chaque couche du réseau, ces cinq images sont utilisées pour estimer le mouvement entre elles avec le réseau FlowNet2 de [Ilg *et al.*, 2017]. À partir de ces champs de vecteurs, les images sont alignées avec l'image courante avant d'être agrégées pour ne plus former qu'une seule image de caractéristiques. Celle-ci est alors additionnée à l'image de caractéristiques courante en ne prenant que la zone à remplir. La partie décodage est opérée en concaténant les nouvelles images de caractéristiques de chaque couche à chaque remontée. La fonction de perte est évaluée sur l'estimation du mouvement entre la prédiction de l'image courante et celle de l'image précédente en plus de la différence entre la prédiction de l'image courante et celle de l'image précédente avec compensation du mouvement.

Un autre modèle de réseau, nommé BVDNet (pour Blind Video Decaptioning Network), a été développé par les mêmes auteurs dans [Kim *et al.*, 2019b]. Cette fois-ci, le mouvement entre images n'est plus estimé. Les deux entrées dans la partie encodeur sont, d'une part, n images autour de l'image courante, elle-même comprise, et, d'autre part, la prédiction de l'image précédente. Une fois les deux encodages effectués, ces derniers sont agrégés avant que ne vienne la partie décodage. La fonction de perte prend toujours en compte la différence entre la prédiction de l'image courante et celle de l'image précédente avec compensation du mouvement, mais aussi la différence entre la prédiction de l'image courante et l'image courante de départ sans défaut.

Chapitre 3

Détection des défauts

Nous nous plaçons dans le contexte de la détection des défauts contenus dans les images de vieux films stockés sur pellicule par la Cinémathèque de Toulouse. Comme évoqué précédemment, ces films ont subi des altérations dues au temps et aux manipulations.

Nous proposons une détection de ces artefacts construite à partir de connaissances d'experts. Nous partons d'un ensemble de données contenant les images d'un film détérioré ainsi que la restauration de ce film par un expert. À partir de ces données, nous souhaitons former un réseau U-Net [Ronneberger *et al.*, 2015], qui prend en entrée plusieurs images successives pour analyser les incohérences temporelles du niveau de gris dues aux défauts. Cette architecture est très utilisée pour ce type de problème pour obtenir une segmentation binaire avec un jeu de données qui ne contient pas énormément d'éléments. Cependant, il faut dans un premier temps, trouver les masques des défauts associés à chaque image détériorée, en la comparant avec l'image restaurée. Ces masques sont souvent inaccessibles, soit parce qu'ils n'ont pas été sauvegardés par le restaurateur, soit parce qu'il s'agit de variables intermédiaires cachées dans le logiciel de restauration. Une fois l'ensemble des masques obtenu, nous formons alors le réseau U-Net. En résumé, nos principales contributions sont les suivantes :

- Nous apprenons à identifier les défauts grâce à l'expertise d'un restaurateur de films aidé par les traitements automatiques d'un logiciel spécialisé, ce qui constitue, à notre connaissance, une nouvelle approche de l'identification des défauts.
- Dans ce but, nous construisons un pipeline générant un ensemble de masques de défauts en comparant les images défectueuses et restaurées.
- Nous construisons un réseau U-Net qui effectuera la détection automatique des défauts.

(a) Séquence de trois images consécutives dans laquelle l'image centrale comporte des rayures et des taches.

(b) La même séquence d'images après la restauration par un expert de la Cinémathèque de Toulouse.

FIGURE 3.1 – Images numérisées d'un vieux film et sa restauration semi-automatique. Notre réseau U-Net calcule, par inférence, un masque de défauts à partir de la seule séquence originale (a). L'étape d'apprentissage du réseau, quant à elle, se fonde sur les séquences défectueuse (a) et restaurée (b) pour récupérer un masque de défauts résultant d'un traitement approprié sur les différences entre images.

3.1 Préparation des données pour le réseau de neurones

Notre jeu de données est un film composé de deux ensembles d'environ 3000 images en niveaux de gris : le premier ensemble contient des images originales, le second est constitué des mêmes images après restauration par un expert de la Cinémathèque de Toulouse (voir FIGURE 3.2) qui s'aide du logiciel de restauration DIAMANT-Film [1]. À partir d'une paire d'images (image défectueuse et image restaurée), nous avons créé des masques de zones défectueuses que nous utilisons comme sortie à prédire par le réseau de neurones.

1. https ://www.hs-art.com/index.php/solutions/diamant-film

(a) Image défectueuse I_A^T. (b) Image défectueuse I_A^F. (c) Image défectueuse I_A^M.

(d) Zoom de (a). (e) Zoom de (b). (f) Zoom de (c).

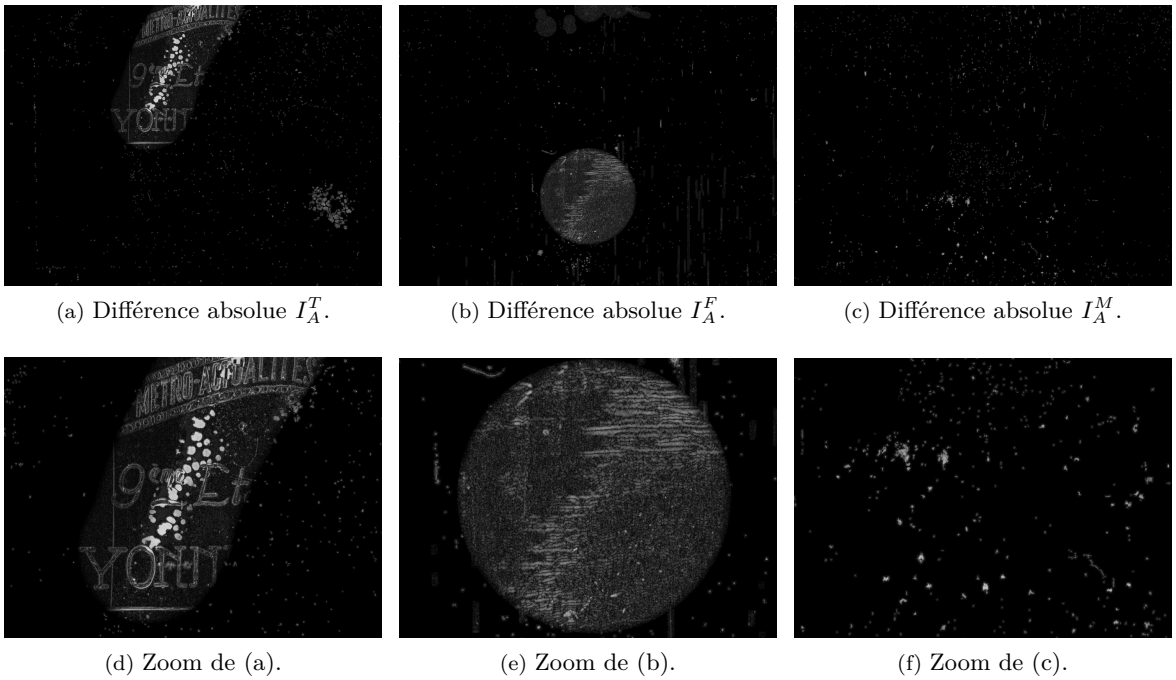
(g) Image restaurée I_A^T. (h) Image restaurée I_A^F. (i) Image restaurée I_A^M.

(j) Zoom de (g). (k) Zoom de (h). (l) Zoom de (i).

FIGURE 3.2 – Exemples de paires d'images défectueuses et restaurées I_A^T, I_A^F et I_A^M. Le restaurateur s'aide du logiciel DIAMANT-Film. Les différentes images du film représentent des scènes de texte (a) ou des scènes naturelles, constituées de plans fixes (b) ou en mouvement (c).

3.1.1 Création des masques par différences seuillées entre images

L'idée pour obtenir les masques est de calculer la différence absolue entre l'image restaurée et l'image défectueuse (voir FIGURE 3.3). Nous effectuons ensuite un seuillage sur cette différence, qui se situe entre 0 et 65535 pour des images en 16 bits.



(a) Différence absolue I_A^T. (b) Différence absolue I_A^F. (c) Différence absolue I_A^M.

(d) Zoom de (a). (e) Zoom de (b). (f) Zoom de (c).

FIGURE 3.3 – Différence absolue entre les images défectueuses et restaurées I_A^T, I_A^F et I_A^M de la FIGURE 3.2. Si les différences les plus importantes sont clairement visibles, les formes géométriques issues des outils du logiciel de restauration DIAMANT-Film peuvent également être facilement reconnues.

Comme on peut le voir sur la FIGURE 3.3, les plus grands défauts sont repérés par l'expert et sélectionnés manuellement, à l'aide de formes géométriques simples. Malheureusement, le logiciel opère la recopie d'une partie d'une image voisine sur l'ensemble de la sélection pour effectuer la restauration de l'image, même si certaines parties n'ont pas lieu d'être restaurées. C'est pourquoi le choix du seuillage, qui n'a rien de trivial, est dicté par un compromis entre la détection d'un nombre suffisant de pixels défectueux et une sur-détection des pixels due à la recopie d'images voisines. Dans cette optique, nous avons étudié la répartition des différents changements de niveau de gris sur l'ensemble des pixels restaurés par le logiciel (voir FIGURE 3.4).

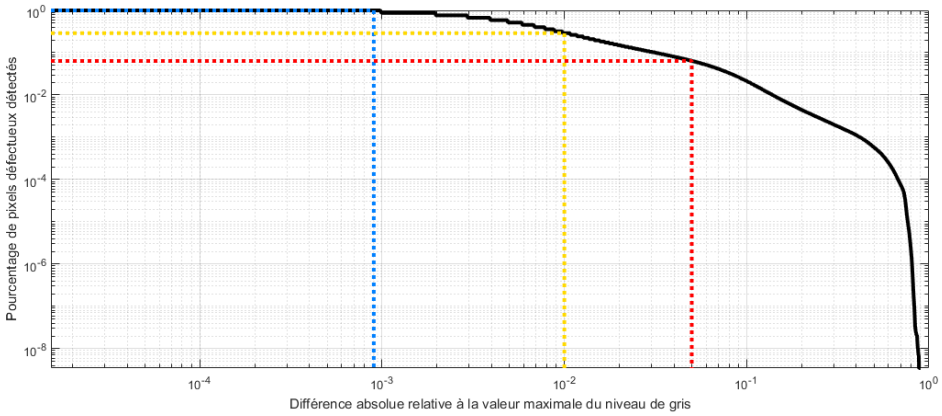
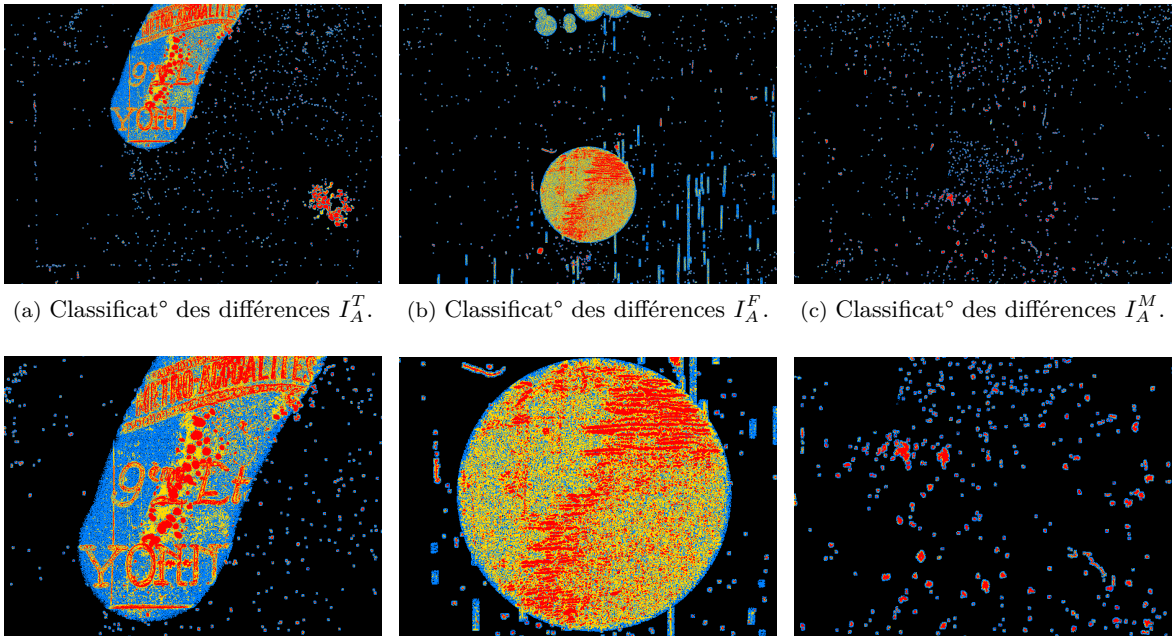


FIGURE 3.4 – Proportion cumulée des pixels défectueux. Le seuil égal à 0,09% de la valeur maximale du niveau de gris (en bleu) est le seuil minimal permettant de détecter tous les changements de niveaux de gris. Un seuil égal à 1% de cette valeur maximale (en jaune) correspond à 30% des pixels restaurés. Un seuil égal à 5% (en rouge) correspond à 6% des pixels restaurés.



(a) Classificat° des différences I_A^T. (b) Classificat° des différences I_A^F. (c) Classificat° des différences I_A^M.

(d) Zoom de (a). (e) Zoom de (b). (f) Zoom de (c).

FIGURE 3.5 – Classification des différences d'images I_A^T, I_A^F et I_A^M de la FIGURE 3.3 correspondant aux trois seuils de la FIGURE 3.4. Les pixels bleus correspondent au seuil minimal et, en ce qui concerne les images originales de la FIGURE 3.2, ne devraient pas être sélectionnés. Par ailleurs, il est difficile de considérer les lettres à l'intérieur de la sélection manuelle comme de véritables défauts, malgré une forte variation de niveau de gris.

Avec les différents seuils choisis expérimentalement, nous avons estimé la limite de la sur-détection à au moins 1% de la valeur maximale du niveau de gris, ce qui ne représente que 30% des pixels qui ont été restaurés. En particulier, avec un seuil à 5% de cette valeur maximale, nous pouvons distinguer les formes des différents défauts (voir FIGURE 3.5), même si la correction apportée aux lettres reste problématique. En revanche, avec un seuil plus élevé, les défauts réels n'ont pas tous été correctement détectés. Même avec un seuil soigneusement choisi, il y a des « trous » dans la détection car certains pixels restaurés peuvent voir leur niveau de gris inchangé.

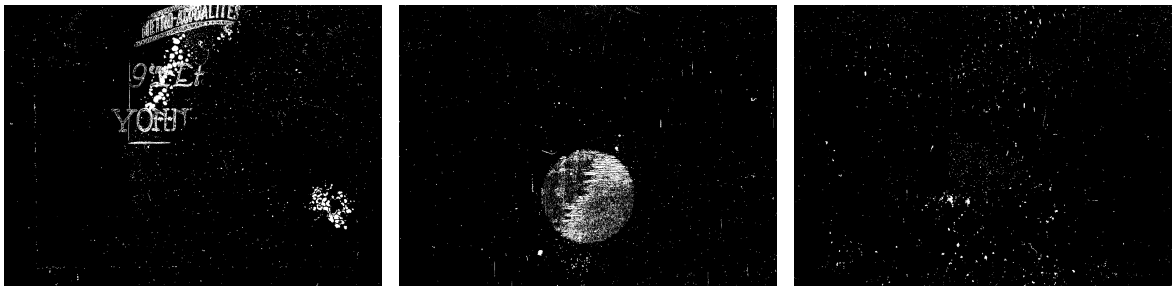
3.1.2 Remplissage automatique des masques par fermeture morphologique

Pour résoudre le problème des pixels défectueux non détectés entourés de pixels défectueux ayant été détectés, nous avons décidé de remplir ces zones en utilisant une fermeture morphologique après l'étape de seuillage, pour récupérer la connectivité spatiale. Les noyaux des filtres morphologiques qui sont utilisés dépendent fortement des types de défauts présents dans les images. Comme expliqué précédemment, il s'agit principalement de taches qui ont des formes globalement rondes, ainsi que de rayures qui constituent généralement des lignes verticales. Le paramètre important à connaître ici est la taille de la fermeture à effectuer sur les masques I_{masque}. Pour le faire automatiquement, nous choisissons des tailles de fermeture initiales assez petites T_{ligne} et T_{disque}, puis nous incrémentons T_{ligne} de 1 à chaque itération. La condition d'arrêt pour la taille de fermeture correcte est atteinte lorsque le nombre de nouveaux pixels qui deviennent des pixels de masque augmente plus que lors de l'itération précédente (voir **Algorithme 1**).

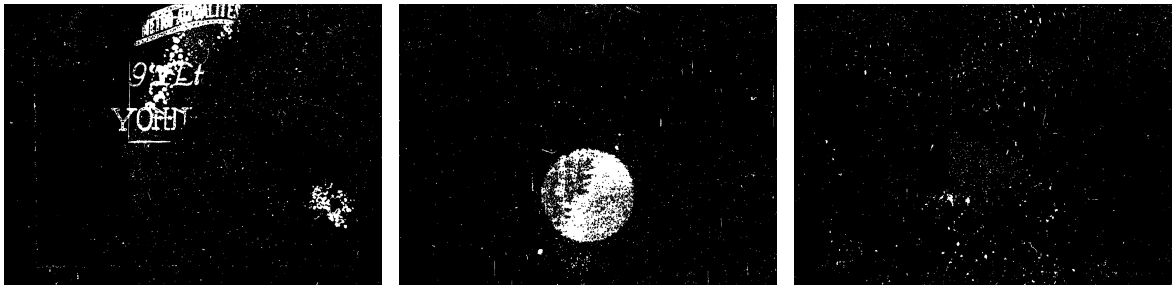
Algorithme 1 - Fermeture morphologique des masques

$T_{\text{ligne}} \leftarrow 3$, $T_{\text{disque}} \leftarrow 2$, I_{masque}
$\Delta \leftarrow +\infty$, $\Delta^* \leftarrow \#(I_{\text{masque}})$
tant que $\Delta > \Delta^*$ **faire**
    $I^*_{\text{masque}} \leftarrow \text{fermeture}\,(I_{\text{masque}}, T_{\text{ligne}})$
    $I^*_{\text{masque}} \leftarrow \text{fermeture}\,\left(I^*_{\text{masque}}, T_{\text{disque}}\right)$
    $\Delta \leftarrow \Delta^*$
    $\Delta^* \leftarrow \#(I^*_{\text{masque}} - I_{\text{masque}})$
    $I_{\text{masque}} \leftarrow I^*_{\text{masque}}$
    $T_{\text{ligne}} \leftarrow T_{\text{ligne}} + 1$
fin tant que

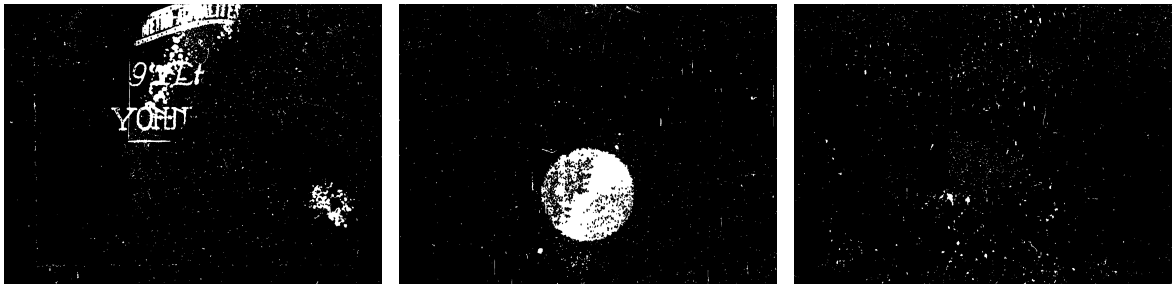
Les résultats de notre algorithme pour la fermeture des masques, toujours avec les mêmes images, sont présentés sur la FIGURE 3.6. Cet algorithme est particulièrement efficace pour combler les gros défauts (voir FIGURE 3.6) et pour relier les différentes parties d'une rayure.



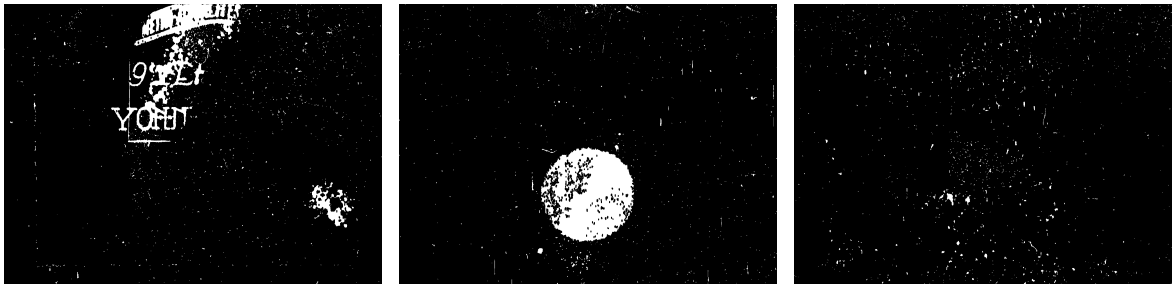
(a) Seuillage des différences absolues (seuil égal à 4% du niveau de gris maximal).



(b) Première itération de la fermeture morphologique.



(c) Deuxième itération de la fermeture morphologique.



(d) Dernière itération de la fermeture morphologique.

FIGURE 3.6 – Fermeture morphologique des masques I_A^T, I_A^F et I_A^M après seuillage. Chaque forme est automatiquement remplie avec notre algorithme.

