Avis de Soutenance

Monsieur Gwendal BERNARDI

Informatique

Soutiendra à huis clos ses travaux de thèse intitulés

Fusion des modalités et des angles de vue avec des méthodes d'apprentissages profonds pour l'inspection verrière

Travaux dirigés par Emmanuel DELLANDREA

Soutenance prévue le ***jeudi 05 novembre 2026*** à 14h00

Lieu : Centrale Lyon, 36 Av. Guy de Collongue, 69130 Écully

Salle : Amphi 1

Composition du jury proposé

M. Emmanuel DELLANDREA	Maître de conférences HDR	Ecole Centrale de Lyon	Directeur de thèse
M. Jean-Yves RAMEL	Professeur	Université Savoie Mont Blanc	Rapporteur
Mme Valérie GOUET-BRUNET	Directrice de recherche	IGN-ENSG	Rapporteure
M. Mohsen ARDABILIAN	Maître de conférences	Ecole Centrale de Lyon	Co-directeur de thèse
Mme Anissa MOKRAOUI	Professeure des universités	Université Sorbonne Paris Nord	Examinatrice
M. Godefroy BRISEBARRE	Docteur	Tiama	Co-encadrant de thèse
M. Alexandre BENOIT	Professeur des universités	Polytech Annecy Chambéry	Examinateur
M. Sébastien ROMAN	Tiama	Invité	

Mots-clés : Apprentissage Automatique, Fusion de modalités et d'angle de vue, Inspection Verrière

Résumé :

Cette thèse s'inscrit dans le contexte de l'inspection automatique des défauts dans l'industrie verrière, un environnement particulièrement contraig- nant caractérisé par des cadences de production élevées, des propriétés optiques complexes du matériau et une grande variabilité des défauts. En raison de la transparence, de la réflectivité et des effets de réfraction du verre, certains défauts ne sont visibles que sous des conditions d'éclairage ou des angles de vue spécifiques. Les systèmes modernes d'inspection reposent ainsi sur des configurations multi-vues et multi-modales, générant des données hétérogènes difficiles à exploiter efficacement avec les approches classiques. L'objectif principal de cette thèse est d'étudier comment la fusion d'images multi-modales et multi-vues basée sur l'apprentissage profond peut améliorer la fiabilité, la robustesse et l'interprétabilité des systèmes d'inspection automa- tique, tout en respectant les contraintes industrielles telles que le temps réel et l'incomplétude des données. Dans un premier temps, un état de l'art appro- fondi des méthodes de fusion d'images est proposé, accompagné d'une nouvelle taxonomie permettant d'unifier les paradigmes supervisés, non supervisés et ori- entés tâche. Cette classification introduit également la distinction entre méthodes mono-catégorie et multi-catégorie, mettant en évidence les limites des approches existantes et la nécessité de modèles génériques capables de traiter simultané- ment plusieurs modalités et points de vue. La contribution principale de cette thèse est l'introduction de EDIF (End-to-End Detection-Driven Image Fusion), une méthode de fusion orientée détection pour les données multi-vues et multi- modales. EDIF s'appuie sur des réseaux de neurones à attention sur graphe afin de modéliser les relations entre des points caractéristiques extraits de différentes images. Cette approche permet de raisonner explicitement sur les correspon- dances inter-vues et inter-modales, et d'unifier dans une même architecture l'extraction de caractéristiques, leur mise en correspondance et la détection des défauts. Une stratégie d'apprentissage en plusieurs étapes est proposée afin d'optimiser conjointement la correspondance, la localisation et la classification. Afin de valider ces approches, un nouveau jeu de données industriel, MMDOD, est développé. Il regroupe plusieurs milliers d'images annotées acquises selon différentes modalités et angles de vue, constituant un benchmark réaliste pour l'évaluation de méthodes de détection en conditions industrielles. Une seconde contribution consiste en un modèle de complétion de points clés basé sur des graphes, permettant de traiter les cas de données manquantes, fréquents dans les contextes industriels. En exploitant des contraintes de cohérence multi-vues et 4 multi-modales via des mécanismes de propagation dans le graphe, cette méthode permet de maintenir des performances élevées même en présence d'informations partielles.Enfin, une architecture compacte basée sur les modèles à états d'espace de type Mamba est proposée pour la détection de marques d'usure de surface sur des contenants en verre réutilisables. Ces défauts, caractérisés par des patterns d'abrasion progressive induits par les cycles d'utilisation industrielle, présentent une forte dépendance au point de vue et constituent un cas particulièrement difficile à modéliser. Cette approche offre un compromis intéressant entre per- formance et coût computationnel, la rendant compatible avec les contraintes de production en temps réel. Ce travail montre que les approches de fusion orien- tées tâche et basées sur des graphes permettent d'améliorer significativement la détection de défauts dans des environnements industriels complexes. Il propose à la fois des contributions méthodologiques et des solutions applicatives visant à rapprocher les avancées en fusion d'images de leur déploiement en production

Summary:

This thesis addresses the problem of automated defect inspection in glass manufacturing, a challenging industrial context characterized by high production rates, complex material properties, and strong variability in defect appearance. Due to the transparency, reflectivity, and refractive behavior of glass, defects are often only visible under specific illumination conditions or viewing angles. As a result, modern inspection systems rely on multi-view and multi-modal imaging configurations, generating heterogeneous data that remain difficult to exploit effectively with conventional methods. The main objective of this work is to investigate how deep-learning–based multi- modal and multi-view image fusion can improve the reliability, robustness, and interpretability of automated inspection systems, while satisfying industrial constraints such as real-time processing and incomplete data acquisition. First, this thesis provides a comprehensive state of the art on image fusion, introducing a novel taxonomy that unifies supervised, unsupervised, and task- driven learning paradigms, and distinguishes between mono-category and multi- category fusion methods. This framework highlights the limitations of existing approaches and motivates the need for generalist fusion architectures capable of jointly handling multiple modalities and viewpoints. The core contribution of this work is the introduction of EDIF (End-to-End Detection-Driven Image Fusion), a task-driven framework designed for multi-view and multi-modal defect detection. EDIF relies on a graph attention neural network to model relationships between keypoints extracted from multiple images, enabling ex- plicit reasoning across views and modalities. The method integrates keypoint detection, feature matching, and object detection into a unified pipeline, al- lowing consistent grouping and classification of defects across all available observations. A dedicated multi-stage training strategy is proposed to jointly learn correspondence, localization, and classification tasks. To support this research, a novel industrial dataset, MMDOD, is constructed, comprising thou- sands of annotated images acquired from multiple viewpoints and modalities, providing a realistic benchmark for multi-view multi-modal detection under real production conditions. Building upon EDIF, a second contribution intro- duces a graph-based multi-view keypoint completion framework, designed to address missing data scenarios frequently encountered in industrial acquisi- tion. By leveraging graph message passing and multi-view consistency con- straints, the proposed approach enables robust inference even when some views or modalities are absent. Finally, this thesis proposes a compact multi-view 2 detection architecture based on Mamba state-space models, specifically de- signed for the detection of surface wearing marks on returnable glass containers. This type of defect, characterized by progressive abrasion patterns induced by repeated industrial usage cycles, is inherently viewpoint-dependent and particularly challenging to model. The proposed approach combines computa- tional efficiency with strong performance, making it suitable for deployment in high-throughput industrial environments. This work demonstrates that task- driven, graph-based fusion strategies significantly improve defect detection performance in complex industrial settings. It provides new methodological insights and practical solutions for bridging the gap between advanced im- age fusion research and real-world deployment in glass inspection systems.