

Ecole Doctorale

*Mathématiques et Sciences et Technologies de l'Information et de la Communication*

*ISEN Laboratoire*

**AVIS DE SOUTENANCE DE THESE**

**Le lundi 18 janvier 2021 à 14h30**

à l'Université de Monastir (Tunisie).

**Monsieur KORTLI YASSIN**

soutiendra une thèse de doctorat sur le sujet suivant :

Conception d'un système de vision embarqué auto-adaptatif et faible-consommation

**Le jury sera ainsi composé :**

- **M. AL FALOU AYMAN, Professeur**  
ISEN Yncréa-Ouest - BREST
- **M. ATRI MOHAMED, Maître de conférences**  
Université de Monastir - 5019 MONASTIR - TUNISIE
- **M. DENIS HAMAD, Professeur**  
ULCO France - CALAIS
- **M. DOUIK ALI, Professeur**  
ENIM Monastir -
- **M. JRIDI MAHER, Enseignant-Chercheur**  
ISEN Brest - BREST
- **M. SAKLY ANIS, Professeur**  
ENIM Monastir -

A BREST, le 14 janvier 2021

Le Président de l'Université de  
Bretagne Occidentale,



A handwritten signature in black ink, appearing to read 'J. Gallo'.

**M. GALLOU**

# Résumé

Aujourd'hui, la détection et la reconnaissance des visages font partie de notre vie quotidienne. Le visage est déjà utilisé dans différentes zones du monde pour déverrouiller les téléphones portables, retirer de l'argent aux distributeurs automatiques, payer dans les magasins, effectuer des contrôles dans les aéroports ou identifier des suspects lors de grands événements tels que des matchs de football ou des concerts. Par conséquent, la demande de systèmes de surveillance, à base de détection et de reconnaissance des visages, a considérablement augmenté dans le secteur de la sécurité.

L'objectif de cette thèse est d'implémenter un système de détection et de reconnaissance des visages robuste à faible coût. Ce système est formé par deux modules : la détection et la reconnaissance des visages.

Dans un premier temps, une recherche approfondie sur les différentes approches de détection et de reconnaissance de visages existantes dans l'état de l'art est réalisée. Nous présentons ensuite une analyse critique et une comparaison entre ces approches (locales, sous-espaces (ou holistiques) et hybrides), tout en discutant les limites et les défis de chacune. A ce propos, nous avons choisi les approches locales pour étudier le module de reconnaissance des visages dans la suite de la thèse.

Dans un deuxième temps, nous proposons et étudions les techniques locales les plus populaires pour la réalisation du module de reconnaissance, telles que la technique HOG, LBP, VLC, SIFT, SURF et BRIEF. Ces techniques ont été simulées sur Visual Studio avec le langage C++ et la bibliothèque OpenCV. La base de données PHPID qui contient des visages humains avec différentes poses, expressions et éclairages est utilisée pour la simulation. Une comparaison est ensuite menée pour évaluer la précision et temps d'exécution dans le but d'étudier l'efficacité de ces techniques. Les résultats de cette comparaison nous ont permis de sélectionner la technique VLC qui a été utilisée.

Les deux parties précédentes nous ont permis de nous rapprocher de notre objectif qui est l'implantation d'une architecture à faible coût pour accélérer un système de détection et de reconnaissance des visages qui fonctionne sur une plate-forme hétérogène. Dans ce contexte, nous avons implémenté un système robuste intégrant l'algorithme Viola-Jones avec la technique VLC sur une plateforme Zynq qui combine un processeur ARM avec la partie logique d'un FPGA. Nous avons constaté que la meilleure façon d'accélérer notre système est de le partitionner en parties logicielles et matérielles en profitant de la nouvelle technologie Zynq SoC de Xilinx. Les parties nécessitant du control ou des calculs

séquentiels sont implantés sur le processeur ARM, les parties gourmandes en ressources de calcul seront dédiées à des exécutions sur le FPGA. L'implémentation de la technique VLC sur des cibles reconfigurables (FPGA) est possible grâce à l'utilisation des algorithmes FFTs qui peuvent être exécutés en parallèle.