3.1.3 Quelques statistiques spatiales et temporelles sur les masques

Afin d'avoir des éléments explicatifs sur les bonnes ou mauvaises détections à venir de notre réseau, nous avons organisé tous les défauts d'un point de vue spatial et d'un point de vue temporel. Les défauts sont divisés manuellement pour définir trois groupes à l'aide de deux seuils : petits défauts, grands défauts verticaux et autres grands défauts (voir FIGURE 3.7). Les petits défauts sont séparés des autres en fonction de leur taille, relativement au nombre de pixels dans une image. Ensuite, les défauts verticaux sont choisis en fonction de leur orientation. Les résultats obtenus avec nos masques sont présentés sur la FIGURE 3.8.

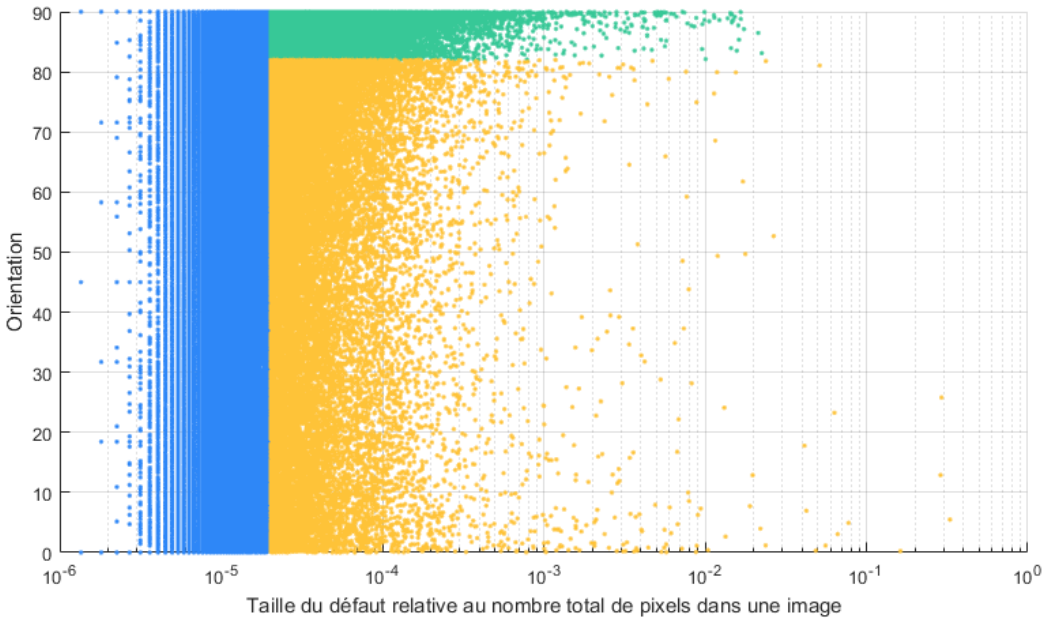
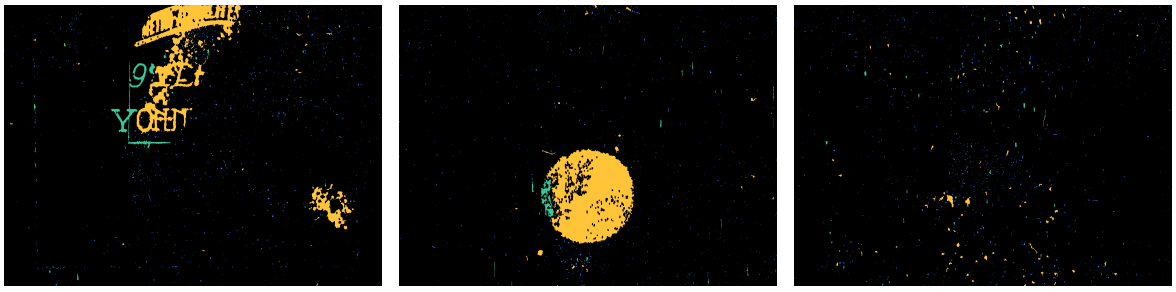
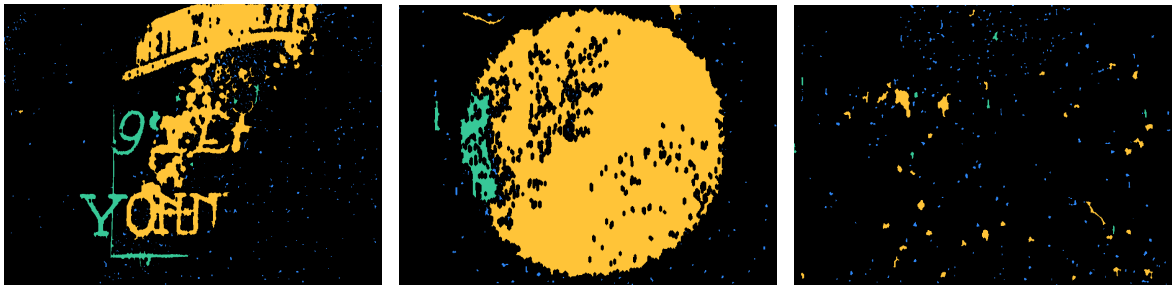


FIGURE 3.7 – Répartition des défauts en fonction de la taille et de l'orientation : petits (en bleu), verticaux (en vert) ou grands (en jaune). Les deux seuils utilisés sont le nombre de pixels pour délimiter les petits défauts, et l'orientation pour séparer les défauts verticaux des autres grands défauts.

Pour l'analyse temporelle, nous avons recherché les différentes « profondeurs » du défaut, pour connaître la corrélation temporelle de certains d'entre eux. En particulier, nous avons construit une image avec les profondeurs maximales des défauts pour chaque pixel, regroupées ensuite sous forme d'histogramme sur la FIGURE 3.9. À l'exception de quelques grands défauts qui peuvent être partiellement superposés dans deux trames successives, les seuls défauts ayant une profondeur importante sont les rayures verticales. Au total, 72% des pixels ont une profondeur de défaut maximale égale à un seulement, et 95% d'entre eux ont une profondeur maximale inférieure ou égale à deux. Par conséquent, l'utilisation de trois images consécutives pour l'entraînement devrait permettre de détecter la grande majorité des défauts.

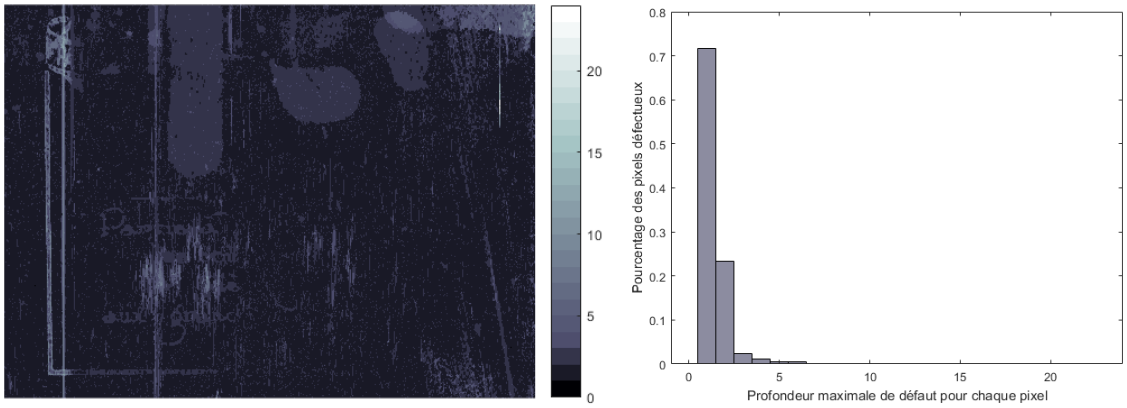


(a) Classification des défauts I_A^T. (b) Classification des défauts I_B^T. (c) Classification des défauts I_C^T.



(d) Zoom de (a). (e) Zoom de (b). (f) Zoom de (c).

FIGURE 3.8 – Exemples de masques (avec un zoom) dont les différentes couleurs correspondent à des petits défauts (en bleu), des lignes verticales ou des défauts orientés verticalement (en vert) et d'autres grands défauts (en jaune).



(a) Profondeur maximale des défauts pour chaque pixel. (b) Histogramme normalisé de (a).

FIGURE 3.9 – Profondeurs temporelles maximales des différents défauts. Les défauts ayant la plus grande profondeur en (a) sont des rayures verticales. On peut également identifier certaines grandes formes sélectionnées par l'expert. Pour 72% des pixels, les défauts ont une profondeur maximale égale à un. Ce pourcentage atteint 95% avec une profondeur maximale inférieure ou égale à deux.

3.2 Expérimentations sur le réseau de neurones proposé

3.2.1 Séparation des données pour l'apprentissage

Pour rappel, notre ensemble de données se compose d'un total de 2974 images de taille 1728×1280, qui sont réparties en 23 scènes. La première étape consiste à les répartir en trois types de scènes : celles comportant seulement un texte explicatif ou descriptif, celles avec un plan fixe et celles où la caméra est en mouvement (voir TABLE 3.1).

TABLE 3.1 – Répartition des scènes et des images en trois types de scènes.

Types de scènes	**Nombre de scènes**	**Nombre d'images**
Texte	7 (30%)	931 (31%)
Caméra fixe	9 (40%)	1105 (37%)
Caméra mobile	7 (30%)	938 (32%)

La deuxième étape consiste à répartir les scènes en trois ensembles de données pour l'apprentissage, la validation et le test. Pour l'ensemble de validation et l'ensemble de test, nous choisissons une scène de chaque type, ce qui signifie trois scènes pour chacun de ces deux ensembles, et 17 pour l'apprentissage (voir TABLE 3.2). En effet, chaque ensemble doit être suffisamment représentatif de la population cible, en sachant que l'ensemble d'apprentissage doit être significativement plus important car c'est lui qui est utilisé pour régler les nombreux paramètres du réseau.

TABLE 3.2 – Répartition des scènes, des images et des triplets de *patchs* entre l'ensemble d'apprentissage, l'ensemble de validation et l'ensemble de test.

Jeu de données	**Nombre de scènes**	**Nombre d'images**	**Nombre de triplets de *patchs* de taille 512×512**
Ensemble d'apprentissage	5+7+5	2296 (77%)	27144
Ensemble de validation	1+1+1	366 (12%)	4320
Ensemble de test	1+1+1	312 (11%)	3672

3.2.2 Modèle U-Net avec des *patchs* spatio-temporels

Nous avons décidé d'utiliser un réseau de neurones de type U-Net, conçu à l'origine pour la segmentation des images biomédicales [Ronneberger *et al.*, 2015]. L'un des problèmes récurrents avec l'utilisation de ces réseaux de neurones est qu'ils impliquent le codage des tailles d'images d'entrée et de sortie en plus des poids. Par conséquent, le réseau ne peut pas être utilisé, une fois formé, avec des images de taille différente de celles des images utilisées pendant l'entraînement. La solution du redimensionnement, qui peut modifier le rapport des images d'entrée, ne peut être considérée comme satisfaisante. Nous nous sommes donc débarrassés de cette contrainte par un entraînement utilisant des *patchs* plutôt que l'image entière. Ce choix n'est pas un problème dans notre cas d'utilisation particulier, car les défauts peuvent se produire de manière aléatoire n'importe où dans l'image. L'étape de prédiction est également effectuée sur l'image découpée en *patchs* se chevauchant. Les *patchs* sont par la suite fusionnés pour reformer l'image complète. Une autre conséquence de cette opération est l'augmentation de la taille du jeu de données (voir TABLE 3.2). Par rapport au réseau U-Net original utilisé avec une seule image en couleur, nous avons utilisé trois *patchs* temporellement consécutifs en entrée (voir TABLE 3.2). Le masque de défauts du *patch* central est utilisé pour la comparaison avec la sortie dans la fonction de perte du réseau (voir FIGURE 3.10).

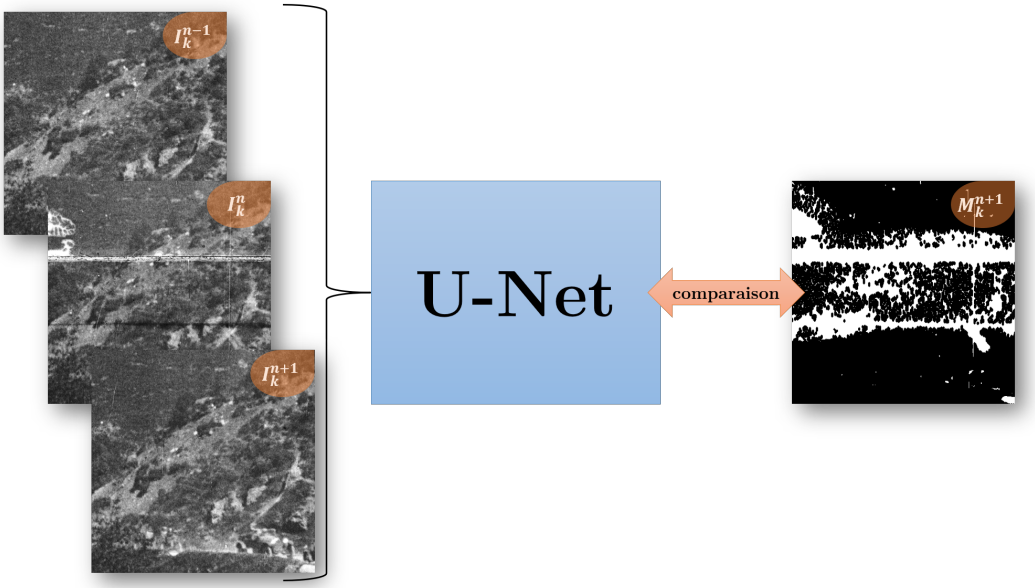


FIGURE 3.10 – Vue globale du réseau U-Net : trois *patchs* d'images avant restauration et temporellement consécutifs, en entrée du réseau, et le masque associé au *patch* central, obtenu par la méthode décrite précédemment, qui permet d'effectuer la comparaison avec la sortie du réseau.

Notre architecture U-Net est illustrée sur la FIGURE 3.11. Elle contient une partie encodage (contraction) et une partie décodage (expansion) comme dans [Ronneberger *et al.*, 2015], mais avec

sept couches dans notre cas. La partie encodage consiste à appliquer successivement deux convolutions de taille 3×3, chacune étant suivie d'une unité linéaire exponentielle (ou ELU pour *Exponential Linear Unit*). Il s'ensuit une opération de *max pooling* de taille 2×2 pour le sous-échantillonnage. À chaque étape de sous-échantillonnage, le nombre de canaux de caractéristiques est doublé. Chaque étape de la partie décodage consiste en un sur-échantillonnage des caractéristiques suivi d'une convolution de taille 2×2 qui divise par deux le nombre de canaux de caractéristiques, suivi d'une concaténation avec les caractéristiques correspondantes de la partie encodage, et de deux convolutions de taille 3×3, chacune étant suivie d'une ELU. À la couche finale, une convolution de taille 1×1 suivie d'une sigmoïde est utilisée pour obtenir une image indiquant les deux classes (défaut ou non) avec des valeurs proches de 1 ou 0.

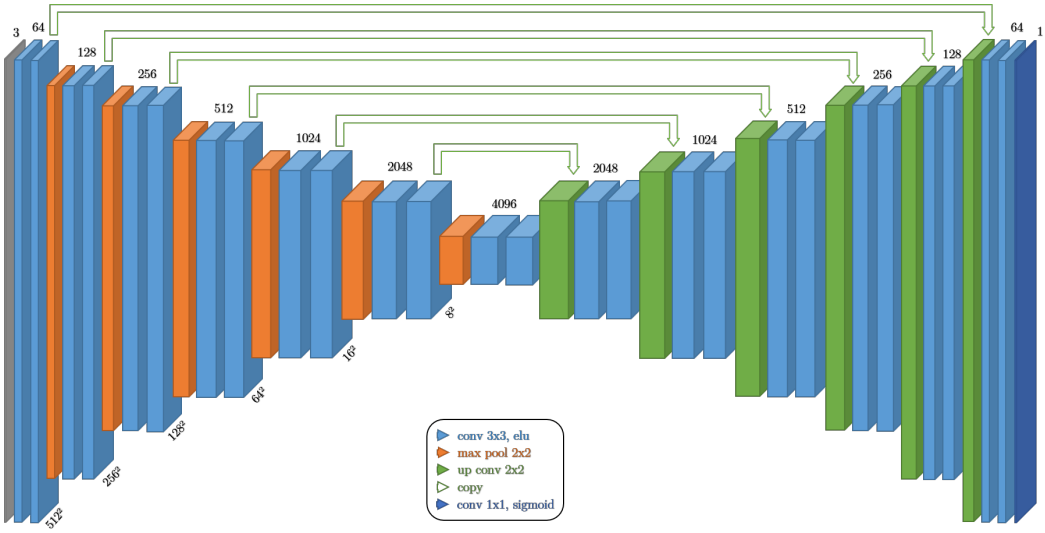


FIGURE 3.11 – Architecture U-Net utilisée pour la détection des défauts, avec trois *patchs* consécutifs de taille 512×512 en entrée, sept couches, et un *patch* des défauts détectés en sortie.

La fonction de perte pour l'entraînement du réseau de neurones est l'inverse de l'approximation linéaire du coefficient de Dice, définie comme suit :

$$\text{Perte}\,(y_C, y_U) = -\frac{2\sum_{i,j} y_C(i,j)y_U(i,j)}{\sum_{i,j} y_C(i,j) + y_U(i,j)} \approx -\frac{2 \times VP}{(VP+FP)+(VP+FN)} \in [-1,0] \tag{3.1}$$

où $y_C(i,j) \in \{0,1\}$ et $y_U(i,j) \in [0,1]$ désignent, respectivement, la valeur du pixel (i,j) du *patch* de masque créé en amont à partir de la restauration de l'expert, et la valeur du pixel (i,j) du *patch* de masque en sortie du réseau U-Net, et où VP, FP et FN représentent, respectivement, le nombre de pixels comptés comme vrais positifs, faux positifs et faux négatifs. Le réseau a été entraîné avec l'optimiseur Adam et un taux d'apprentissage de 5.10^{-5}. Les résultats, en termes de fonction de perte, après avoir prédit toutes les images possibles sur l'ensemble des données de la TABLE 3.3, montrent que, plus les scènes sont complexes, plus il est difficile pour le réseau de détecter précisément les défauts.

En effet, les scènes ne comportant que du texte blanc sur fond noir obtiennent de meilleurs scores que les scènes comportant un plan fixe, qui elles-mêmes obtiennent de meilleurs scores que les scènes où la caméra est en mouvement. Les scores peuvent ne pas sembler pleinement satisfaisants. L'une des raisons est que l'expertise du restaurateur ne constitue pas à proprement parler une vérité terrain, à cause de la sur-détection découlant des grandes sélections manuelles effectuées par le restaurateur.

TABLE 3.3 – Fonction de perte associée aux différents jeux de données et aux différents types de scènes après apprentissage.

Jeu de données \ **Type de scène**	**Texte**	**Caméra fixe**	**Caméra mobile**	**Total**
Ensemble d'entraînement	-0,6252	-0,4564	-0,3629	-0,4323
Ensemble de validation	-0,8422	-0,2899	-0,2018	-0,2737
Ensemble de test	-0,8487	-0,5468	-0,2479	-0,4184

Comme le montre la matrice de confusion de la TABLE 3.4, le pourcentage de VP / FP / FN est vraiment faible par rapport au nombre total des pixels des images. Il est assez difficile d'arriver à une conclusion sur ce qui est vraiment bien détecté ou non, selon la qualité des masques que nous utilisons pour l'entraînement. Mais les bonnes estimations représentent tout de même 99,58% de l'ensemble des pixels.

TABLE 3.4 – Matrice de confusion entre le masque d'entrée, créé à partir de la restauration de l'expert, et le masque en sortie du réseau de neurones.

Entrée \ **Sortie**	**0**	**1**
0	99,44%	0,10%
1	0,32%	0,14%

TABLE 3.5 – Fonction de perte sur l'ensemble de test en utilisant de un à cinq *patchs* consécutifs pour l'entraînement du réseau.

Nombre de *patchs*	1	3	5
Fonction de perte	-0.,3521	**-0,4184**	-0,3743

Nous avons également étudié différents ensembles d'entrée (différents nombres de *patchs*) pour établir la quantité d'information temporelle nécessaire à une détection correcte. Comme prévu, il est plus efficace de considérer trois *patchs* plutôt qu'un seul, car la plupart des défauts n'apparaissent que sur une seule image. Plus surprenant, l'utilisation de cinq *patchs* consécutifs est moins efficace que trois, tout en gardant à l'esprit les limites concernant les détections en amont utilisées pour l'entraînement.

3.2.3 Résultats avec le jeu de données - Scènes de texte

L'utilisation de formes géométriques pour sélectionner les différentes zones à restaurer implique que toute la sélection englobant les défauts est remplacée par une copie des images voisines, même pour les pixels considérés comme non défectueux. Malgré le seuillage et la fermeture morphologique, nous avions indiqué que les lettres des textes restaient toujours présentes dans le masque. En conséquence, cela conduit à de nombreux pixels considérés comme des faux négatifs sur la FIGURE 3.12. Le même problème se pose pour les autres scènes avec les grandes sélections manuelles qui ne sont pas complètement détectées comme défectueuses par le réseau.

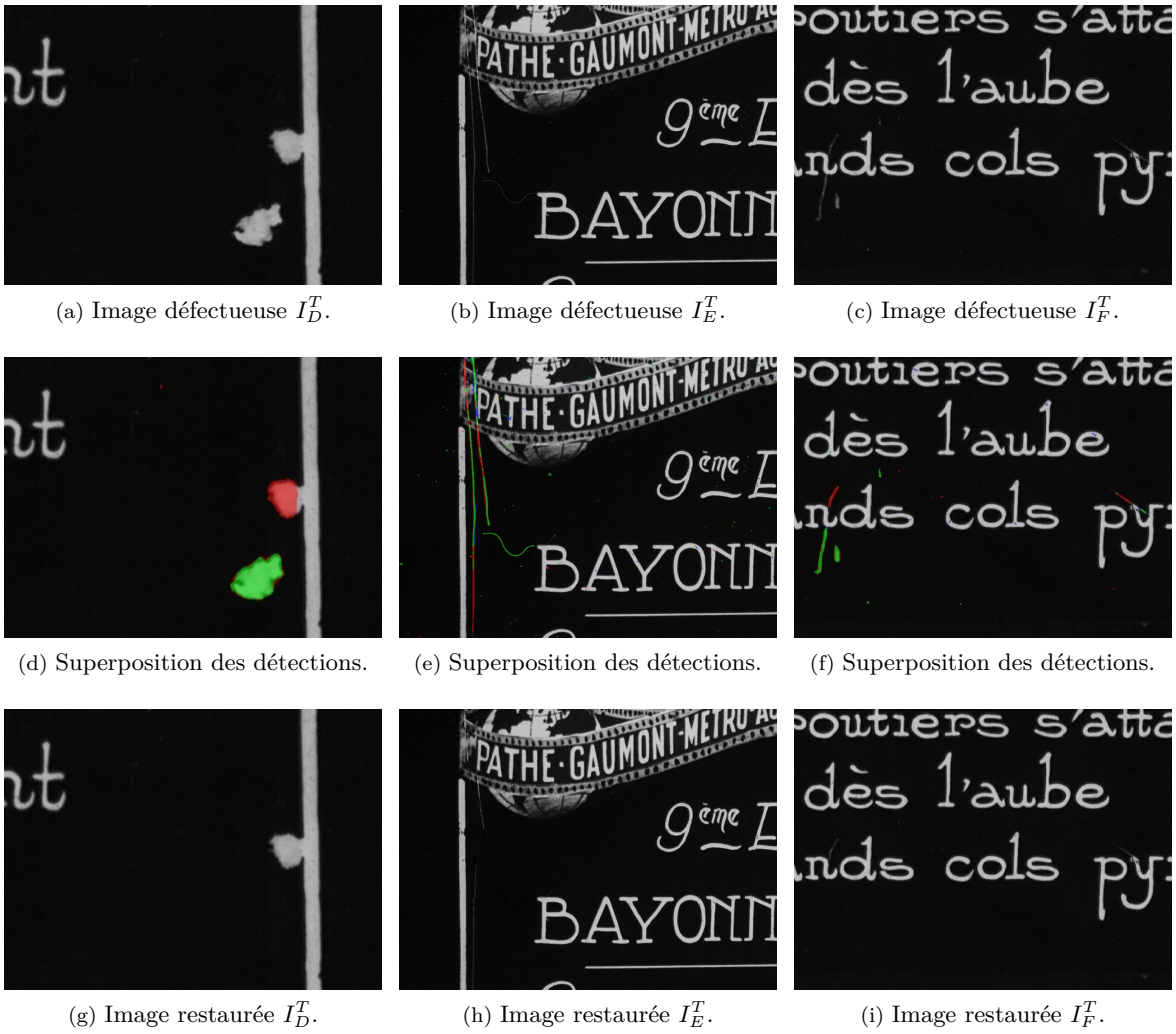
(a) Image défectueuse I_A^T. (b) Image défectueuse I_B^T. (c) Image défectueuse I_C^T.

(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_A^T. (h) Image restaurée I_B^T. (i) Image restaurée I_C^T.

FIGURE 3.12 – Comparaison des différents masques dans des scènes de texte. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. La sur-restauration entraîne l'apparition de larges parties bleues pour des zones considérées comme saines en comparant les images avant et après restauration.

D'autre part, certains défauts n'ont malheureusement pas été détectés par l'expert, en raison du grand nombre d'images à traiter et du temps limité qu'il peut y consacrer. Dans ce cas, le réseau parvient à améliorer la détection manuelle de la restauration en trouvant de nouveaux défauts qui n'avaient pas été détectés auparavant (voir FIGURE 3.13). Cela remet également en perspective les résultats quantitatifs vis-à-vis des faux positifs pour la fonction de perte.



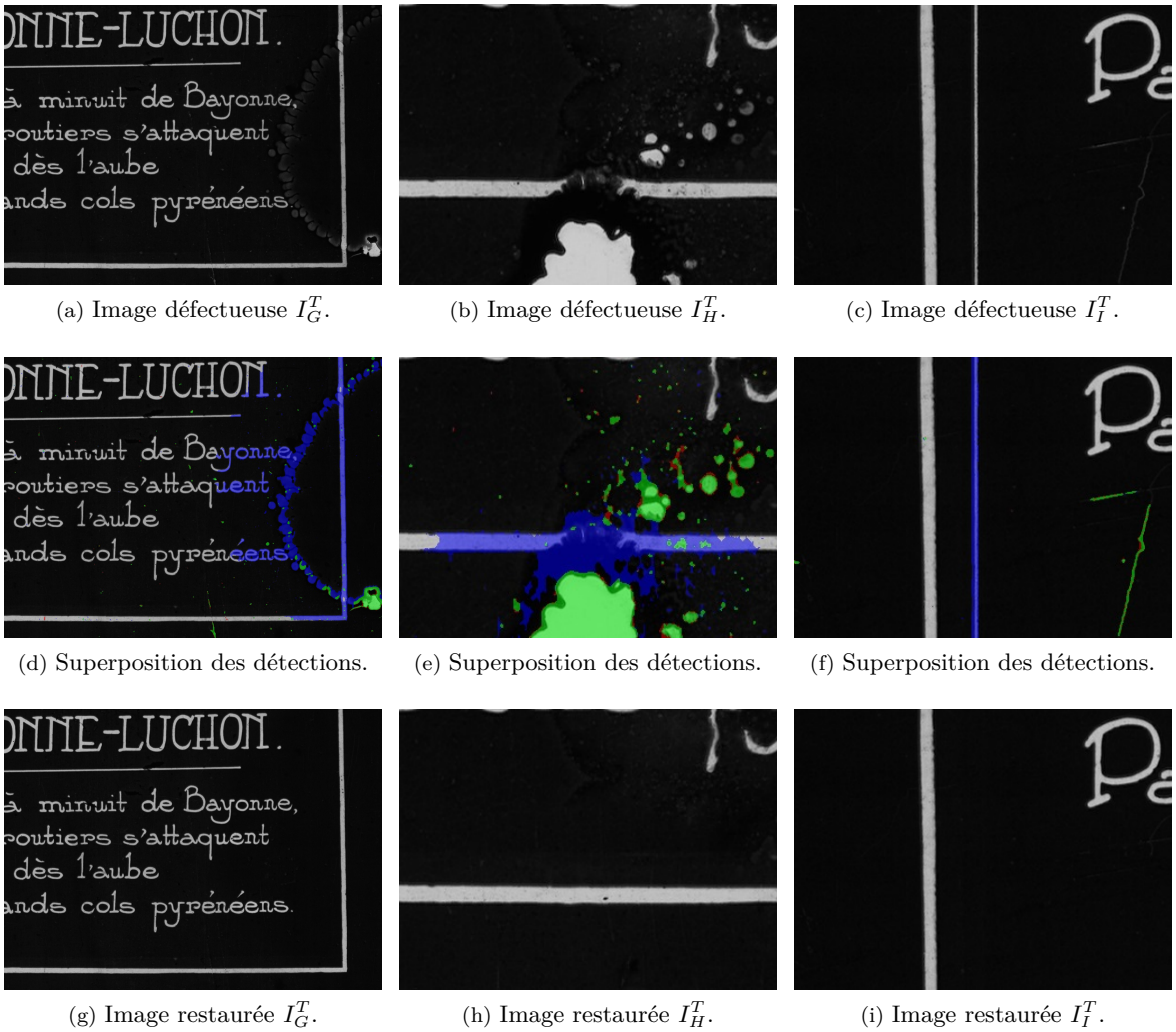
(a) Image défectueuse I_D^T. (b) Image défectueuse I_E^T. (c) Image défectueuse I_F^T.

(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_D^T. (h) Image restaurée I_E^T. (i) Image restaurée I_F^T.

FIGURE 3.13 – Comparaison des différents masques dans des scènes de texte. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Des défauts de différentes formes qui n'avaient pas été restaurés sont tout de même détectés par notre réseau de neurones.

Malgré ces résultats visuellement bons, qui ne se reflètent pas dans la fonction de perte, notre méthode comporte encore certaines limites. Par exemple, même s'il y en a vraiment peu dans le jeu de données, la détection des rayures verticales reste limitée par le fait de n'utiliser que trois images consécutives (ou *patchs* pour être précis) afin de détecter une incohérence temporelle dans le niveau de gris des pixels. En effet, dans l'exemple de la FIGURE 3.14, la rayure est présente au même endroit dans les images successives, de sorte que le réseau proposé n'arrive pas à détecter d'incohérence temporelle. D'autres défauts, plus sombres, ne sont pas détectés non plus. Ce biais est sans doute dû à l'apprentissage des scènes de texte qui comportent essentiellement des défauts blancs ou clairs.



(a) Image défectueuse I_G^T. (b) Image défectueuse I_H^T. (c) Image défectueuse I_I^T.

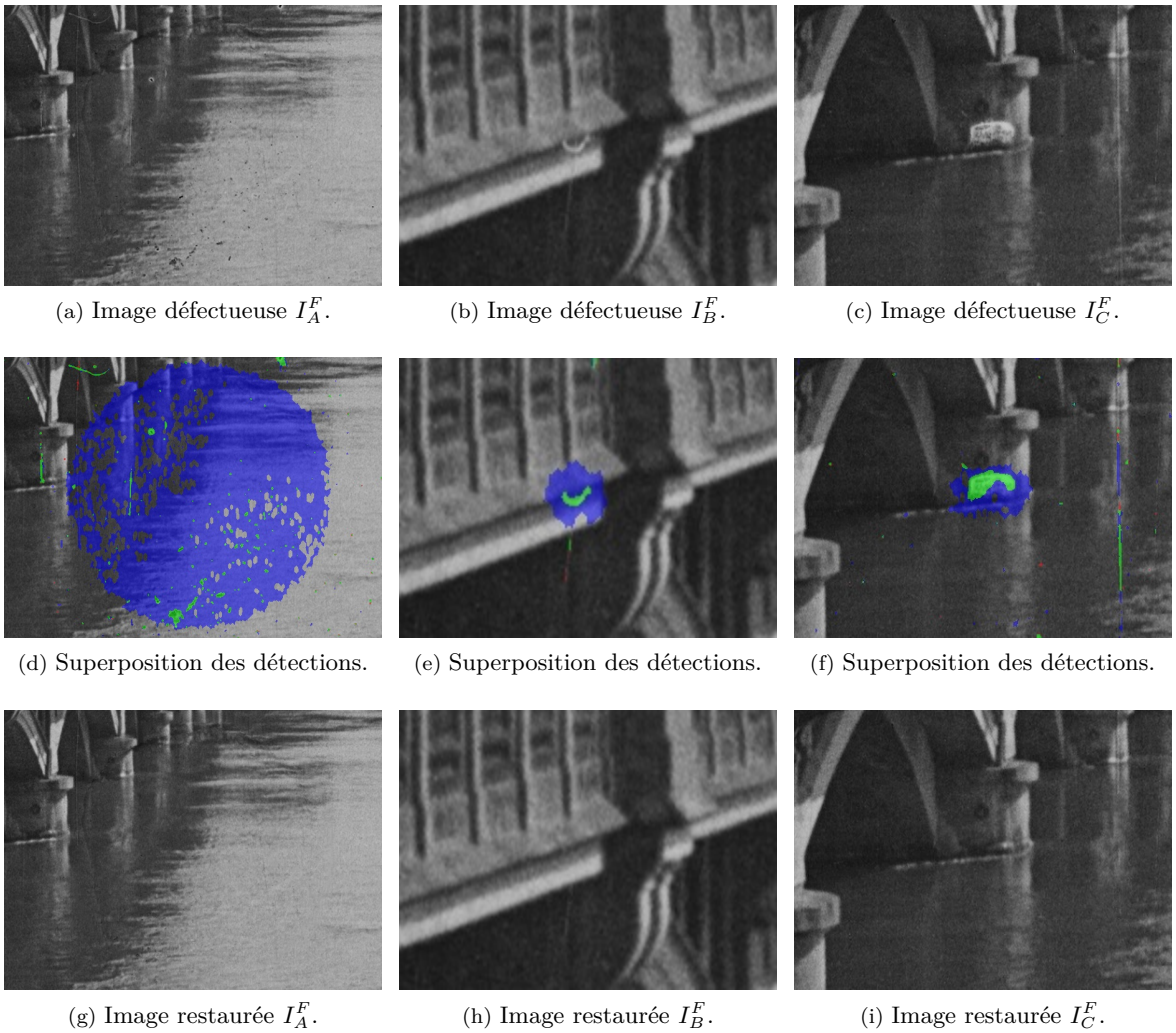
(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_G^T. (h) Image restaurée I_H^T. (i) Image restaurée I_I^T.

FIGURE 3.14 – Comparaison des différents masques dans des scènes de texte. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Les défauts plus sombres ne sont pas détectés, de même que certaines rayures qui sont cohérentes temporellement.

3.2.4 Résultats avec le jeu de données - Scènes fixes

On retrouve sur la FIGURE 3.15 les mêmes formes englobantes typiques de la sur-restauration dans le cas des images naturelles. Notre réseau ne détecte que la forme du défaut en lieu et place du cercle de sélection positionné par le restaurateur. Le cas particulier de la présence d'une tache due à une goutte de liquide qui laisse transparaître les bonnes données avec une déformation, est cependant beaucoup plus compliqué à repérer pour le réseau.



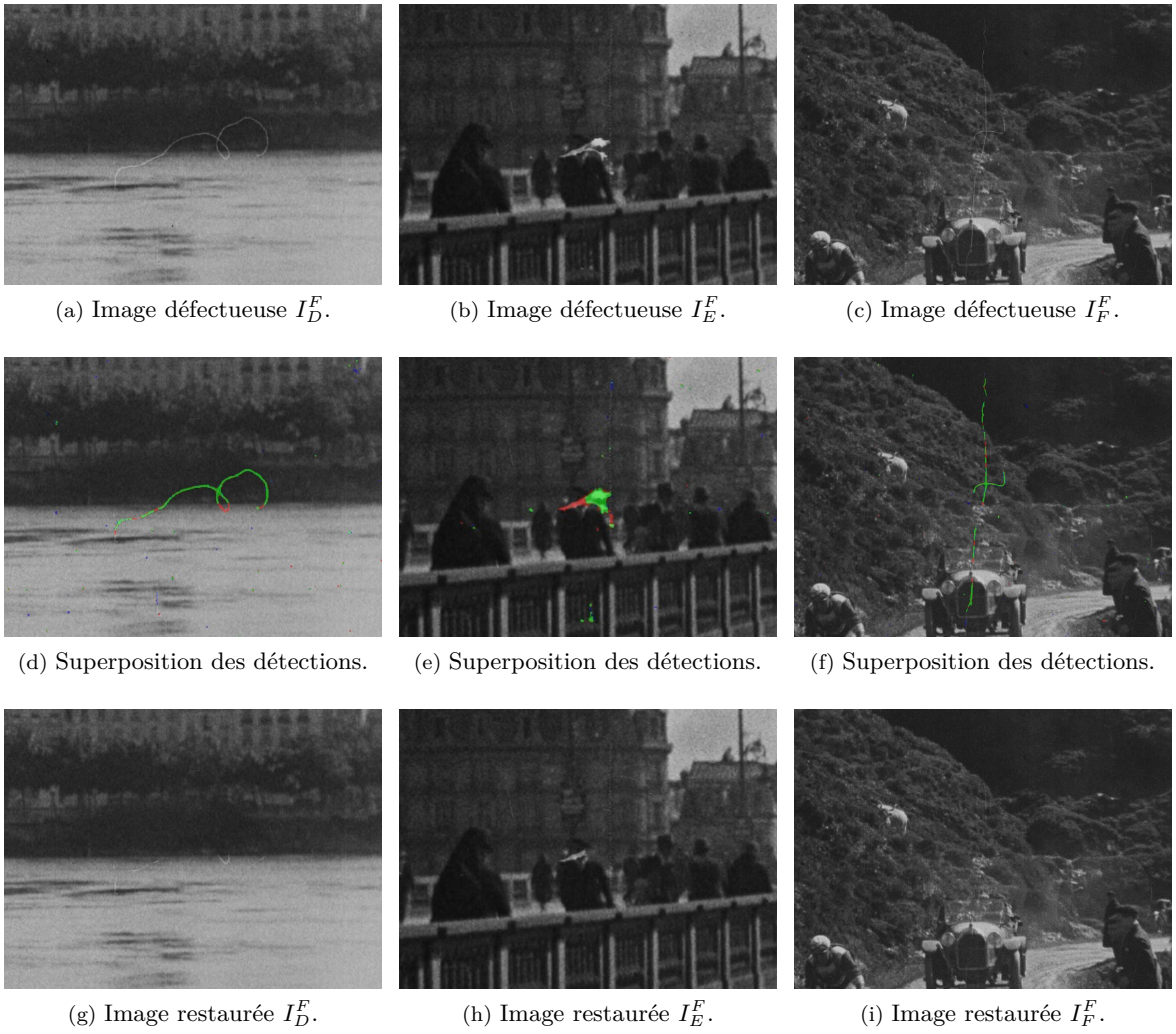
(a) Image défectueuse I_A^F. (b) Image défectueuse I_B^F. (c) Image défectueuse I_C^F.

(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_A^F. (h) Image restaurée I_B^F. (i) Image restaurée I_C^F.

FIGURE 3.15 – Comparaison des différents masques dans des scènes fixes. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Les formes réelles des défauts sont retrouvées par le réseau à l'intérieur de la sélection englobante effectuée par le restaurateur.

Là où les défauts non détectés dans les scènes de texte étaient dus à un oubli de sélection manuelle par le restaurateur, dans le cas des scènes fixes, les défauts non détectés sont globalement des parties de défauts qui ont été détectées de manière automatique par le logiciel, mais pas entièrement. Les formes des défauts mal détectés par le logiciel sont variables, de la rayure à la tache ou à la poussière. Notre réseau arrive à détecter la partie manquante dans tous les cas (voir FIGURE 3.16).



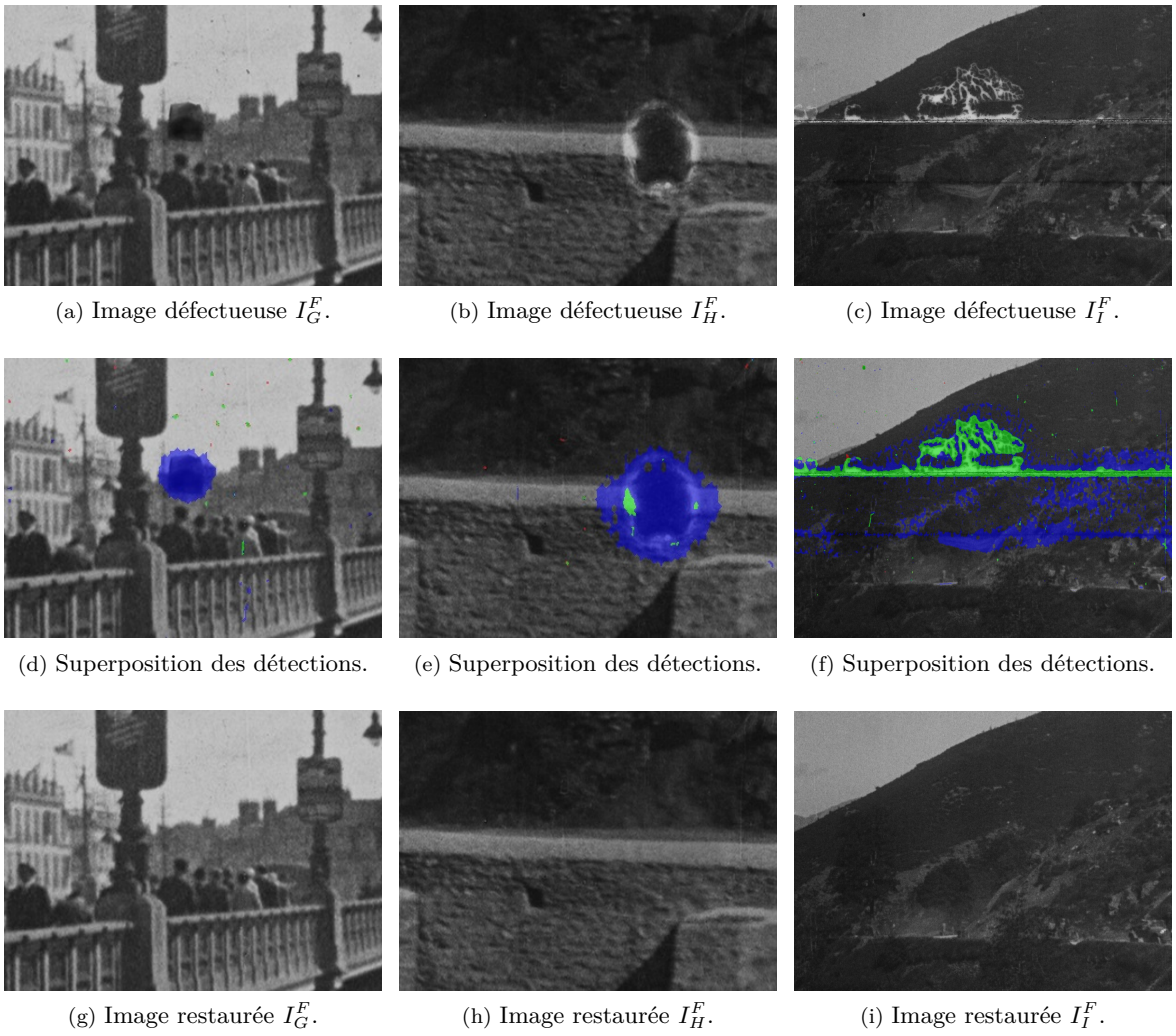
(a) Image défectueuse I_D^F. (b) Image défectueuse I_E^F. (c) Image défectueuse I_F^F.