Nous avons d'abord conçu une architecture matérielle optimisée pour la technique VLC et l'avons implémentée sur la logique programmable permettant d'obtenir de meilleurs résultats par rapport à l'architecture matérielle traditionnelle. L'adaptation de l'architecture proposée pour qu'elle convienne aux FPGA a été un défi. Ce défi consiste à conserver les optimisations de haut niveau pour obtenir une implémentation efficace tout en permettant d'adapter l'architecture sur FPGA de sorte que le calcul est également parallélisé. Pour permettre cette flexibilité, nous avons utilisé l'outil de synthèse Vivado HLS de Xilinx à partir d'un ensemble des *directives* et *pragmas* d'optimisation pour générer la description HDL de l'architecture proposée. L'architecture matérielle de la technique VLC optimisée offre un bon compromis entre l'utilisation des ressources et le temps d'exécution en fournissant une accélération de 4,44x et 39,22x respectivement par rapport à l'architecture traditionnelle et le module logiciel correspondant sur le processeur ARM avec une fréquence de fonctionnement de 190 MHz pour une image de visage de 128x128 pixels.

Fort de ces constats, nous avons étudié trois différentes implémentations de pour notre système proposé sur Zynq SoC. La première est entièrement logicielle sur le processeur ARM et utilise des fonctions OpenCV. Les deux autres implémentations sont des solutions matérielles/logicielles, la première utilise l'architecture matérielle de la technique VLC traditionnelle et la deuxième utilise l'architecture matérielle de la technique VLC optimisée. Cette implémentation permet une accélération de 3x par rapport à l'implémentation logicielle et 1,13x par rapport à la première conception matérielle/logicielle. En outre, pour valider la précision des conceptions matérielles/logicielles proposées, nous analysons également le taux d'erreur par rapport à l'implémentation logicielle qui montre qu'il ne dépasse pas 2%.

# Abstract

Today, face detection and recognition have become part of modern life. The face is already used in different areas of the world to unlock mobile phones, withdraw money from ATMs, pay in shops, make checks in airports or identify persons during major events such as football matches or concerts. Consequently, the demand for surveillance systems, especially for face detection and recognition, has increased considerably in the security field.

The objective of this thesis is to implement a robust and low-cost face detection and recognition system. This system is formed by two modules: face detection and recognition.

Initially, an extensive research on the different approaches for the face detection and recognition systems that exist in the state-of-the-art is realized. We then present a critical analysis and comparison between these approaches (local, holistic and hybrid), while discussing the limitations, challenges and weaknesses of each one. In this respect, we have chosen local approaches to study the face recognition module in the rest of the thesis.

Secondly, we propose and study the most popular local techniques for the implementation of the recognition module, such as HOG, LBP, VLC, SIFT, SURF and BRIEF. These techniques have been simulated on Visual Studio with C++ and OpenCV. The PHPID database that contains human faces with different poses, expressions and lighting is used for the simulation. A comparison is then performed on the accuracy and execution time in order to evaluate the effectiveness of these techniques. The results of this comparison allow to select the VLC technique, which will be used for the rest of the thesis.

The two previous parts allowed to get closer to our objective is to develop a low-cost architecture to accelerate a face detection and recognition system that operates on a heterogeneous platform. In this context, we have implemented a robust system integrating the Viola-Jones algorithm with the VLC technique that has already been selected on a Zynq platform, which combines an ARM processor with an FPGA. We found that the best way to accelerate our system is to split it into software and hardware parts by taking the advantage of the new Zynq SoC technology from Xilinx. This choice was made to provide a parallel processing of all parts of our system at the same time. The implementation of the VLC technique on reconfigurable targets (FPGAs) is possible thanks to the use of FFTs algorithms, which can be executed in parallel.

We first designed an optimized hardware architecture for the VLC technique and implemented it on programmable logic allowing better results compared to traditional hardware architecture. The adaptation of the proposed architecture to be suitable to FPGAs

has been a challenge. This challenge consists to preserve the high-level optimizations to obtain an efficient implementation while allowing to adapt the architecture on FPGAs while the processing is also parallelized. To achieve this flexibility, we used Vivado HLS synthesis tool from Xilinx using a set of optimization directives and pragmas to generate the HDL description of the proposed architecture. The hardware architecture of the optimized VLC technique provides a good trade-off between resources utilization and execution time with an acceleration of 4.44x and 39.22x compared to the traditional architecture and the corresponding software module on the ARM processor with an operation frequency of 190 MHz for a 128x128 pixel face image, respectively.

Thirdly, three different implementations on Zynq SoC were studied for our proposed system. The first one is completely software based on the ARM processor and uses OpenCV functions. The two other implementations are hardware/software co-designs, where the first one uses the hardware architecture of the traditional VLC technique and the second one uses the hardware architecture of the optimized VLC technique. This implementation provides an acceleration of 3x compared to the software implementation and 1.13x compared to the first hardware/software co-design. Additionally, to validate the accuracy of the proposed hardware/software co-designs, we also analyze the error rate compared to the software implementation that shows that it does not exceed 2%.