(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_D^F. (h) Image restaurée I_E^F. (i) Image restaurée I_F^F.

FIGURE 3.16 – Comparaison des différents masques dans des scènes fixes. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. La détection des défauts est améliorée par notre réseau, qui récupère les parties manquantes.

Comme dans le cas de scènes de texte, les défauts sombres ne sont quant à eux pas détectés par le réseau. En particulier, dans le troisième exemple de la FIGURE 3.17, on voit que la partie claire des défauts est bien détectée (ligne horizontale supérieure), alors que la partie foncée ne l'est pas (ligne horizontale inférieure). Une fois encore, le biais semble venir de l'apprentissage des scènes de texte, qui oriente la détection vers les défauts les plus clairs.



(a) Image défectueuse I_G^F. (b) Image défectueuse I_H^F. (c) Image défectueuse I_I^F.

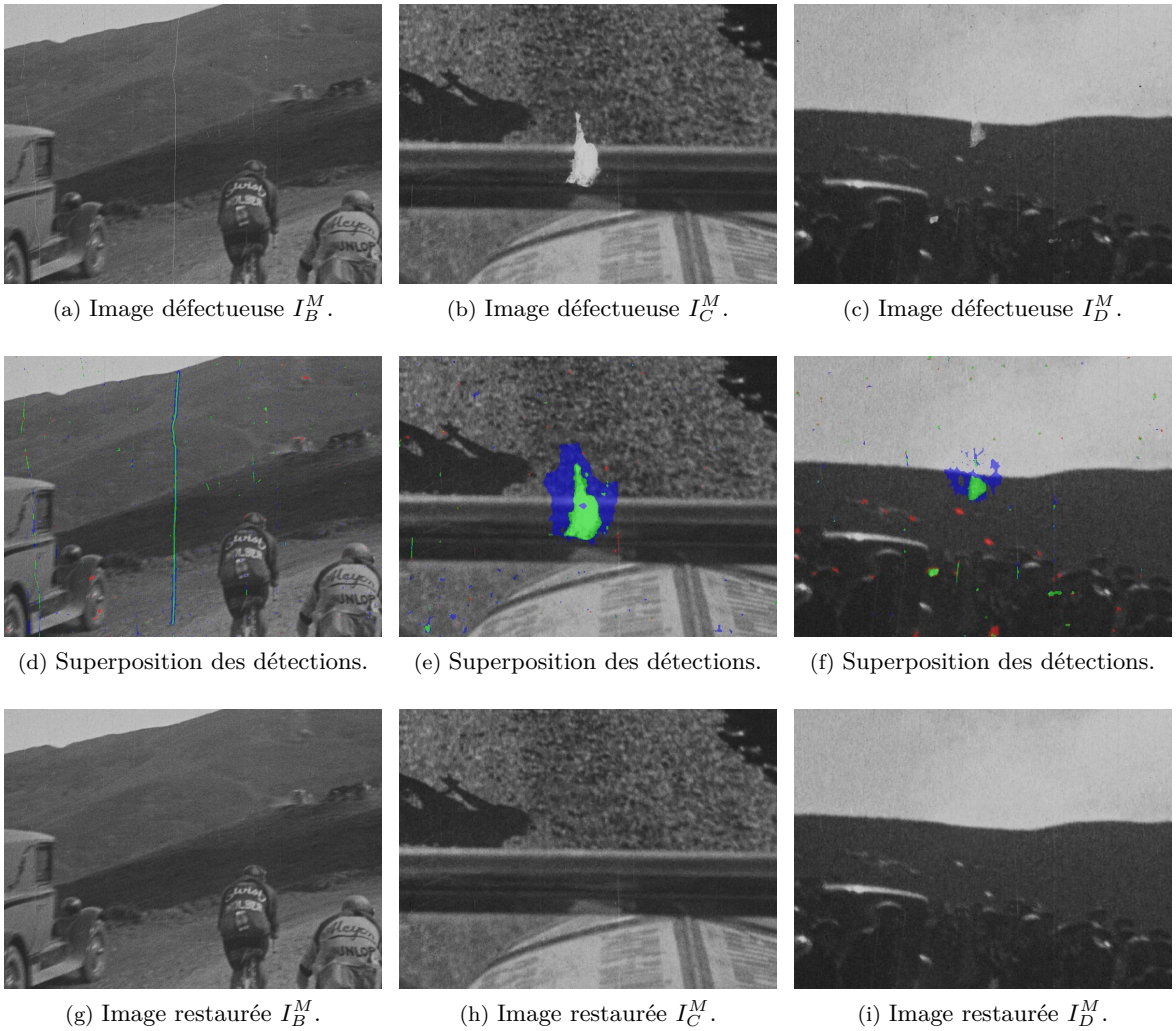
(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_G^F. (h) Image restaurée I_H^F. (i) Image restaurée I_I^F.

FIGURE 3.17 – Comparaison des différents masques dans des scènes fixes. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Les défauts les plus sombres ne sont pas détectés.

3.2.5 Résultats avec le jeu de données - Scènes en mouvement

Si le cas des scènes comportant un mouvement de la caméra est plus compliqué à appréhender par le réseau, qui n'a pas connaissance de la compensation de mouvement pour différencier les incohérences temporelles, ce dernier parvient tout de même à améliorer la précision des détections, dans les cas de sélections effectuées manuellement par le restaurateur (voir FIGURE 3.18), pour certaines rayures ou certaines taches.



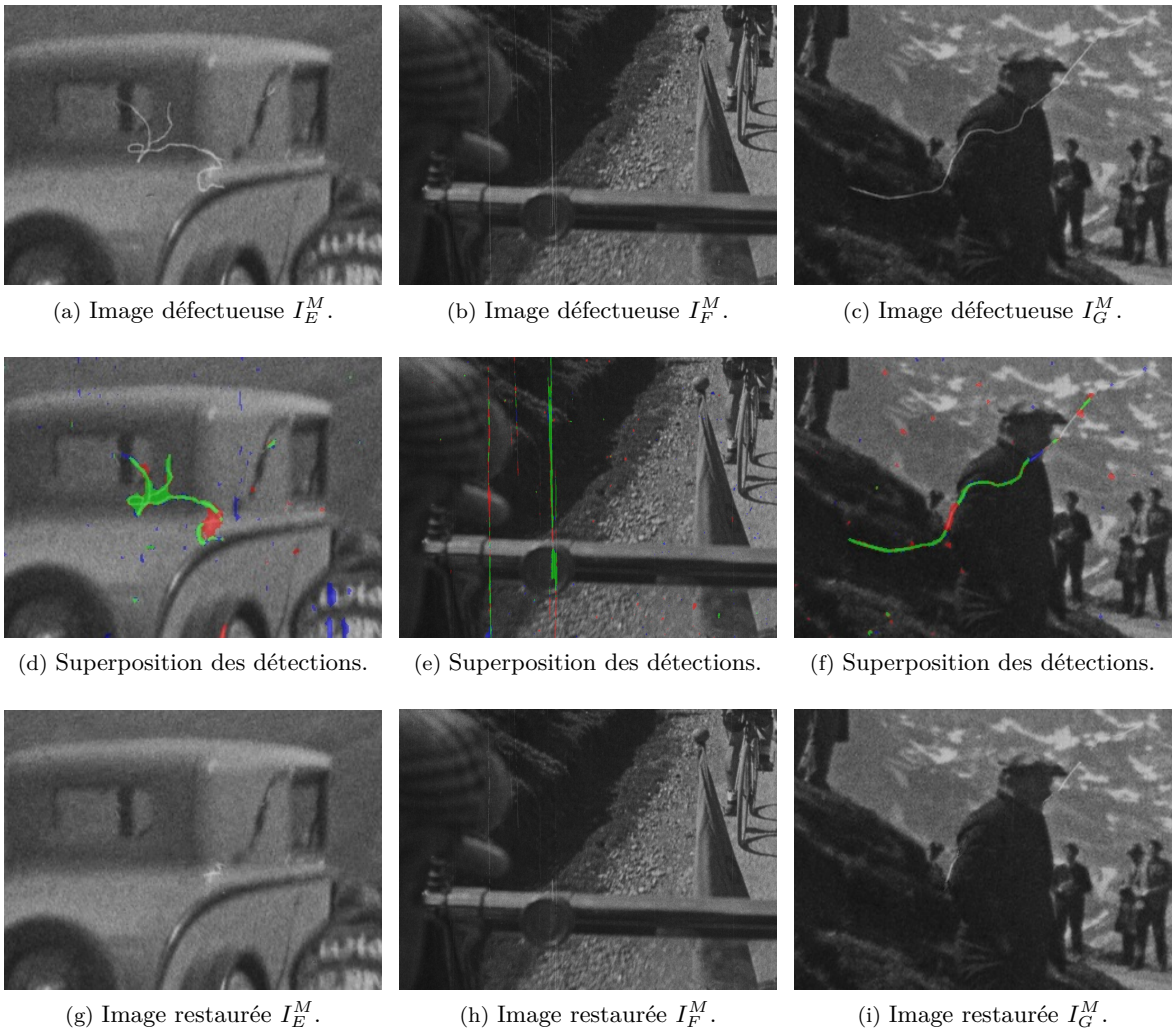
(a) Image défectueuse I_B^M. (b) Image défectueuse I_C^M. (c) Image défectueuse I_D^M.

(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_B^M. (h) Image restaurée I_C^M. (i) Image restaurée I_D^M.

FIGURE 3.18 – Comparaison des différents masques dans des scènes en mouvement. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Les formes réelles des défauts sont encore retrouvées par le réseau à l'intérieur de la sélection englobante effectuée par le restaurateur.

Comme dans le cas des scènes fixes, certains défauts détectés de manière automatique par apprentissage profond avec le logiciel DIAMANT-Film ne sont pas complètement détectés. Notre réseau arrive quant à lui à retrouver les parties manquantes, quelle que soit la forme du défaut (voir FIGURE 3.19).



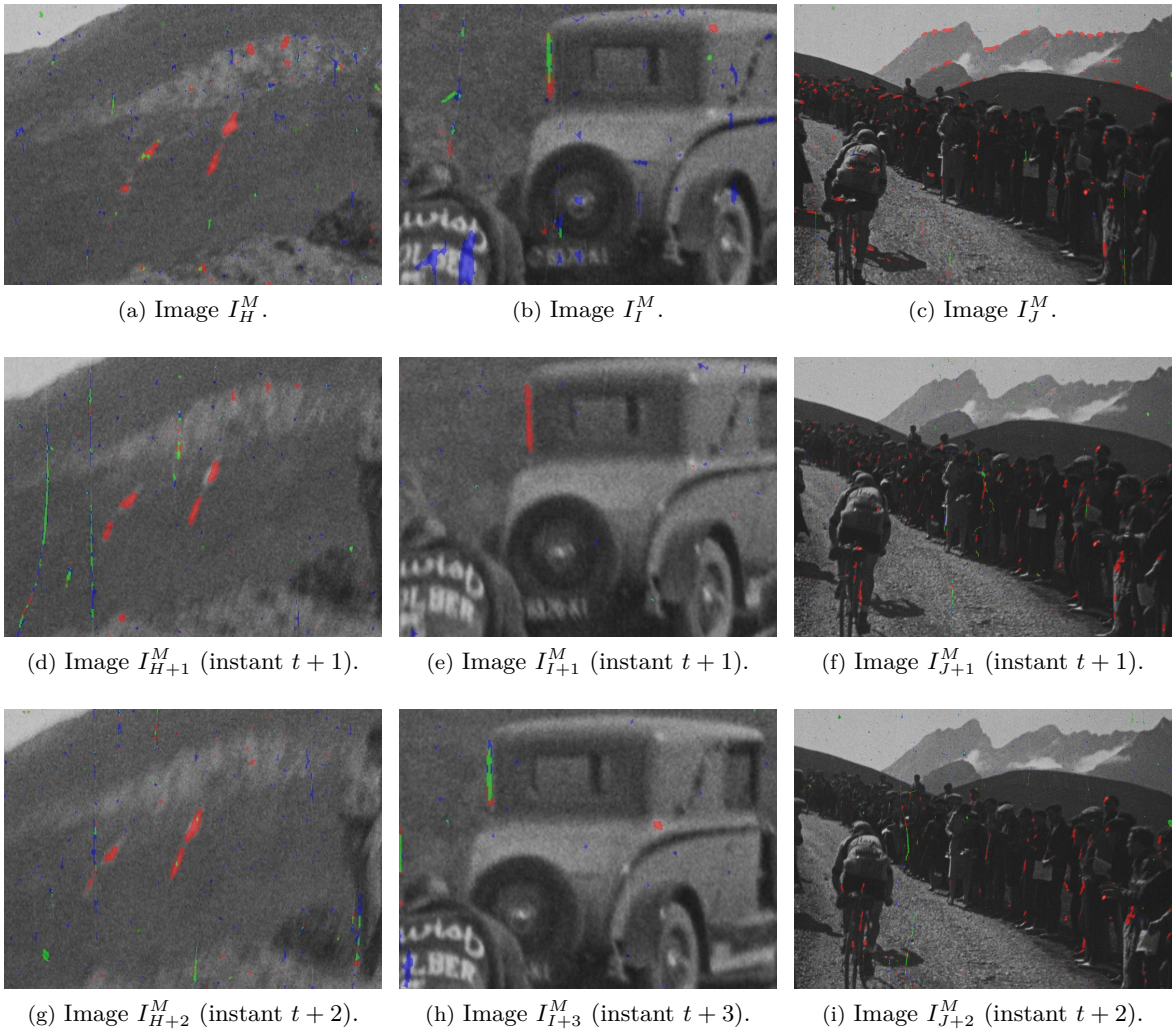
(a) Image défectueuse I_E^M. (b) Image défectueuse I_F^M. (c) Image défectueuse I_G^M.

(d) Superposition des détections. (e) Superposition des détections. (f) Superposition des détections.

(g) Image restaurée I_E^M. (h) Image restaurée I_F^M. (i) Image restaurée I_G^M.

FIGURE 3.19 – Comparaison des différents masques dans des scènes en mouvement. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Les défauts sont complètement détectés par notre réseau, ce qui n'est pas le cas de la détection automatique effectuée par le logiciel DIAMANT-Film.

Une autre limitation du réseau concerne la détection des défauts en présence d'un mouvement de caméra important. En effet, dans ce cas, le réseau ne parvient pas à suffisamment compenser le mouvement pour récupérer les pixels correspondants dans les images voisines. Par conséquent, il assimile certaines parties des images à des taches, comme on peut le voir sur la FIGURE 3.20. Cependant, en regardant l'arrière gauche de la voiture, on peut remarquer que même le logiciel DIAMANT-Film détecte des défauts (indiqués en vert) sur certaines images là où il n'y en a pas.



(a) Image I_H^M. (b) Image I_I^M. (c) Image I_J^M.

(d) Image I_{H+1}^M (instant $t+1$). (e) Image I_{I+1}^M (instant $t+1$). (f) Image I_{J+1}^M (instant $t+1$).

(g) Image I_{H+2}^M (instant $t+2$). (h) Image I_{I+3}^M (instant $t+3$). (i) Image I_{J+2}^M (instant $t+2$).

FIGURE 3.20 – Comparaison des différents masques dans des scènes en mouvement. Les vrais positifs sont indiqués en vert, les faux négatifs en bleu et les faux positifs en rouge. Les pixels indiqués en rouge ne devraient pas être détectés comme pixels défectueux dans ces images. En raison d'un mouvement de caméra important, le réseau les détecte à tort, à cause de leur ressemblance avec des taches. D'un point de vue temporel sans information sur le mouvement, il s'agit en effet d'incohérences.

3.2.6 Comparaison avec un détecteur de rayures

Nous avons utilisé une autre séquence provenant de [Newson *et al.*, 2012] (voir FIGURE 3.21), afin de comparer notre détection des défauts à celle qui est décrite dans cet article. Même si les rayures verticales ne sont pas entièrement détectées par notre réseau, alors que cela est le but de la méthode de [Newson *et al.*, 2012], la quasi-totalité des taches semblent quand même avoir été détectées au-dessus du niveau de la mer. En dessous, les variations de niveaux de gris des vagues font moins ressortir les défauts.

(a) Newson et al. [Newson *et al.*, 2012]. (b) Notre détection.

FIGURE 3.21 – Comparaison sur la séquence « Star » entre la détection des rayures de [Newson *et al.*, 2012] en rouge (a), et notre méthode générique de détection des défauts en bleu (b).

3.3 Vers la correction de défauts

En comparant un film comportant des défauts et le même film restauré de manière semi-automatique par un expert de la Cinémathèque de Toulouse, nous avons réussi à mettre en place un apprentissage automatique via un réseau de neurones de type U-net pour effectuer une détection automatique performante des défauts dans les films. Le fait de ne pas disposer d'une vérité terrain fiable correspondant au film sans défaut nous a amené à devoir rechercher au préalable les masques de défauts entre les deux versions du film, ces derniers n'étant pas directement disponibles. Cependant, même avec une version approximative des « vrais » masques de défauts, notre réseau arrive à compenser les erreurs de détection du logiciel DIAMANT-Film, ou les sur-restaurations qui avaient été effectuées. Une prise en compte plus explicite du mouvement de la caméra et un rééquilibrage entre défauts clairs et sombres pourraient sans doute nous permettre d'améliorer encore notre détecteur. Une fois la détection des défauts effectuée, l'étape suivante, dans notre processus de restauration de films, est la correction de ces défauts par *inpainting* vidéo, qui, comme pour notre détecteur, utilise également les images adjacentes dans la reconstruction.

Chapitre 4

Correction des défauts

Notre objectif est de corriger les défauts précédemment détectés dans les films. Pour y parvenir, notre approche consiste à combiner deux modèles d'*inpainting* vidéo. Le premier effectue l'*inpainting* vidéo par diffusion en calculant conjointement le flux optique. Ce modèle est très efficace pour récupérer la structure géométrique de la partie manquante. Le second recopie des mélanges de pixels à partir de *patchs* 2D recherchés en utilisant des cartes de correspondance vers les images temporellement voisines. Ce modèle est, quant à lui, très performant pour restaurer la texture. Grâce à cette approche, notre modèle n'a besoin que des deux images adjacentes pour reconstruire la zone endommagée. Si chaque modèle pris séparément présente des inconvénients en termes de qualité de reconstruction, la combinaison des deux donne de bien meilleurs résultats. Les contributions liées à ce chapitre sont :

— un nouveau modèle combiné utilisant alternativement l'*inpainting* vidéo par diffusion ainsi que l'*inpainting* vidéo par recopie de *patchs* ;
— des explications sur l'utilisation d'une approche multirésolution, en particulier sur les masques des parties défectueuses qui doivent rester binaires et coïncider avec les défauts dans les images ;
— des tests sur des films présentant des défauts synthétiques et des comparaisons avec d'autres approches d'*inpainting* vidéo.

(a) Image comportant un défaut synthétique (en rouge).

(b) Notre restauration.

FIGURE 4.1 – Restauration par notre approche d'une image de la séquence « Flamant rose » dans laquelle un défaut synthétique de type tache a été ajouté.

4.1 Énoncé du problème

Définissons une séquence de trois images couleur successives $\{u_{_b}, u, u_{_f}\}$ comme des fonctions à variation bornée [Giusti et Williams, 1984] de $\Omega \subset \mathbb{R}^2 \to \mathbb{R}^3$, où $u_{_b}$ désigne l'image arrière, u l'image courante (ou centrale) contenant le défaut à corriger et $u_{_f}$ l'image avant. La zone de défaut est définie par $O \subset \Omega$. Le problème est alors le suivant :

$$u^* = \underset{u}{\operatorname{argmin}} \left\{ E_S(v_{_b}, u, v_{_f}) + E_T(\Gamma_{_b}, u, \Gamma_{_f}) \right\} \tag{4.1}$$

où E_S désigne l'énergie de reconstruction de la structure, minimisée canal par canal (ou en utilisant la luminance lors de l'estimation du mouvement), et E_T l'énergie de reconstruction de la texture, minimisée en utilisant directement l'image couleur. Les différentes variables apparaissant dans (4.1) sont explicitées dans les prochains paragraphes.

4.1.1 Énergie de reconstruction de la structure

Le modèle choisi pour E_S se fonde sur les travaux de [Lauze et Nielsen, 2018] et [Burger *et al.*, 2018], en ajoutant un traitement symétrique des flux optiques avant et arrière. Pour plus de lisibilité, nous considérons, dans les équations sur la reconstruction de la structure, que u comporte un seul canal. Ainsi, le modèle pour E_S prend la forme suivante :

$$\begin{aligned} E_S(v_{_b}, u, v_{_f}) = & \int_\Omega |u_{_b}(x + v_{_b}(x)) - u(x)| \, \mathrm{d}x + \lambda \int_\Omega |Jv_{_b}(x)| \, \mathrm{d}x && \text{(flux arrière régularisé)} \\ & + \int_\Omega \left| u_{_f}(x + v_{_f}(x)) - u(x) \right| \mathrm{d}x + \lambda \int_\Omega \left| Jv_{_f}(x) \right| \mathrm{d}x && \text{(flux avant régularisé)} \\ & + \mu \int_\Omega |\nabla u(x)| \, \mathrm{d}x && \text{(régularisation spatiale)} \end{aligned} \tag{4.2}$$

sous la contrainte $u = u^0$ sur $O^c = \Omega \backslash O$, afin de préserver la partie saine de l'image. Les termes faisant intervenir le champ de déplacement $v_{_b} : \Omega \to \mathbb{R}^2$ (respectivement $v_{_f}$) traduisent la contrainte de flux optique régularisée, proposée par [Zach *et al.*, 2007], entre l'image courante u et l'image arrière $u_{_b}$ (respectivement l'image avant $u_{_f}$), mais en utilisant des matrices jacobiennes, comme dans [Chambolle et Pock, 2011], pour conserver un caractère isotrope à la régularisation des déplacements. Chaque intégrale contient une norme euclidienne $|.|$ définie par $|M| = \sqrt{\sum_{i,j} m_{i,j}^2}$, qui représente une valeur absolue, une norme vectorielle (ici pour le gradient ∇, aussi appelée variation totale [Rudin *et al.*, 1992]) ou une norme de Frobenius, selon le cas. Les paramètres λ et μ sont utilisés pour définir le compromis entre attache aux données et termes de régularisation.

4.1.2 Énergie de reconstruction de la texture

La deuxième énergie E_T est une extension à la vidéo du travail de [Arias *et al.*, 2011] utilisant directement les images couleur, et non canal par canal comme pour la reconstruction de la structure. Ici la recherche de *patchs* optimaux n'est plus effectuée dans un voisinage spatial autour du défaut, mais dans un voisinage temporel (dans les images précédente $u_{_b}$ et suivante $u_{_f}$). Alors que, dans [Newson *et al.*, 2014a] et [Le *et al.*, 2017], la recherche de *patchs* 3D n'est pas limitée dans le temps, nous nous restreignons ici aux *patchs* 2D dans les images avant et arrière, de la même manière que pour la structure. Dans l'image courante u, le pixel central x du *patch* $p_u(x)$ pour lequel on recherche le *patch* optimal dans les images voisines $u_{_b}$ et $u_{_f}$ se situe dans une zone $\widetilde{O}$ égale à l'extension de O d'une demi-largeur de *patch*, afin de propager les *patchs* contenant suffisamment de données « saines » :

$$\begin{aligned} E_T(\Gamma_{_b}, u, \Gamma_{_f}) = &\int_{\widetilde{O}} \omega(x)\ \varepsilon\Big[p_{u_{_b}}(\Gamma_{_b}(x)) - p_u(x)\Big]\,\mathrm{d}x \qquad \text{(dissimilarités arrière)} \\ &+ \int_{\widetilde{O}} (1-\omega(x))\ \varepsilon\Big[p_{u_{_f}}(\Gamma_{_f}(x)) - p_u(x)\Big]\,\mathrm{d}x \qquad \text{(dissimilarités avant)} \end{aligned} \tag{4.3}$$

où $\Gamma_{_b}$ et $\Gamma_{_f}$ désignent les cartes de correspondance, respectivement, entre u et $u_{_b}$ et entre u et $u_{_f}$, où $\omega(x) \in [0,1]$ est un poids entre les deux reconstructions possibles de u à partir des images avant ou arrière (voir Section 4.2.4) et où la fonction ε désigne la distance entre *patchs*. En pratique, ε désigne une pondération de la différence au carré entre *patchs* $(p_{u_{_b}}(\Gamma_{_b}(x)) - p_u(x))^2$ par un noyau gaussien g_a d'écart-type a pour donner plus d'importance au centre du *patch* qu'à son bord. La reconstruction obtenue est alors :

$$\begin{aligned} \varepsilon\Big[p_{u_{_b}}(\Gamma_{_b}(x)) - p_u(x)\Big] &= \int_{\Omega_p} g_a(x_p)\ [u_{_b}(\Gamma_{_b}(x) - x_p) - u(x - x_p)]^2\,\mathrm{d}x_p \\ \varepsilon\Big[p_{u_{_f}}(\Gamma_{_f}(x)) - p_u(x)\Big] &= \int_{\Omega_p} g_a(x_p)\ [u_{_f}(\Gamma_{_f}(x) - x_p) - u(x - x_p)]^2\,\mathrm{d}x_p \end{aligned} \tag{4.4}$$

où $x_p \in \Omega_p$ désigne un pixel dans le *patch* $p_u(x)$, relativement à son centre x.

Minimiser la somme de E_S et E_T constitue un problème complexe, sans preuve d'existence ni d'unicité d'une solution, mais nous cherchons ici seulement à obtenir une approximation numérique de la solution en minimisant E_S et E_T de manière alternée, en utilisant le résultat de la minimisation en u de l'un des deux termes pour initialiser la minimisation de l'autre terme.

4.2 Optimisation

Pour corriger des défauts de grande taille ou estimer le mouvement, une approche multirésolution est nécessaire, de la résolution la plus grossière ($L = L_{\max}$) vers la résolution la plus fine ($L = 0$). Dans le contexte de la vidéo, il est d'autant plus important de suivre cette stratégie pour pouvoir estimer correctement le mouvement. Ceci permet en particulier de se positionner dans de bonnes conditions d'initialisation pour une minimisation locale d'un problème non convexe.

4.2.1 Combinaison des deux modèles

L'idée est de suffisamment sous-échantillonner les images pour pouvoir considérer que les mouvements dans la scène sont petits. À un niveau de résolution L proche de $L_{\max}$, en utilisant un filtrage gaussien pour éliminer les hautes fréquences dans l'étape de sous-échantillonnage, la reconstruction de la structure fonctionne bien, alors que la reconstruction de la texture n'est pas efficace. En revanche, à plus haute résolution (L proche de 0), nous voulons mettre l'accent sur la texture. C'est pourquoi l'algorithme peut choisir un niveau maximal de résolution $L_{\max}^{\text{texture}}$ pour démarrer la reconstruction de la texture et un niveau minimal de résolution $L_{\min}^{\text{structure}}$ pour arrêter la reconstruction de la structure, avec $L_{\max}^{\text{texture}} \geq L_{\min}^{\text{structure}} - 1$ pour garantir qu'au moins une des reconstructions est effectuée à chaque résolution. Par conséquent, à chaque niveau de résolution, notre algorithme applique une seule reconstruction ou les deux à la fois.

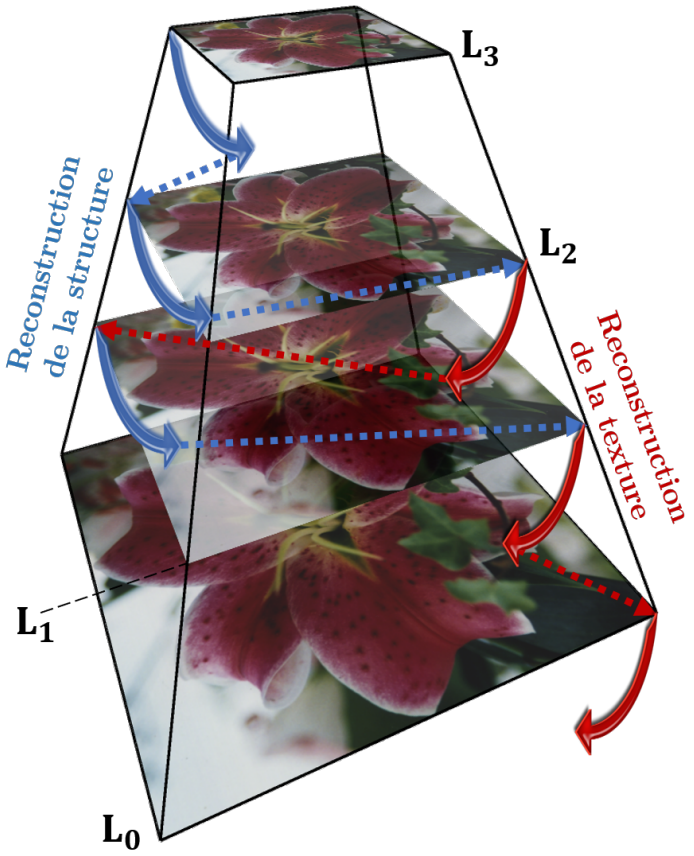


FIGURE 4.2 – Illustration sur quatre niveaux de notre algorithme. À faible résolution, seule la structure de l'image est reconstruite. Aux résolutions intermédiaires, la reconstruction de la structure suivie de la reconstruction de la texture sont appliquées à chaque étage. Aux plus hautes résolutions, seule la texture de l'image est reconstruite.

Algorithme 2 - Reconstruction aux niveaux L et $L-1$

si $L \geq L_{\min}^{\text{structure}}$ **alors**
  $v_b^*, u^*, v_f^* \leftarrow \underset{v_b,u,v_f}{\operatorname{argmin}} \left\{ E_S\left(v_b, u, v_f\right) \right\}$
  $v_b, v_f \leftarrow$ sur-échantillonnage(v_b^*, v_f^*)
  $u \leftarrow u^*$
fin si

si $L \leq L_{\max}^{\text{texture}}$ **alors**
  $\Gamma_b^*, \Gamma_f^* \leftarrow \underset{\Gamma_b,\Gamma_f}{\operatorname{argmin}} \left\{ E_T\left(\Gamma_b, u, \Gamma_f\right) \right\}$
  $\Gamma_b, u, \Gamma_f \leftarrow$ sur-échantillonnage$(\Gamma_b^*, u, \Gamma_f^*)$
  $u^* \leftarrow \underset{u}{\operatorname{argmin}} \left\{ E_T\left(\Gamma_b, u, \Gamma_f\right) \right\}$
  $u \leftarrow u^*$
sinon
  $u \leftarrow$ sur-échantillonnage(u)
fin si

Lors des étapes de sous-échantillonnage, une attention particulière a été portée aux images et aux masques, qui doivent recevoir des traitements différents. Un filtrage gaussien est appliqué aux images, pour éliminer les hautes fréquences avant de descendre en résolution. Quant aux masques, ceux-ci doivent rester binaires à chaque niveau de résolution. Cependant, par un simple sous-échantillonnage en interpolant au plus proche voisin, le bord du masque ne coïncide plus avec le bord du défaut dans l'image. C'est pourquoi nous avons choisi d'appliquer le filtre gaussien au masque, puis d'effectuer à nouveau un seuillage pour récupérer tous les pixels ayant une valeur supérieure à 0, quitte à ce que le masque soit légèrement plus grand que la zone de défauts dans l'image.

Les étapes de sur-échantillonnage sont également différenciées entre les images, les flux optiques et les cartes de correspondance. Pour les images, seule la partie restaurée est sur-échantillonnée avec une interpolation bicubique, et copiée dans la partie défectueuse de l'image à la résolution supérieure. Pour les flux optiques, qui sont denses, l'interpolation bicubique a été été choisie pour passer à la résolution supérieure. Pour les cartes de correspondance, comme deux coordonnées voisines peuvent pointer vers des *patchs* très éloignés spatialement, la seule interpolation possible pour rester proche d'un « bon » *patch* à la résolution supérieure est celle au plus proche voisin.

Les différents problèmes de minimisations permettant de mettre à jour u, v_b, v_f, Γ_b et Γ_f (voir **Algorithme 2**) sont détaillés dans les prochains paragraphes.

4.2.2 Estimation du mouvement

Afin de minimiser E_S par rapport au vecteur de mouvement v_b, nous procédons comme dans [Zach *et al.*, 2007] en linéarisant l'argument des deux différences absolues dans (4.2). Cependant, cette linéarisation n'est possible que si les déplacements sont petits. C'est pourquoi un autre vecteur de mouvement constant v_b^0 (respectivement v_f^0) est introduit, proche de v_b, autour duquel ce dernier est estimé :

$$\begin{aligned} E_S(v_b, u, v_f) = &\int_\Omega \left|\nabla u_b(x + v_b^0(x)) \cdot [v_b - v_b^0](x) + u_b(x + v_b^0(x)) - u(x)\right| \mathrm{d}x + \lambda \int_\Omega |Jv_b(x)| \,\mathrm{d}x \\ &+ \int_\Omega \left|\nabla u_f(x + v_f^0(x)) \cdot [v_f - v_f^0](x) + u_f(x + v_f^0(x)) - u(x)\right| \mathrm{d}x + \lambda \int_\Omega \left|Jv_f(x)\right| \mathrm{d}x \\ &+ \mu \int_\Omega |\nabla u(x)| \,\mathrm{d}x \end{aligned} \tag{4.5}$$

où $\nabla u_b(x + v_b^0(x)) \cdot [v_b - v_b^0](x) + u_b(x + v_b^0(x)) - u(x)$ sera noté $\rho(u, v_b, u_b)$ dorénavant (idem pour v_f). Minimiser E_S relativement aux vecteurs de mouvement v_b et v_f conduit aux expressions suivantes :

$$\begin{aligned} v_b^* &= \underset{v_b}{\operatorname{argmin}} \max_y \left\{ \int_\Omega |\rho(u, v_b, u_b)| \,\mathrm{d}x + \langle\, Jv_b \,|\, y \,\rangle - \iota_{B^\infty}\left(\frac{y}{\lambda}\right) \right\} \\ v_f^* &= \underset{v_f}{\operatorname{argmin}} \max_y \left\{ \int_\Omega \left|\rho(u, v_f, u_f)\right| \mathrm{d}x + \left\langle\, Jv_f \,|\, y \,\right\rangle - \iota_{B^\infty}\left(\frac{y}{\lambda}\right) \right\} \end{aligned} \tag{4.6}$$

où nous avons introduit la variable duale de v_b, $y : \Omega \to \mathbb{R}^{2\times 2}$ et la fonction indicatrice ι_{B^∞}, issues de la transformée de Legendre-Fenchel de la variation totale de v_b, définie ici par :

$$\left(\lambda \int_\Omega |Jv_b(x)| \,\mathrm{d}x\right)^* = \max_y \left\{ \langle\, Jv_b \,|\, y \,\rangle - \iota_{B^\infty}\left(\frac{y}{\lambda}\right) \right\} \tag{4.7}$$

faisant intervenir la fonction indicatrice définie par :

$$\iota_{B^\infty}(y) = \begin{cases} 0 & \text{si } y \in B^\infty \\ +\infty & \text{si } y \notin B^\infty \end{cases} \tag{4.8}$$

où B^∞ est la boule unité associée à la norme infinie. Les problèmes convexes (4.6) peuvent être résolus par l'algorithme primal-dual de [Chambolle et Pock, 2011]. En remarquant que $Jv_b = [\nabla v_{b,1}, \nabla v_{b,2}]^\top$, l'opérateur adjoint de la jacobienne de v_b prend alors la forme $J^*y = -[\mathrm{div}([y_{1,1}, y_{1,2}]^\top), \mathrm{div}([y_{2,1}, y_{2,2}]^\top)]^\top$. Alors que l'opérateur proximal, défini de manière générale dans [Rockafellar, 1976] par :

$$\mathrm{prox}_h(x) = \underset{z}{\operatorname{argmin}} \left\{ h(x) + \frac{1}{2}\|x - z\|^2 \right\} \tag{4.9}$$

associé à y/λ est une projection sur B^∞, l'opérateur proximal associé à v_b est un seuillage doux [Zach *et al.*, 2007] :

$$\mathrm{prox}_{\rho(u,-,u_b)}(v_b) = v_b + \begin{cases} \nabla u_b & \text{si } \rho(u, v_b, u_b) < -|\nabla u_b|^2 \\ -\rho(u, v_b, u_b)\dfrac{\nabla u_b}{|\nabla u_b|} & \text{si } |\rho(u, v_b, u_b)| \le |\nabla u_b|^2 \\ -\nabla u_b & \text{si } \rho(u, v_b, u_b) > |\nabla u_b|^2 \end{cases} \tag{4.10}$$

Les différentes étapes de l'algorithme primal-dual mènent finalement à la forme suivante :

$$\begin{cases} y^{(n+1)} \leftarrow \text{prox}_{\lambda\sigma\iota_{B^\infty}}\left(y^{(n)} + \sigma J\bar{v}_{b}^{(n)}\right) \\ v_{b}^{(n+1)} \leftarrow \text{prox}_{\tau\rho(u,-,u_{b})}\left(v_{b}^{(n)} - \tau J^{*}y^{(n+1)}\right) \\ \bar{v}_{b}^{(n+1)} \leftarrow v_{b}^{(n+1)} + \theta\left(v_{b}^{(n+1)} - v_{b}^{(n)}\right) \end{cases} \tag{4.11}$$

où $\sigma, \tau > 0$ sont des pas de temps. La troisième affectation dans (4.11) sert d'accélération proportionnelle ($\theta \in [0,1]$) à la différence deux dernières itérations de v_{b}. Le même algorithme est utilisé pour v_{f}.

4.2.3 Reconstruction de la structure

Une fois le mouvement estimé, le processus d'*inpainting* est obtenu grâce à la minimisation suivante :

$$u^{*} = \underset{u}{\text{argmin}} \left\{ \int_{\Omega} |u_{b}(x + v_{b}(x)) - u(x)| \, \mathrm{d}x + \int_{\Omega} \left|u_{f}(x + v_{f}(x)) - u(x)\right| \mathrm{d}x + \mu \int_{\Omega} |\nabla u(x)| \, \mathrm{d}x \right\} \tag{4.12}$$

Pour réécrire le problème convexe (4.12) avec des variables duales, les dépendances temporelles de u, v_{b} et v_{f} doivent être clarifiées. En effet, en fixant à 1 le pas de temps entre images successives, et en introduisant la variable de temps t dans u, les fonctions u_{b} et u_{f} prennent la forme suivante :

$$\begin{aligned} u_{b}(x + v_{b}(x)) &= u(x + v_{b}(x,t), t-1) = u\left(\varphi_{b}(x,t)\right) \\ u_{f}(x + v_{f}(x)) &= u(x + v_{f}(x,t), t+1) = u\left(\varphi_{f}(x,t)\right) \end{aligned} \tag{4.13}$$

où φ_{b} et φ_{f} sont deux transformations supposées différentiables et inversibles, similaires aux cartes de correspondance Γ_{b} et Γ_{f}, en faisant l'hypothèse que $\varphi_{b} \circ \varphi_{f} = \varphi_{f} \circ \varphi_{b} = I_{d}$ presque partout. Avec ces nouvelles notations, (4.12) peut être réécrite :

$$u^{*} = \underset{u}{\text{argmin}} \max_{z} \left\{ \left\langle \begin{array}{c} u \circ \varphi_{b} - u \\ u \circ \varphi_{f} - u \\ \nabla u \end{array} \middle| z \right\rangle - \iota_{B^\infty}(z_{1}) - \iota_{B^\infty}(z_{2}) - \iota_{B^\infty}\left(\frac{1}{\mu}[z_{3}, z_{4}]^{\top}\right) \right\} \tag{4.14}$$

où nous introduisons la variable duale de u, $z = [z_{1}, z_{2}, z_{3}, z_{4}]^{\top} : \Omega \to \mathbb{R}^{4}$. En notant K l'opérateur relatif à u dans le produit scalaire, qui prend la forme suivante :

$$Ku = \begin{pmatrix} u \circ \varphi_{b} - u \\ u \circ \varphi_{f} - u \\ \nabla u \end{pmatrix} \tag{4.15}$$

cela conduit à l'opérateur adjoint K^{*} décrit dans [Lauze et Nielsen, 2018] :

$$K^{*}z = \det\left(J\varphi_{f}\right) z_{1} \circ \varphi_{f} - z_{1} + \det\left(J\varphi_{b}\right) z_{2} \circ \varphi_{b} - z_{2} - \text{div}([z_{3}, z_{4}]^{\top}) \tag{4.16}$$

La minimisation (4.14) peut également être effectuée en utilisant l'algorithme primal-dual de [Chambolle et Pock, 2011] :

$$\begin{cases} z_1^{(n+1)} \leftarrow \text{prox}_{\sigma \iota_{B^\infty}}\left(z_1^{(n)} + \sigma\left(u \circ \varphi_b - \bar{u}^{(n)}\right)\right) \\ z_2^{(n+1)} \leftarrow \text{prox}_{\sigma \iota_{B^\infty}}\left(z_2^{(n)} + \sigma\left(u \circ \varphi_f - \bar{u}^{(n)}\right)\right) \\ \begin{bmatrix} z_3^{(n+1)} \\ z_4^{(n+1)} \end{bmatrix} \leftarrow \text{prox}_{\mu\,\sigma' \iota_{B^\infty}}\left(\begin{bmatrix} z_3^{(n)} \\ z_4^{(n)} \end{bmatrix} + \sigma' \nabla \bar{u}^{(n)}\right) \\ u^{(n+1)} \leftarrow u^{(n)} - \tau K^* z^{(n+1)} \\ \bar{u}^{(n+1)} \leftarrow u^{(n+1)} + \theta\left(u^{(n+1)} - u^{(n)}\right) \end{cases} \tag{4.17}$$

où $\sigma, \sigma', \tau > 0$ sont des pas de temps et $\theta \in [0, 1]$. Pour minimiser une énergie similaire à (4.2), [Burger *et al.*, 2018] répète les minimisations alternées pour l'*inpainting* de u et l'estimation d'un unique champ de déplacement v jusqu'à la convergence. A contrario, dans notre cas, les trois variables (v_b, u, v_f) sont minimisées simultanément, en appliquant des étapes proximales successives à chaque variable.

4.2.4 Reconstruction de la texture avec estimation des cartes de correspondance

La minimisation de l'énergie de texture E_T se fonde sur les travaux de [Arias *et al.*, 2011], où les cartes de correspondance Γ_b et Γ_f sont estimées en utilisant l'algorithme PatchMatch de [Barnes *et al.*, 2009] pour opérer une recherche non locale efficace sans être exhaustive, avec une distance L^2 entre *patchs*. Par conséquent, pour Γ_b (respectivement Γ_f), nous tirons de (4.4), $\forall x \in \widetilde{O}$:

$$\Gamma_b(x) = \underset{x_b \in \Omega}{\text{argmin}} \int_{\Omega_p} g_a(x_p)\, \left[u_b(x_b - x_p) - u(x - x_p)\right]^2 \mathrm{d}x_p \tag{4.18}$$

Pour chacune des images voisines u_b et u_f, chacun des termes du membre droit de (4.3) peut être réécrit, sans prendre en compte le poids ω dans un premier temps, comme le cas extrême où l'on choisit seulement le *patch* le plus proche pour chaque pixel de la zone de défaut (voir [Arias *et al.*, 2011] pour plus de détails), en introduisant une fonction de Dirac δ :

$$E_T^b(u, \Gamma_b) = \int_{\widetilde{O}} \int_{\Omega} \delta(\Gamma_b(x) - x_b)\, \varepsilon[p_{u_b}(x_b) - p_u(x)]\, \mathrm{d}x_b\, \mathrm{d}x \tag{4.19}$$

Avec les changements de variables $x := x - x_p$ et $x_b := x_b - x_p$, qui opèrent deux translations de telle sorte que $x \in O$ et $x_b \in O$ désormais, nous tirons de (4.4) et (4.19) :

$$E_T^b(u, \Gamma_b) = \int_O \int_{\Omega} m(x, x_b)\, \left[u_b(x_b) - u(x)\right]^2 \mathrm{d}x_b\, \mathrm{d}x \tag{4.20}$$

où :

$$m(x, x_b) = \int_{\Omega_p} g_a(x_p)\ \delta(\Gamma_{_b}(x + x_p) - (x_b + x_p))\,\mathrm{d}x_p \tag{4.21}$$

dont l'intégrale sur Ω est égale à 1, puisque g_a est, par hypothèse, une gaussienne normalisée. En développant la différence au carré dans (4.20), u est donc aussi le minimiseur de l'énergie suivante, qui est égale à $E_T^b(u, \Gamma_{_b})$ à une constante près :

$$\widetilde{E}_T^b(u, \Gamma_{_b}) = \int_O \left[u(x) - \int_\Omega m(x, x_b)\, u_{_b}(x_b)\,\mathrm{d}x_b\right]^2 \mathrm{d}x \tag{4.22}$$

ce qui nous fournit directement la solution des moyennes non locales définies, $\forall x \in O$, par :

$$u(x) = \int_\Omega m(x, x_b)\, u_{_b}(x_b)\,\mathrm{d}x_b = \int_{\Omega_p} g_a(x_p)\, u_{_b}(\Gamma_{_b}(x + x_p) - x_p))\,\mathrm{d}x_p \tag{4.23}$$

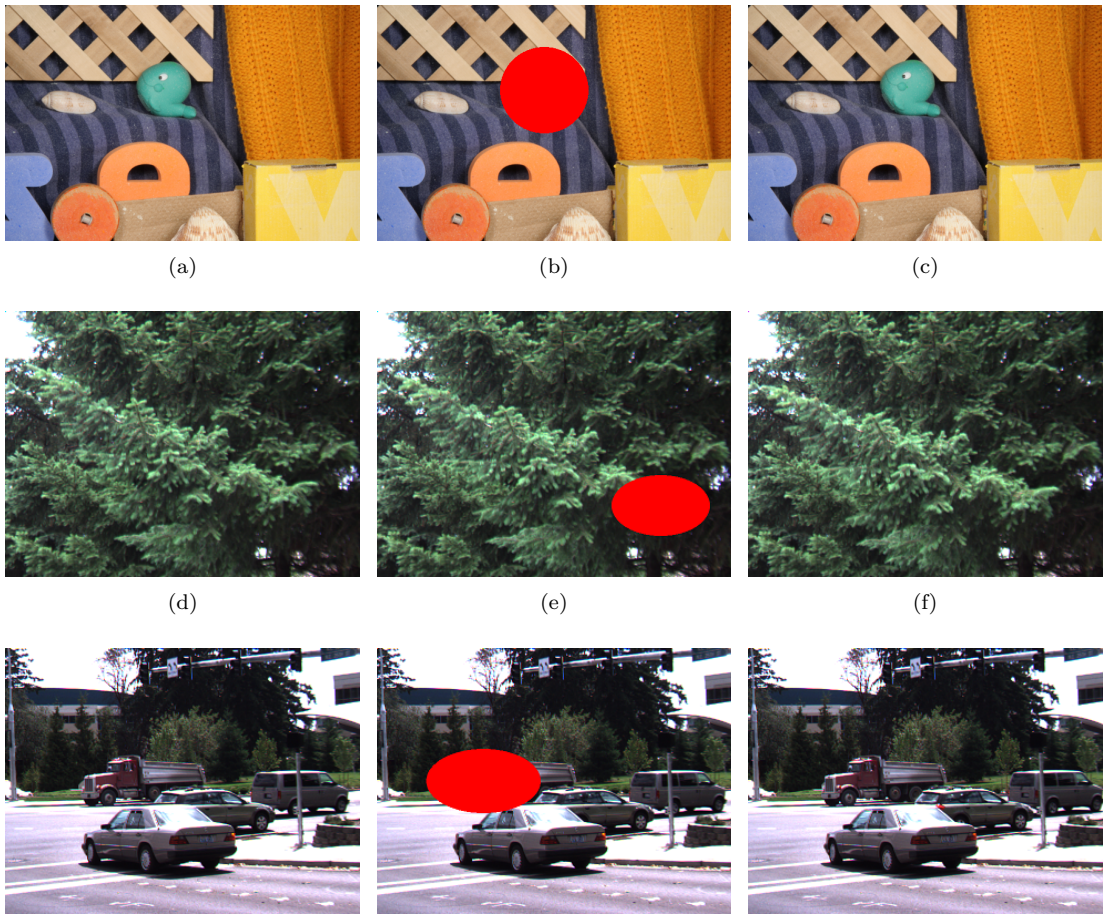
Il s'agit ici du résultat pour une seule des deux images voisines. Pour prendre en compte les deux images, des moyennes pondérées peuvent être effectuées avec les poids $\omega(x)$ et $(1 - \omega(x))$. Moyenner les deux résultats avec $\omega(x) = 0,5$ (variante notée T_H pour *Half*) provoquera l'apparition de flou. À l'inverse, choisir le meilleur des deux résultats avec $\omega(x) \in \{0, 1\}$ (variante notée T_B pour *Best*) fera apparaître des artefacts spatiaux. Dans ce cas, le choix entre 0 et 1 pour les poids est effectué après avoir estimé les deux cartes de correspondance $\Gamma_{_b}$ et $\Gamma_{_f}$. Ensuite, pour chaque pixel à reconstruire, le poids $\omega(x)$ est égal à 1 si la distance estimée $\varepsilon\left[p_{u_{_b}}(\Gamma_{_b}(x)) - p_u(x)\right]$ au meilleur *patch* dans l'image arrière est inférieure à la distance estimée $\varepsilon\left[p_{u_{_f}}(\Gamma_{_f}(x)) - p_u(x)\right]$ au meilleur *patch* dans l'image avant, et $\omega(x)$ est égal à 0 sinon. Une solution intermédiaire consiste à appliquer une moyenne pondérée (variante notée T_W, pour *Weighted mean*) en utilisant le ratio entre la distance estimée au meilleur *patch* dans l'image avant et la somme des deux distances estimées :

$$\omega(x) = \frac{\varepsilon\left[p_{u_{_f}}(\Gamma_{_f}(x)) - p_u(x)\right]^\alpha}{\varepsilon\left[p_{u_{_f}}(\Gamma_{_f}(x)) - p_u(x)\right]^\alpha + \varepsilon\left[p_{u_{_b}}(\Gamma_{_b}(x)) - p_u(x)\right]^\alpha} \tag{4.24}$$

où $\alpha \in [0, +\infty[$ permet de pondérer le choix entre le pixel provenant de l'image avant et celui provenant de l'image arrière. Les cas extrêmes T_H et T_B sont atteints, respectivement, pour $\alpha = 0$ et α tendant vers $+\infty$.

4.3 Expérimentations avec les variantes de notre algorithme

Notre algorithme a été implémenté en choisissant un niveau maximal de la pyramide multirésolution $L_{\max} = 10$, et un facteur entre étages successifs égal à $\sqrt{2}$. L'algorithme est d'abord testé sur trois images consécutives issues de séquences Middlebury, où un défaut artificiel a été introduit dans l'image centrale (voir FIGURE 4.3).



(a) (b) (c)

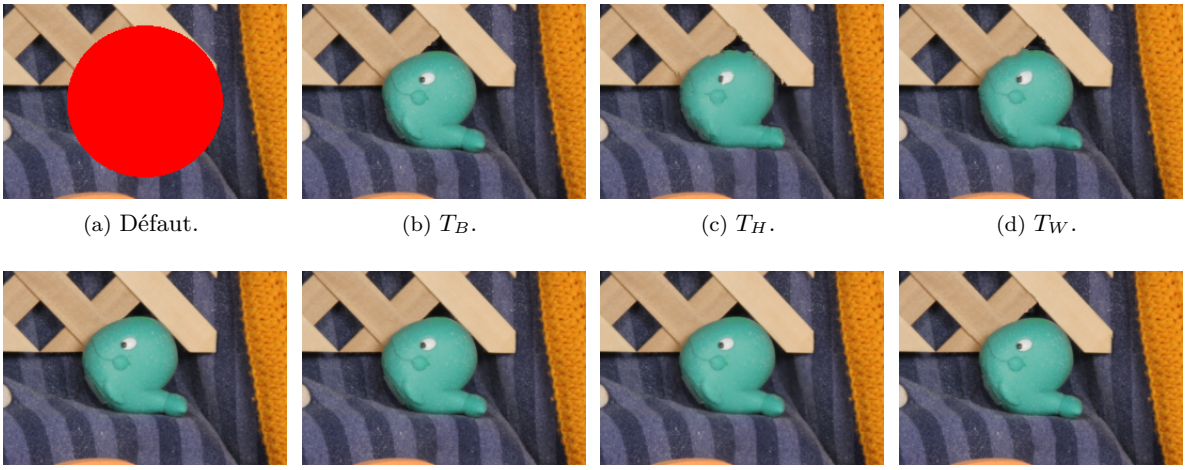
(d) (e) (f)

(g) (h) (i)

FIGURE 4.3 – Images 8 à 10 des séquences « RubberWhale » (en haut), « Evergeen » (au milieu) et « Dumptruck » (en bas) de la base de données Middlebury. Le défaut est surligné en rouge dans l'image centrale (colonne du milieu).

4.3.1 Séquence « RubberWhale »

Pour les images de la séquence « RubberWhale », les résultats de la FIGURE 4.4 montrent que les méthodes de restauration de la texture T_B, T_H et T_W (4.4b, 4.4c, 4.4d) sont inadéquates pour reconstruire correctement la marionnette et le fond en bois, quel que soit le choix du poids ω, car cette image est peu texturée. En revanche, la reconstruction structurelle (4.4e) fournit à elle seule un résultat de bonne qualité. Cependant, notre algorithme permet d'améliorer ce résultat (4.4f, 4.4g, 4.4h), en particulier au sommet de la tête de la marionnette où il y a une fausse couleur due à la séparation des canaux de couleur avec la seule reconstruction de la structure, et au coin de la barre en bois à l'arrière-plan qui est moins arrondie.



(a) Défaut. (b) T_B. (c) T_H. (d) T_W.

(e) Structure. (f) Structure + T_B. (g) Structure + T_H. (h) Structure + T_W.

FIGURE 4.4 – Image centrale de la séquence « RubberWhale » : zooms sur les reconstructions de la zone défectueuse.

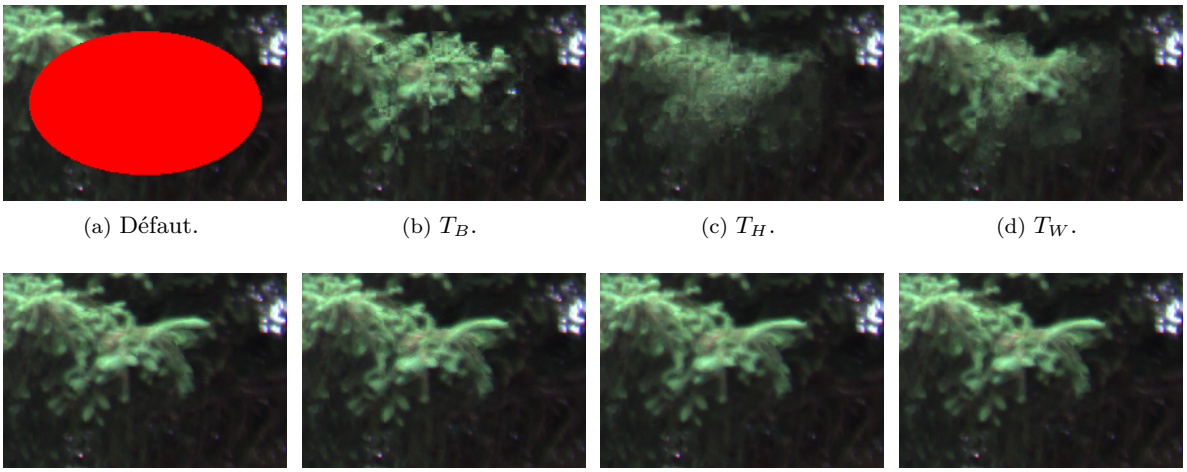
Du point de vue quantitatif, avec les métriques que sont le PSNR (*Peak Signal to Noise Ratio*) et la SSIM (*Structural Similarity*) dans la TABLE 4.1, les différentes méthodes qui reconstruisent la structure sont toutes équivalentes, avec une léger mieux pour la méthode qui combine structure et texture en moyennant les pixels des images avant et arrière.

	Reconstruction de la texture			
	Aucune	T_B	T_H	T_W
Sans reconstruction de la structure	16,34 - 0,915	38,46 - 0,994	41,12 - 0,996	39,92 - 0,995
Avec reconstruction de la structure	**48,49 - 0,999**	46,99 - 0,999	**48,81 - 0,999**	**48,59 - 0,999**

TABLE 4.1 – Mesures de PSNR - SSIM pour les différentes reconstructions de l'image centrale de la séquence « RubberWhale ». L'ordre des scores est le même que celui des images de la FIGURE 4.4 : (4.4a, 4.4b, 4.4c, 4.4d) sur la première ligne et (4.4e, 4.4f, 4.4g, 4.4h) sur la seconde ligne.

4.3.2 Séquence « Evergreen »

Pour le test sur la séquence « Evergreen », les résultats de la FIGURE 4.5 montrent que les méthodes où l'on ne reconstruit que la texture (4.5b, 4.5c, 4.5d) ne parviennent pas à restaurer correctement les branches de l'arbre. L'approche avec le meilleur des deux *patchs* (4.5b) provoque des artefacts, alors que celles utilisant un moyennage (4.5c et 4.5d) conduisent à un résultat où la texture est perdue au profit d'une homogénéité non désirée. La reconstruction de la structure (4.5e) est meilleure, mais encore un peu floue à cause de la diffusion, et certaines branches semblent un peu trop étirées. Enfin, la combinaison des deux reconstructions (4.5f, 4.5g, 4.5h) réussit à superposer une texture à la structure.



(a) Défaut. (b) T_B. (c) T_H. (d) T_W.

(e) Structure. (f) Structure + T_B. (g) Structure + T_H. (h) Structure + T_W.

FIGURE 4.5 – Image centrale de la séquence « Evergreen » : zooms sur les reconstructions de la zone défectueuse.

Ces interprétations visuelles sont confirmées par les mesures de qualité reportées dans la TABLE 4.2. La reconstruction de la structure est plus performante que celle de la texture, tandis que notre algorithme, qui combine les deux, donne les meilleurs résultats, avec un écart non négligeable. Par ailleurs, le choix de l'une quelconque des trois reconstructions de la texture, associée à la reconstruction de la structure, n'a pas de réel impact sur le résultat.

	Reconstruction de la texture			
	Aucune	T_B	T_H	T_W
Sans reconstruction de la structure	18,66 - 0,943	30,94 - 0,978	33,37 - 0,980	33,56 - 0,982
Avec reconstruction de la structure	37,24 - 0.991	**40,37 - 0,994**	**40,47 - 0,994**	**40,50 - 0,994**

TABLE 4.2 – Mesures de PSNR - SSIM pour les différentes reconstructions de l'image centrale de la séquence « Evergreen ». L'ordre des scores est le même que celui des images de la FIGURE 4.5 : (4.5a, 4.5b, 4.5c, 4.5d) sur la première ligne et (4.5e, 4.5f, 4.5g, 4.5h) sur la seconde ligne.

4.3.3 Séquence « Dumptruck »

Pour le test sur la séquence « Dumptruck », les résultats de la FIGURE 4.6 montrent que la reconstruction de la structure (4.6e) entraîne une déformation du camion, qui semble osciller verticalement dans la vidéo. De plus, une trace apparaît derrière la voiture située à droite, au niveau des roues du camion. Les reconstructions de la texture à l'aide de moyennes (4.6c et 4.6d) conduisent à une reconstruction floue du camion. Même en utilisant le meilleur *patch* des deux images adjacentes (4.6b), dont le résultat semble être bon sur cette seule image, la vidéo montre que l'algorithme choisit l'image la plus proche, en terme de distance entre *patchs*, et reste « verrouillé ». C'est pourquoi il semble y avoir un manque de mouvement dans la vidéo, à l'intérieur de la zone de défaut. Notre algorithme (4.6f, 4.6g, 4.6h) fournit de meilleurs résultats (il reste quand même une trace à l'arrière de la voiture de droite), grâce à l'apport de la texture, qui permet de compenser les erreurs dans le flux optique.

(a) Défaut. (b) T_B. (c) T_H. (d) T_W.

(e) Structure. (f) Structure + T_B. (g) Structure + T_H. (h) Structure + T_W.

FIGURE 4.6 – Image centrale de la séquence « Dumptruck » : zooms sur les reconstructions de la zone défectueuse.

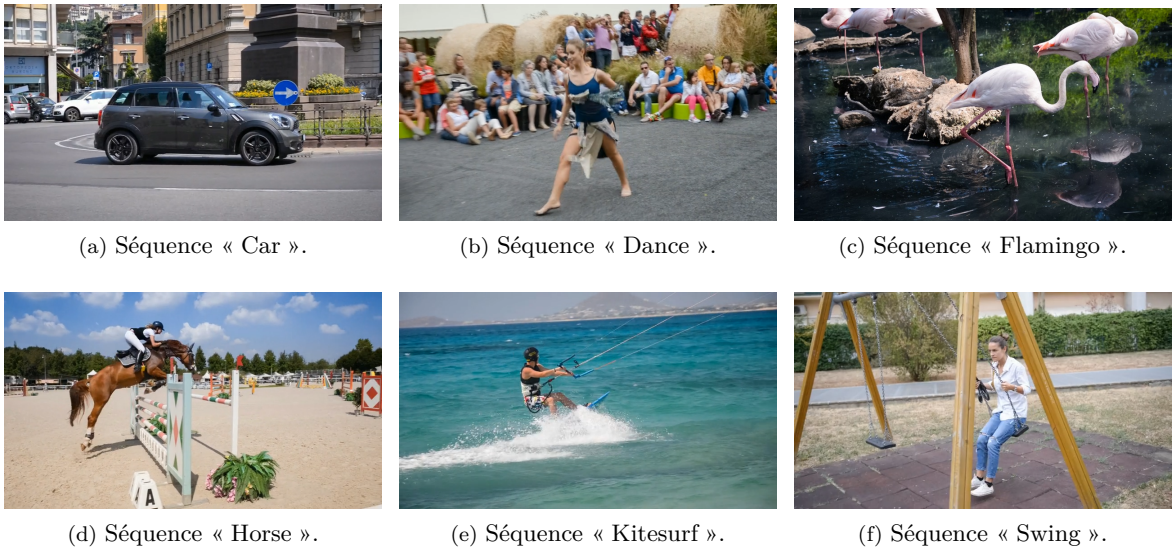
D'autre part, les mesures de qualité reportées dans la TABLE 4.3 montrent que notre algorithme améliore significativement les résultats, en comparaison de ceux obtenus avec la reconstruction de la structure seule ou de la texture seule.

	Reconstruction de la texture			
		T_B	T_H	T_W
Sans reconstruction de la structure	17,53 - 0,937	33,16 - 0,987	34,35 - 0,989	34,31 - 0,988
Avec reconstruction de la structure	27,76 - 0,975	**38,11 - 0,994**	**38,87 - 0,994**	**38,58 - 0,994**

TABLE 4.3 – Mesures de PSNR - SSIM pour les différentes reconstructions de l'image centrale de la séquence « Dumptruck ». L'ordre des scores est le même que celui des images de la FIGURE 4.6 : (4.6a, 4.6b, 4.6c, 4.6d) sur la première ligne et (4.6e, 4.6f, 4.6g, 4.6h) sur la seconde ligne.

4.4 Comparaisons avec d'autres modèles

Ayant choisi la reconstruction de la texture T_W, qui offre le meilleur compromis entre mélange de pixels et valorisation du meilleur pixel entre l'avant et l'arrière, nous souhaitons comparer notre algorithme avec d'autres modèles : celui de [Newson *et al.*, 2014a] qui effectue la recopie de *patchs*, celui de [Burger *et al.*, 2018] qui opère la diffusion, et celui par apprentissage profond de [Xu *et al.*, 2019]. Pour ce faire, nous avons choisi six séquences vidéo de 20 images (voir FIGURE 4.7).



(a) Séquence « Car ». (b) Séquence « Dance ». (c) Séquence « Flamingo ».

(d) Séquence « Horse ». (e) Séquence « Kitesurf ». (f) Séquence « Swing ».

FIGURE 4.7 – Première image de chacune des six séquences utilisées pour comparer notre algorithme à d'autres modèles.

Des défauts synthétiques ont été ajoutés dans un des quatre quarts de chacune des images (de l'image 2 à l'image 19). Le quart de l'image est tiré aléatoirement en empêchant de choisir le même quart pour deux images successives, ce qui semble réaliste quant à la disposition des défauts de type tache dans les films. La forme de ces défauts est issue d'images de taches de peintures qui ont été binarisées (voir FIGURE 4.8).

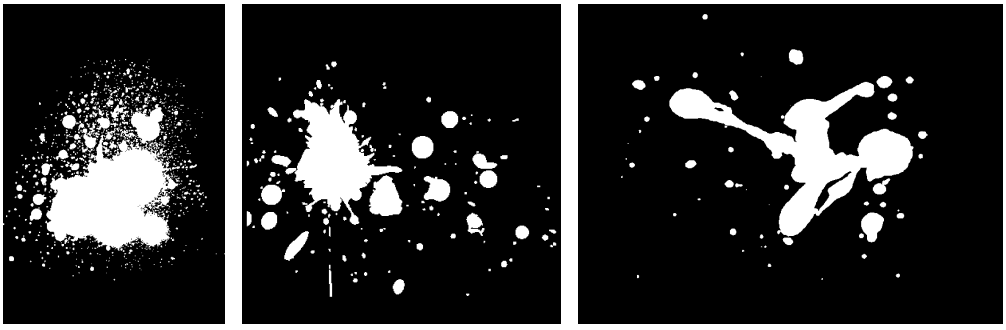


FIGURE 4.8 – Exemples de masques utilisés pour ajouter des défauts artificiels dans les séquences.

4.4.1 Séquence « Flamingo »

(a) Image défectueuse. (b) Newson *et al.* (c) Burger *et al.*

(d) Image saine. (e) Xu *et al.* (f) Notre algorithme.

FIGURE 4.9 – Zoom sur la zone restaurée dans l'image 11 de la séquence « Flamingo ». Les principales difficultés des reconstructions se concentrent sur les pattes du flamant situé à l'arrière-plan (4.9c) et sur le cou du flamant situé au premier plan (4.9e). Si les pattes perdent en verticalité pour les algorithmes utilisant la structure (4.9c et 4.9f), notre reconstruction semble visuellement la meilleure (4.9f).

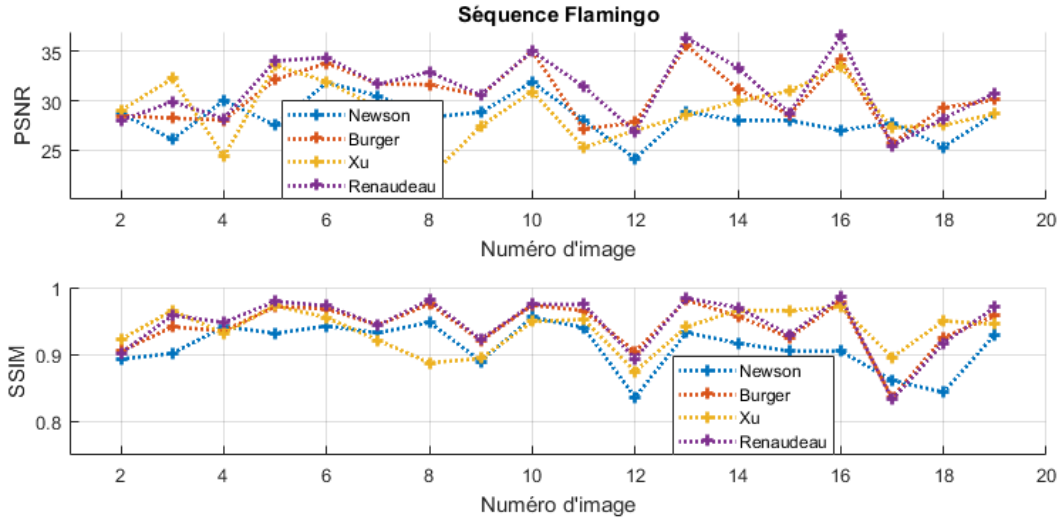
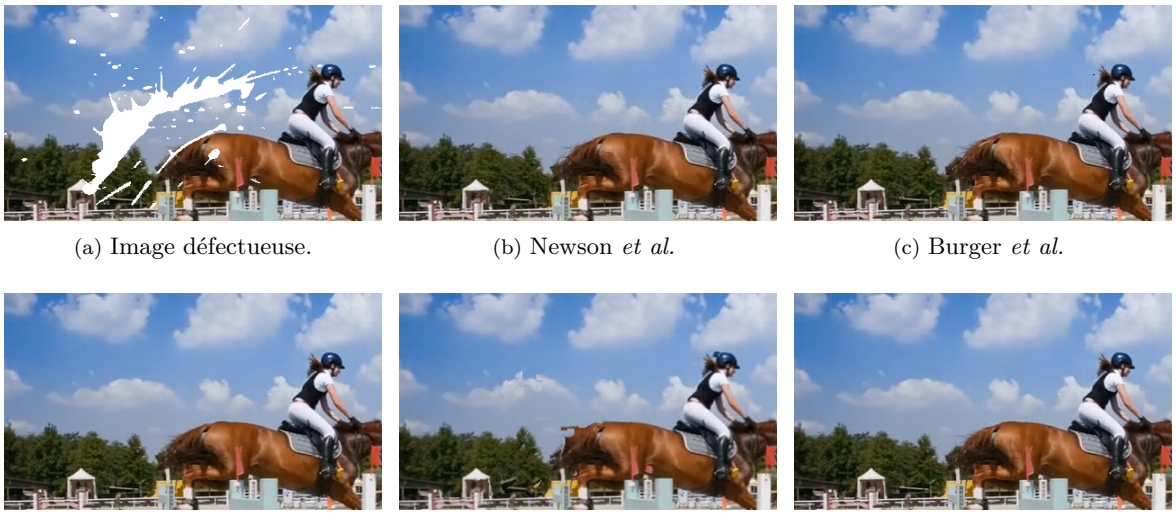


FIGURE 4.10 – Valeurs du PSNR et de la SSIM pour les 18 images restaurées de la séquence « Flamingo » en utilisant quatre méthodes différentes. Notre algorithme (en violet) obtient la plupart du temps les meilleurs résultats.

4.4.2 Séquence « Horse »



(a) Image défectueuse. (b) Newson *et al.* (c) Burger *et al.*

(d) Image saine. (e) Xu *et al.* (f) Notre algorithme.

FIGURE 4.11 – Zoom sur la zone restaurée dans l'image 10 de la séquence « Horse ». Les méthodes qui ne reconstruisent pas la structure (4.11b et 4.11e) ne parviennent pas à restaurer correctement les arbres et les nuages. L'ajout de la texture dans notre algorithme (4.11f) permet une meilleure reconstruction de l'arrière-train du cheval par rapport à la seule reconstruction de la structure (4.11c).

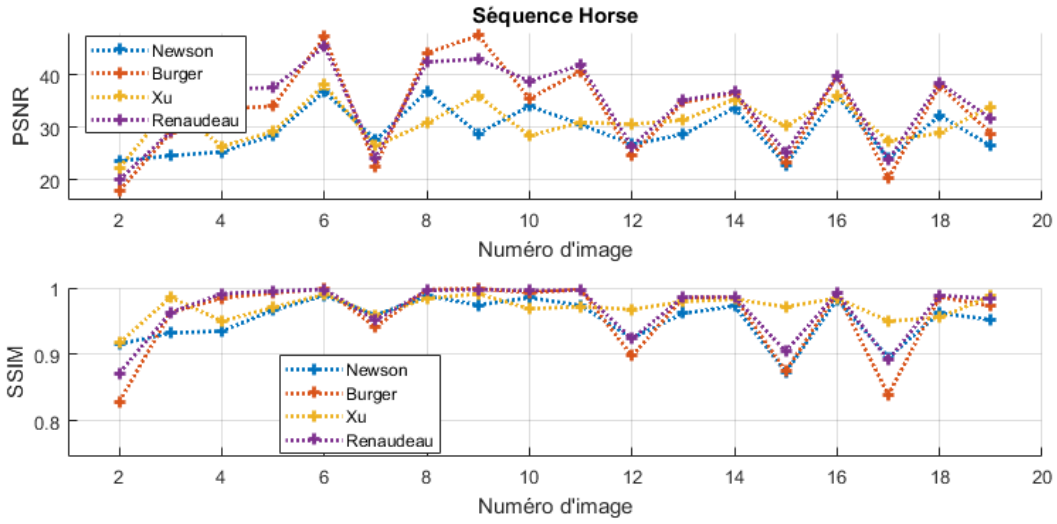


FIGURE 4.12 – Valeurs du PSNR et de la SSIM pour les 18 images restaurées de la séquence « Horse » en utilisant quatre méthodes différentes. Aucune de ces méthodes ne se détache nettement des autres, mais la nôtre obtient quand même les meilleurs résultats.

4.4.3 Comparaison quantitative pour les six séquences

Afin de comparer les différentes méthodes sur l'ensemble des images restaurées des différentes séquences, nous avons calculé la moyenne du PSNR (voir TABLE 4.4) et la moyenne de la SSIM (voir TABLE 4.5) pour effectuer un classement. Pour le PSNR, notre modèle arrive dans les deux premières places dans tous les cas. La SSIM est plus avantageuse pour l'approche par apprentissage profond de [Xu *et al.*, 2019], mais notre méthode reste tout de même bien placée.

	Car	Dance	Flamingo	Horse	Kitesurf	Swing
Newson *et al.*	**29,71**	28,17	28,31	29,32	**31,61**	28,97
Burger *et al.*	25,61	32,00	**30,52**	**33,25**	29,36	29,50
Xu *et al.*	27,46	**33,31**	28,88	30,93	29,28	**32,51**
Notre algorithme	**28,19**	**32,37**	**31,26**	**34,28**	**30,75**	**29,93**

TABLE 4.4 – Moyenne du PSNR sur les 18 images restaurées de chacune des six séquences, en utilisant les quatre méthodes déjà citées (première place en vert foncé, deuxième place en vert clair). Notre méthode arrive en tête ou en seconde position dans tous les cas.

	Car	Dance	Flamingo	Horse	Kitesurf	Swing
Newson *et al.*	**0,942**	0,880	0,911	0,952	**0,953**	0,908
Burger *et al.*	0,862	0,920	**0,942**	0,957	0,940	0,901
Xu *et al.*	**0,935**	**0,961**	0,937	**0,971**	0,951	**0,958**
Notre algorithme	0,903	**0,931**	**0,947**	**0,967**	**0,952**	**0,920**

TABLE 4.5 – Moyenne de la SSIM sur les 18 images restaurées de chacune des six séquences, en utilisant les quatre méthodes comparées (première place en vert foncé, deuxième place en vert clair). Notre algorithme semble moins performant vis-à-vis de cette métrique, même s'il reste dans les deux premières places dans cinq cas sur six. La méthode par apprentissage profond, quant à elle, obtient de meilleurs scores, en particulier pour les scènes « Car » et « Horse », là où le PSNR n'était pas si concluant.

4.5 Vers la restauration de films contenant des défauts réels

Dans notre modèle, la partie qui traite de la structure effectue une diffusion au niveau pixellique. Les formes et les couleurs sont alors bien reconstruites. L'ajout au modèle d'une partie traitant la texture, qui utilise une moyenne non locale à partir de *patchs*, permet de récupérer du contenu de hautes fréquences pour apporter des détails à l'image. Par une approche multirésolution, les différents modèles apportent chacun leur contribution à une bonne correction des défauts.

Si les résultats sont convaincants lorsqu'il s'agit de défauts synthétiques, l'application visée réside dans l'élimination de défauts réels présents dans les films anciens. Il nous reste donc maintenant à mettre en place un pipeline complet de restauration de films, comportant une étape de détection des défauts, et une étape de correction de ces derniers.

Chapitre 5

Conclusion

Une fois réalisée l'étape de détection des défauts, décrite dans le chapitre 3, et l'étape de correction de ces derniers, décrite dans le chapitre 4, qui ont été traitées séparément, nous pouvons mettre en place un pipeline de restauration permettant d'effectuer les deux étapes en séquence. Nous pourrons alors tester ce pipeline sur une autre séquence provenant de la Cinémathèque de Toulouse, afin de valider son bon fonctionnement. Les résultats en sortie du pipeline nous permettront d'évaluer visuellement la qualité des restaurations effectuées sur les images défectueuses.

Après un bilan sur les travaux réalisés au cours de cette thèse, nous envisagerons ensuite les différentes pistes d'amélioration de notre algorithme. Ces dernières concernent autant la partie détection que la partie correction. Nous conclurons enfin sur les futurs travaux liés à la restauration de films, que nous poursuivrons en partenariat avec la Cinémathèque de Toulouse, sous la forme d'un post-doctorat de 18 mois. Ces travaux sont censés porter sur les suites de notre algorithme de restauration, mais aussi sur la colorisation et sur la super-résolution.

5.1 Pipeline de restauration : détection et correction des défauts

Notre pipeline de détection et de correction des défauts (voir FIGURE 5.1) prend en entrée trois images défectueuses consécutives. Dans un premier temps, les trois images sont découpées en *patchs* pour former des triplets, qui constituent l'entrée du réseau de neurones U-net visant à détecter les défauts (voir chapitre 3). Une fois tous les *patchs* des masques de défauts obtenus par inférence du réseau, ces derniers sont alors fusionnés pour former le masque complet des défauts. La dernière étape du pipeline consiste à corriger les défauts, à partir des trois mêmes images défectueuses et du masque de défauts. Ce masque devient alors une variable intermédiaire cachée à l'intérieur du pipeline de restauration.

La partie opérant la détection est réalisée en Python, alors que la partie opérant la correction est réalisée en Matlab, avec des accélérations en C par le biais de fonctions MEX. Dans l'optique de rendre ce pipeline plus homogène, nous envisageons d'uniformiser l'ensemble du code dans le même langage, et de créer une interface pour les utilisateurs.

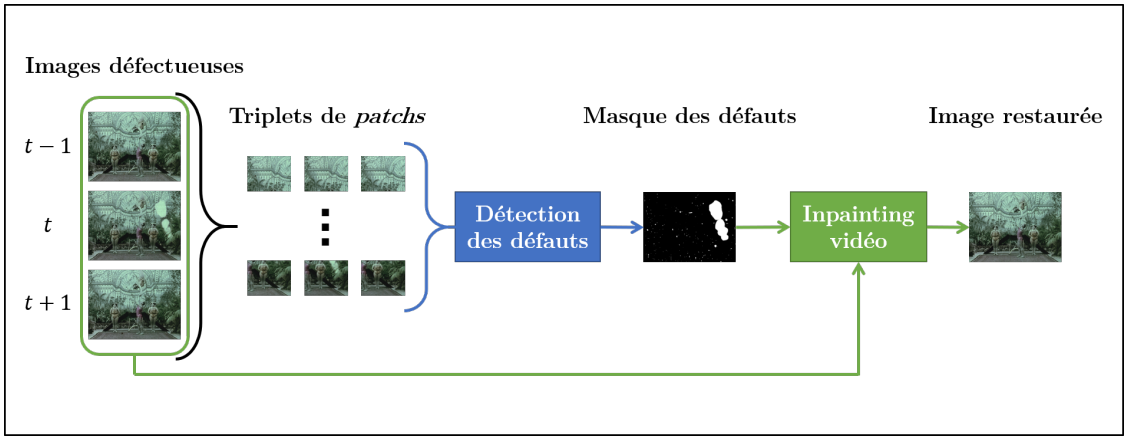


FIGURE 5.1 – Pipeline de restauration de films constitué d'une étape de détection des défauts par apprentissage profond à partir de trois images consécutives, ce qui permet de former un masque binaire, suivie d'une étape de correction des défauts par *inpainting* vidéo à partir de ces trois mêmes images et du masque des défauts.

5.2 Restauration d'une séquence de la Cinémathèque

Nous utilisons une autre séquence détériorée de la Cinémathèque de Toulouse, nommée « Acrobates », afin de tester la qualité de la restauration obtenue avec notre pipeline. Cette séquence étant en couleur, des adaptations ont dû être effectuées, par exemple la transformation des images en niveaux de gris, transformation nécessaire à la partie détection.

Comme cela est illustré sur la FIGURE 5.2, les défauts, même de grande taille comme ceux de l'image 11 (5.2a et 5.2b), sont détectés et corrigés (5.2c). Si certaines rayures n'ont pas été complètement détectées, comme dans l'image 19 (5.2e et 5.2f), des défauts sur des personnes en mouvement, comme cela est le cas de l'un des personnages féminins dans l'image 19 (5.2g) sont quand même détectés et restaurés de manière convaincante (5.2h).

(a) Image défectueuse 11. (b) Zoom de (a). (c) Restauration de (b).

(d) Image défectueuse 19. (e) Zoom 1 de (d). (f) Restauration de (e).

(g) Zoom 2 de (d). (h) Restauration de (g).

FIGURE 5.2 – Restaurations de deux images de la séquence « Acrobates ».

5.3 Bilan et perspectives

Les résultats de notre algorithme de restauration sont encourageants, que ce soit pour la détection des défauts ou pour la correction de ces derniers. Le réseau de neurones parvient à détecter les défauts à partir d'une restauration semi-automatique opérée par un expert de la Cinémathèque de Toulouse, qui s'aide pour cela d'un logiciel professionnel. Notre correcteur, quant à lui, combine les avantages de deux types de reconstruction, celle de la structure et celle de la texture, pour obtenir de très bons résultats. Cependant, quelques pistes sont envisagées pour aller plus loin.

Pour le détecteur, l'augmentation des données d'apprentissage pourrait donner plus de variété quant aux types de défauts à repérer. Même avec le seul jeu de données en notre possession, il serait possible par exemple d'inverser les niveaux de gris des images et de les utiliser soit sous la forme de données supplémentaires, soit pour augmenter le nombre de *patchs* en entrée, qui passeraient de trois à six (les trois *patchs* d'origine et ces trois mêmes *patchs* avec le niveau de gris inversé), pour améliorer la détection par le réseau des défauts de couleur foncée. D'autre part, la prise en compte du mouvement entre images semble nécessaire pour éviter les fausses détections dues à un déplacement trop important entre images successives.

Pour le correcteur, le flux optique a besoin d'être très régularisé pour remplacer les mauvaises valeurs de ce dernier au niveau de la localisation des défauts, ce qui peut entraîner de nouvelles incohérences dans l'estimation des déplacements, que les cartes de correspondance ne peuvent pas toujours compenser. De plus, le temps de calcul par image n'est pas négligeable, en comparaison des approches par apprentissage profond. Pour corriger dix-huit images d'une séquence qui en compte vingt, l'algorithme de [Newson *et al.*, 2014a] nécessite deux heures de temps de calcul, contre environ trente minutes pour le nôtre et celui de [Burger *et al.*, 2018], quand celui de [Xu *et al.*, 2019] opère en moins d'une minute. Une idée pour résoudre ces deux problèmes serait d'intégrer un calcul de flux optique par apprentissage profond qui permettrait à la fois d'éviter certaines incohérences, tout en diminuant significativement le temps de calcul.

La qualité d'une restauration doit aussi prendre en compte des critères de perception, et ceux-ci ne sont pas forcément les mêmes entre le grand public et un public averti composé d'experts travaillant dans le milieu du cinéma. Par exemple, la restauration du grain n'est pas souhaitée par le public amateur, alors qu'elle fait partie intégrante de l'œuvre pour l'œil d'un professionnel. La validation d'un certain niveau de qualité doit être menée prochainement de concert avec la Cinémathèque de Toulouse, ce qui nous permettra de bénéficier d'échanges avec différents experts du milieu de la restauration cinématographique. Le choix de cette validation « experte » est motivée par une volonté des professionnels de la restauration de films de se rapprocher le plus possible de la version initiale du film, avec son cachet de l'époque du tournage.

Dans la continuité des travaux menés durant cette thèse, l'après-thèse va s'articuler autour du projet FLAVIA (restauration de FiLms par Approches Variationnelles et Intelligence Artificielle) au sein de l'IRIT, mais toujours en collaboration avec la Cinémathèque de Toulouse. Ce projet porte sur les trois axes de recherche que sont la restauration, la colorisation (sans le concours de la Cinématèque de Toulouse ici, dont la politique est de privilégier l'authenticité des films en noir et blanc) et la super-résolution.

Le premier axe vise à améliorer les deux étapes de traitement permettant la restauration de films argentiques numérisés, à savoir : la détection automatique, image par image, des défauts présents sur la pellicule avant numérisation (rayures, taches, brûlures) et la restauration de ces défauts par *inpainting* vidéo. Les pistes d'amélioration ont déjà été évoquées dans le paragraphe précédent.

Comme la restauration de films anciens ne se limite pas à la simple correction des défauts (voir le paragraphe 2 du chapitre 1), l'un des deux autres axes du projet FLAVIA est la colorisation des films en noir et blanc. Comme pour l'*inpainting*, l'approche la plus classique pour résoudre le problème inverse que constitue la colorisation est l'approche variationnelle [Levin *et al.*, 2004, Pierre *et al.*, 2017], mais les travaux les plus récents se tournent de plus en plus vers l'apprentissage profond [He *et al.*, 2018].

Enfin, un autre problème identifié par nos partenaires de la Cinémathèque de Toulouse est celui des films enregistrés sur bande magnétique par les caméscopes des années 80 (sous des formats tels que VHS, Video 8, ou Betamax). Le nombre de lignes de ces formats analogiques, de l'ordre de quelques centaines, est adapté à la lecture sur petit écran mais pas à la projection sur grand écran. De fait, toute une génération de films amateurs, dont la Cinémathèque est dépositaire, sont impossibles à projeter tels quels dans une salle de cinéma. Il a alors été convenu que le seul moyen de remédier à ce problème était d'augmenter la résolution des images numérisées. C'est pourquoi la super-résolution intervient en tant que troisième axe du projet FLAVIA. Ce problème est très similaire à celui de l'*inpainting* vidéo, si l'on considère que les zones à reconstruire sont des inter-lignes supplémentaires. La super-résolution fait partie des trois problèmes qui ont été très en vogue dans la communauté des mathématiques appliquées jusqu'à l'émergence de l'apprentissage profond, avec l'*inpainting* et l'acquisition comprimée (*compressive sensing*). Et là encore, les méthodes provenant de l'approche variationnelle [Mitzel *et al.*, 2009, Liu et Sun, 2013] cèdent peu à peu la place aux approches utilisant des réseaux de neurones [Shi *et al.*, 2016, Lucas *et al.*, 2019].

Bibliographie

[Anandan, 1989] Anandan, P. (1989). A Computational Framework and an Algorithm for the Measurement of Visual Motion. *International Journal of Computer Vision*, 2(3):283–310.

[Arias *et al.*, 2011] Arias, P., Facciolo, G., Caselles, V. et Sapiro, G. (2011). A Variational Framework for Exemplar-based Image Inpainting. *International Journal of Computer Vision*, 93(3): 319–347.

[Aubert *et al.*, 1999] Aubert, G., Deriche, R. et Kornprobst, P. (1999). Computing Optical Flow via Variational Techniques. *SIAM Journal on Applied Mathematics*, 60(1):156–182.

[Aujol *et al.*, 2010] Aujol, J.-F., Ladjal, S. et Masnou, S. (2010). Exemplar-based Inpainting from a Variational Point of View. *SIAM Journal on Mathematical Analysis*, 42(3):1246–1285.

[Barnes *et al.*, 2009] Barnes, C., Shechtman, E., Finkelstein, A. et Goldman, D. B. (2009). Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24.

[Bergen *et al.*, 1992] Bergen, J. R., Anandan, P., Hanna, K. J. et Hingorani, R. (1992). Hierarchical Model-based Motion Estimation. *In Proceedings of the 2nd European Conference on Computer Vision*, pages 237–252. Springer.

[Bertalmio *et al.*, 2000] Bertalmio, M., Sapiro, G., Caselles, V. et Ballester, C. (2000). Image Inpainting. *In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co.

[Biemond *et al.*, 1999] Biemond, J., van Roosmalen, P. M. B. et Lagendijk, R. L. (1999). Improved Blotch Detection by Postprocessing. *In Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3101–3104. IEEE.

[Bruni *et al.*, 2008] Bruni, V., Ferrara, P. et Vitulano, D. (2008). Color Scratches Removal using Human Perception. *In Proceedings of the 5th International Conference on Image Analysis and Recognition*, volume 5112 de *Lecture Notes in Computer Science*, pages 33–42.

[Bruni et Vitulano, 2004] Bruni, V. et Vitulano, D. (2004). A Generalized Model for Scratch Detection. *IEEE Transactions on Image Processing*, 13(1):44–50.

[Bruni *et al.*, 2004] Bruni, V., Vitulano, D. et Kokaram, A. C. (2004). Fast Removal of Line Scratches in Old Movies. *In Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 827–830.

[Buades *et al.*, 2010] Buades, A., Delon, J., Gousseau, Y. et Masnou, S. (2010). Adaptive Blotches Detection for Film Restoration. *In Proceedings of the 17th International Conference on Image Processing*, pages 3317–3320. IEEE.

[Bugeau et Bertalmio, 2009] Bugeau, A. et Bertalmio, M. (2009). Combining Texture Synthesis and Diffusion for Image Inpainting. *In Proceedings of the 4th International Conference on Computer Vision Theory and Applications*, pages 26–33.

[Burger *et al.*, 2018] Burger, M., Dirks, H. et Schonlieb, C.-B. (2018). A Variational Model for Joint Motion Estimation and Image Reconstruction. *SIAM Journal on Imaging Sciences*, 11(1):94–128.

[Buyssens *et al.*, 2015] Buyssens, P., Daisy, M., Tschumperlé, D. et Lézoray, O. (2015). Exemplar-based Inpainting: Technical Review and New Heuristics for Better Geometric Reconstructions. *IEEE Transactions on Image Processing*, 24(6):1809–1824.

[Cao *et al.*, 2011] Cao, F., Gousseau, Y., Masnou, S. et Pérez, P. (2011). Geometrically Guided Exemplar-based Inpainting. *SIAM Journal on Imaging Sciences*, 4(4):1143–1179.

[Chambolle et Pock, 2011] Chambolle, A. et Pock, T. (2011). A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145.

[Chan, 2001] Chan, T. (2001). Local Inpainting Models and TV Inpainting. *SIAM Journal on Applied Mathematics*, 62(3):1019–1043.

[Cheung *et al.*, 2006] Cheung, S.-C. S., Zhao, J. et Venkatesh, M. V. (2006). Efficient Object-based Video Inpainting. *In Proceedings of the 2006 International Conference on Image Processing*, pages 705–708. IEEE.

[Chishima et Arakawa, 2009] Chishima, K. et Arakawa, K. (2009). A Method of Scratch Removal from Old Movie Film using Variant Window by Hough Transform. *In Proceedings of the 9th International Symposium on Communications and Information Technology*, pages 1559–1563.

[Cocquerez *et al.*, 2003] Cocquerez, J. P., Chanas, L. et Blanc-Talon, J. (2003). Simultaneous Inpainting and Motion Estimation of Highly Degraded Video-sequences. *In Proceedings of the 13th Scandinavian Conference on Image Analysis*, pages 685–692. Springer.

[Criminisi *et al.*, 2004] Criminisi, A., Pérez, P. et Toyama, K. (2004). Region Filling and Object Removal by Exemplar-based Image Inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212.

[Daisy *et al.*, 2015] Daisy, M., Buyssens, P., Tschumperlé, D. et Lézoray, O. (2015). Exemplar-based Video Completion with Geometry-guided Space-time Patch Blending. *In Proceedings of the 2015 SIGGRAPH Asia Conference*, page 3. ACM.

[Delon et Desolneux, 2010] Delon, J. et Desolneux, A. (2010). Stabilization of Flicker-like Effects in Image Sequences through Local Contrast Correction. *SIAM Journal on Imaging Sciences*, 3(4):703–734.

[Efros et Leung, 1999] Efros, A. A. et Leung, T. K. (1999). Texture Synthesis by Non-parametric Sampling. *In Proceedings of the 7th International Conference on Computer Vision*, volume 2, pages 1033–1038. IEEE.

[Giusti et Williams, 1984] Giusti, E. et Williams, G. H. (1984). *Minimal Surfaces and Functions of Bounded Variation*, volume 80. Springer.

[Guillemot et Le Meur, 2013] Guillemot, C. et Le Meur, O. (2013). Image Inpainting: Overview and Recent Advances. *IEEE Signal Processing Magazine*, 31(1):127–144.

[Gullu *et al.*, 2008] Gullu, M. K., Urhan, O. et Erturk, S. (2008). Blotch Detection and Removal for Archive Film Restoration. *AEU - International Journal of Electronics and Communications*, 62(7):534–543.

[He *et al.*, 2018] He, M., Chen, D., Liao, J., Sander, P. V. et Yuan, L. (2018). Deep Exemplar-based Colorization. *ACM Transactions on Graphics*, 37(4):1–16.

[Horn et Schunck, 1981] Horn, B. K. et Schunck, B. G. (1981). Determining Optical Flow. *Artificial Intelligence*, 17(1-3):185–203.

[Huang *et al.*, 2016] Huang, J.-B., Kang, S. B., Ahuja, N. et Kopf, J. (2016). Temporally Coherent Completion of Dynamic Video. *ACM Transactions on Graphics*, 35(6):196:1–196:11.

[Iizuka *et al.*, 2017] Iizuka, S., Simo-Serra, E. et Ishikawa, H. (2017). Globally and Locally Consistent Image Completion. *ACM Transactions on Graphics*, 36(4):1–14.

[Ilg *et al.*, 2017] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A. et Brox, T. (2017). Flownet 2.0: Evolution of Optical Flow Estimation with Deep Networks. *In Proceedings of the 30th Conference on Computer Vision and Pattern Recognition*, pages 2462–2470. IEEE.

[Jiang *et al.*, 2020] Jiang, Y., Xu, J., Yang, B., Xu, J. et Zhu, J. (2020). Image Inpainting Based on Generative Adversarial Networks. *IEEE Access*, 8:22884–22892.

[Joyeux *et al.*, 1999] Joyeux, L., Buisson, O., Besserer, B. et Boukir, S. (1999). Detection and Removal of Line Scratches in Motion Picture Films. *In Proceedings of the 2nd Conference on Computer Vision and Pattern Recognition*, volume 1, pages 548–553. IEEE.

[Kao *et al.*, 2007] Kao, Y.-T., Shih, T. K., Zhong, H.-Y. et Dai, L.-K. (2007). Scratch Line Removal on Aged Films. *In Proceedings of the 9th International Symposium on Multimedia*, pages 147–151. IEEE.

[Keller *et al.*, 2008] Keller, S. H., Lauze, F. et Nielsen, M. (2008). Deinterlacing using Variational Methods. *IEEE Transactions on Image Processing*, 17(11):2015–2028.

[Keller *et al.*, 2011] Keller, S. H., Lauze, F. et Nielsen, M. (2011). Video Super-Resolution using Simultaneous Motion and Intensity Calculations. *IEEE Transactions on Image Processing*, 20(7): 1870–1884.

[Khriji *et al.*, 2005] Khriji, L., Meribout, M. et Gabbouj, M. (2005). Detection and Removal of Video Defects using Rational-based Techniques. *Advances in Engineering Software*, 36(7):487–495.

[Kim *et al.*, 2019a] Kim, D., Woo, S., Lee, J.-Y. et Kweon, I. S. (2019a). Deep Video Inpainting. *In Proceedings of the 32nd Conference on Computer Vision and Pattern Recognition*, pages 5792–5801. IEEE.

[Kim *et al.*, 2019b] Kim, D., Woo, S., Lee, J.-Y. et Kweon, I. S. (2019b). Recurrent Temporal Aggregation Framework for Deep Video Inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5):1038–1052.

[Kim et Kim, 2007] Kim, K.-T. et Kim, E. Y. (2007). Automatic Film Line Scratch Removal System Based on Spatial Information. *In Proceedings of the 2007 International Symposium on Consumer Electronics*, pages 1–5. IEEE.

[Kokaram, 1996] Kokaram, A. C. (1996). Detection and Removal of Line Scratches in Degraded Motion Picture Sequences. *In Proceedings of the 8th European Signal Processing Conference*, pages 1–4.

[Kokaram, 2004] Kokaram, A. C. (2004). On Missing Data Treatment for Degraded Video and Film Archives: A Survey and a New Bayesian Approach. *IEEE Transactions on Image Processing*, 13(3):397–415.

[Kokaram *et al.*, 2003] Kokaram, A. C., Dahyot, R., Pitie, F. et Denman, H. (2003). Simultaneous Luminance and Position Stabilization for Film and Video. *In Image and Video Communications and Processing 2003*, volume 5022, pages 688–699. International Society for Optics and Photonics.

[Kokaram *et al.*, 1995] Kokaram, A. C., Morris, R. D., Fitzgerald, W. J. et Rayner, P. J. (1995). Detection of Missing Data in Image Sequences. *IEEE Transactions on Image Processing*, 4(11):1496–1508.

[Kokaram et Rayner, 1992] Kokaram, A. C. et Rayner, P. J. (1992). System for the Removal of Impulsive Noise in Image Sequences. *In Visual Communications and Image Processing'92*, volume 1818 de *Proceedings of the SPIE*, pages 322–331.

[Kornprobst *et al.*, 1998] Kornprobst, P., Deriche, R. et Aubert, G. (1998). Image Sequence Restoration: A PDE Based Coupled Method for Image Restoration and Motion Segmentation. *In Proceedings of the 5th European Conference on Computer Vision*, pages 548–562. Springer.

[Lauze et Nielsen, 2004] Lauze, F. et Nielsen, M. (2004). A Variational Algorithm For Motion Compensated Inpainting. *In Proceedings of the 2004 British Machine Vision Conference*, pages 1–11.

[Lauze et Nielsen, 2018] Lauze, F. et Nielsen, M. (2018). On Variational Methods for Motion Compensated Inpainting. *arXiv preprint (from Technical Report of 2009).*

[Le *et al.*, 2017] Le, T. T., Almansa, A., Gousseau, Y. et Masnou, S. (2017). Motion-Consistent Video Inpainting. *In Proceedings of the 24th International Conference on Image Processing*, pages 2094–2098. IEEE.

[Levin *et al.*, 2004] Levin, A., Lischinski, D. et Weiss, Y. (2004). Colorization using Optimization. *In ACM SIGGRAPH 2004 Papers*, pages 689–694.

[Li *et al.*, 2013] Li, H., Lu, Z., Wang, Z., Ling, Q. et Li, W. (2013). Detection of Blotch and Scratch in Video Based on Video Decomposition. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(11):1887–1900.

[Liu et Sun, 2013] Liu, C. et Sun, D. (2013). On Bayesian Adaptive Video Super Resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):346–360.

[Lucas *et al.*, 2019] Lucas, A., Lopez-Tapia, S., Molina, R. et Katsaggelos, A. K. (2019). Generative Adversarial Networks and Perceptual Losses for Video Super-Resolution. *IEEE Transactions on Image Processing*, 28(7):3312–3327.

[Lucas et Kanade, 1981] Lucas, B. D. et Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. *In Proceedings of the 7th International Joint Conference on Artificial Intelligence - volume 2*, pages 674–679.

[Masnou et Morel, 1998] Masnou, S. et Morel, J.-M. (1998). Level Lines Based Disocclusion. *In Proceedings of the 1998 International Conference on Image Processing*, pages 259–263. IEEE.

[Matsushita *et al.*, 2006] Matsushita, Y., Ofek, E., Ge, W., Tang, X. et Shum, H.-Y. (2006). Full-Frame Video Stabilization with Motion Inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1150–1163.

[Mitzel *et al.*, 2009] Mitzel, D., Pock, T., Schoenemann, T. et Cremers, D. (2009). Video Super Resolution using Duality Based TV-L1 Optical Flow. *In Proceedings of the 31st Joint Pattern Recognition Symposium*, pages 432–441. Springer.

[Nadenau et Mitra, 1997] Nadenau, M. J. et Mitra, S. K. (1997). Blotch and Scratch Detection in Image Sequences Based on Rank Ordered Differences. *In Time-Varying Image Processing and Moving Object Recognition, 4*, pages 27–35. Elsevier.

[Newson *et al.*, 2014a] Newson, A., Almansa, A., Fradet, M., Gousseau, Y. et Pérez, P. (2014a). Video Inpainting of Complex Scenes. *SIAM Journal on Imaging Sciences*, 7(4):1993–2019.

[Newson *et al.*, 2013] Newson, A., Almansa, A., Gousseau, Y. et Pérez, P. (2013). Temporal Filtering of Line Scratch Detections in Degraded Films. *In Proceedings of the 20th International Conference on Image Processing*, pages 4088–4092. IEEE.

[Newson *et al.*, 2014b] Newson, A., Almansa, A., Gousseau, Y. et Pérez, P. (2014b). Robust Automatic Line Scratch Detection in Films. *IEEE Transactions on Image Processing*, 23(3):1240–1254.

[Newson *et al.*, 2017] Newson, A., Delon, J. et Galerne, B. (2017). A Stochastic Film Grain Model for Resolution-Independent Rendering. *In Proceedings of the Computer Graphics Forum*, volume 36, pages 684–699. Wiley Online Library.

[Newson *et al.*, 2012] Newson, A., Pérez, P., Almansa, A. et Gousseau, Y. (2012). Adaptive Line Scratch Detection in Degraded Films. *In Proceedings of the 9th European Conference on Visual Media Production*, pages 66–74.

[Niklaus *et al.*, 2017a] Niklaus, S., Mai, L. et Liu, F. (2017a). Video Frame Interpolation via Adaptive Convolution. *In Proceedings of the 30th Conference on Computer Vision and Pattern Recognition*, pages 670–679. IEEE.

[Niklaus *et al.*, 2017b] Niklaus, S., Mai, L. et Liu, F. (2017b). Video Frame Interpolation via Adaptive Separable Convolution. *In Proceedings of the 16th International Conference on Computer Vision*, pages 261–270. IEEE.

[Pathak *et al.*, 2016] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. et Efros, A. A. (2016). Context Encoders: Feature Learning by Inpainting. *In Proceedings of the 29th Conference on Computer Vision and Pattern Recognition*. IEEE.

[Patwardhan *et al.*, 2005] Patwardhan, K. A., Sapiro, G. et Bertalmio, M. (2005). Video Inpainting of Occluding and Occluded Objects. *In Proceedings of the 2005 International Conference on Image Processing*, volume 2, pages II–69. IEEE.

[Patwardhan *et al.*, 2007] Patwardhan, K. A., Sapiro, G. et Bertalmío, M. (2007). Video Inpainting Under Constrained Camera Motion. *IEEE Transactions on Image Processing*, 16(2):545–553.

[Pierre *et al.*, 2017] Pierre, F., Aujol, J.-F., Bugeau, A. et Ta, V.-T. (2017). Interactive Video Colorization within a Variational Framework. *SIAM Journal on Imaging Sciences*, 10(4):2293–2325.

[Ren et Vlachos, 2007] Ren, J. et Vlachos, T. (2007). Segmentation-Assisted Detection of Dirt Impairments in Archived Film Sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(2):463–470.

[Rockafellar, 1976] Rockafellar, R. T. (1976). Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898.

[Ronneberger *et al.*, 2015] Ronneberger, O., Fischer, P. et Brox, T. (2015). U-net: Convolutional Networks for Biomedical Image Segmentation. *In Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention*, volume 9351 de *Lecture Notes in Computer Science*, pages 234–241. Springer.

[Rudin *et al.*, 1992] Rudin, L. I., Osher, S. et Fatemi, E. (1992). Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268.

[Shi *et al.*, 2016] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D. et Wang, Z. (2016). Real-Time Single Image and Video Super-Resolution using an Efficient Sub-pixel Convolutional Neural Network. *In Proceedings of the 29th Conference on Computer Vision and Pattern Recognition*, pages 1874–1883. IEEE.

[Shih *et al.*, 2004] Shih, T. K., Chang, R.-C., Lu, L.-C. et Huang, H.-C. (2004). Multi-layer Inpainting on Chinese Artwork [Restoration Applications]. *In Proceedings of the 2004 International Conference on Multimedia and Expo*, volume 1, pages 21–24. IEEE.

[Shih *et al.*, 2006] Shih, T. K., Lin, L. H. et Lee, W. (2006). Detection and Removal of Long Scratch Lines in Aged Films. *In Proceedings of the 2006 International Conference on Multimedia and Expo*, pages 477–480. IEEE.

[Shiratori *et al.*, 2006] Shiratori, T., Matsushita, Y., Tang, X. et Kang, S. B. (2006). Video Completion by Motion Field Transfer. *In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition*, volume 1, pages 411–418. IEEE.

[Sizyakin *et al.*, 2017] Sizyakin, R., Gapon, N., Shraifel, I., Tokareva, S. et Bezuglov, D. (2017). Defect Detection on Videos using Neural Network. *In Proceedings of the 13th International Scientific-Technical Conference "Dynamic of Technical Systems"*, volume 132 de *MATEC Web of Conferences*, page 05014.

[Sizyakin *et al.*, 2019] Sizyakin, R., Voronin, V., Gapon, N., Pismenskova, M. et Nadykto, A. (2019). A Blotch Detection Method for Archive Video Restoration using a Neural Network. *In Proceedings of the 11th International Conference on Machine Vision*, volume 11041, pages 230–237.

[Storey, 1985] Storey, R. (1985). Electronic Detection and Concealment of Film Dirt. *SMPTE Journal*, 94(6):642–647.

[Tilie *et al.*, 2007] Tilie, S., Bloch, I. et Laborelli, L. (2007). Fusion of Complementary Detectors for Improving Blotch Detection in Digitized Films. *Pattern Recognition Letters*, 28(13):1735–1746.

[Tschumperlé, 2006] Tschumperlé, D. (2006). Fast Anisotropic Smoothing of Multi-Valued Images using Curvature-preserving PDE's. *International Journal of Computer Vision*, 68(1):65–82.

[Wang et Mirmehdi, 2012] Wang, X. et Mirmehdi, M. (2012). Archive Film Defect Detection and Removal: An Automatic Restoration Framework. *IEEE Transactions on Image Processing*, 21(8): 3757–3769.

[Wexler *et al.*, 2004] Wexler, Y., Shechtman, E. et Irani, M. (2004). Space-Time Video Completion. *In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–I. IEEE.

[Wexler *et al.*, 2007] Wexler, Y., Shechtman, E. et Irani, M. (2007). Space-Time Completion of Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):463–476.

[Xu *et al.*, 2007] Xu, J., Guan, J., Wang, X., Sun, J., Zhai, G. et Li, Z. (2007). An OWE-Based Algorithm for Line Scratches Restoration in Old Movies. *In Proceedings of the 2007 International Symposium on Circuits and Systems*, pages 3431–3434. IEEE.

[Xu *et al.*, 2019] Xu, R., Li, X., Zhou, B. et Loy, C. C. (2019). Deep Flow-Guided Video Inpainting. *In Proceedings of the 32nd Conference on Computer Vision and Pattern Recognition*, pages 3723–3732. IEEE.

[Xu *et al.*, 2015] Xu, Z., Wu, H. R., Yu, X. et Qiu, B. (2015). Features-Based Spatial and Temporal Blotch Detection for Archive Video Restoration. *Journal of Signal Processing Systems*, 81(2):213–226.

[Yous et Serir, 2016] Yous, H. et Serir, A. (2016). Blotch Detection in Archived Video Based on Regions Matching. *In Proceedings of the 2016 International Symposium on Signal, Image, Video and Communications*, pages 379–383. IEEE.

[Yous et Serir, 2017] Yous, H. et Serir, A. (2017). Spatio-temporal Blotches Detection and Removal in Archive Video. *In Proceedins of the 2017 Conference on Intelligent Systems and Computer Vision*, pages 1–6. IEEE.

[Yous *et al.*, 2019] Yous, H., Serir, A. et Yous, S. (2019). CNN-based Method for Blotches and Scratches Detection in Archived Videos. *Journal of Visual Communication and Image Representation*, 59:486–500.

[Yu *et al.*, 2018] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. et Huang, T. S. (2018). Generative Image Inpainting with Contextual Attention. *In Proceedings of the 31st Conference on Computer Vision and Pattern Recognition*, pages 5505–5514. IEEE.

[Zach *et al.*, 2007] Zach, C., Pock, T. et Bischof, H. (2007). A Duality Based Approach for Realtime TV-L1 Optical Flow. *In Proceedings of the 29th Joint Pattern Recognition Symposium*, pages 214–223. Springer.