UNIVERSITÀ DEGLI STUDI DI SALERNO

Dottorato in Informatica e Ingegneria dell'Informazione
Curriculum Informatica

Coordinatore:
Prof. Alfredo De Santis
XV Nuovo Ciclo

About the Development of Visual Search Algorithms and their Hardware Implementations

Relatori:
Ch.mo Prof. Giancarlo Raiconi
Ph.D. Mario Vigliar

Candidato:
Luca Puglia
Matr. 8888100007

Anno Accademico 2016/2017

Abstract

The main goal of my work is to exploit the benefits of a hardware implementation of a 3D visual search pipeline. The term visual search refers to the task of searching objects in the environment starting from the real world representation. Object recognition today is mainly based on scene descriptors, an unique description for special spots in the data structure. This task has been implemented traditionally for years using just plain images: an image descriptor is a feature vector used to describe a position in the images. Matching descriptors present in different viewing of the same scene should allows the same spot to be found from different angles, therefore a good descriptor should be robust with respect to changes in: scene luminosity, camera affine transformations (rotation, scale and translation), camera noise and object affine transformations. Clearly, by using 2D images it is not possible to be robust with respect to the change in the projective space, e.g. if the object is rotated with respect to the up camera axes its 2D projection will dramatically change. For this reason, alongside 2D descriptors, many techniques have been proposed to solve the projective transformation problem using 3D descriptors that allow to map the shape of the objects and consequently the surface real appearance. This category of descriptors relies on 3D Point Cloud and Disparity Map to build a reliable feature vector which is invariant to the projective transformation. More sophisticated techniques are needed to obtain the 3D representation of the scene and, if necessary, the texture of the 3D model and obviously these techniques are also more computationally intensive than the simple image capture. The field of 3D model acquisition is very broad, it is possible to distinguish between two main categories: active and passive methods. In the active methods category we can find special devices able to obtain 3D information projecting special light and. Generally an infrared projector is coupled with a camera: while the infrared light projects a well known and fixed pattern, the camera will receive the information of the patterns reflection on a certain surface and the distortion in the pattern will give the precise depth of every point in the scene. These kind of sensors are of

course expensive and not very efficient from the power consumption point of view, since a lot of power is wasted projecting light and the use of lasers also imposes eye safety rules on frame rate and transmissed power. Another way to obtain 3D models is to use passive stereo vision techniques, where two (or more) cameras are required which only acquire the scene appearance. Using the two (or more) images as input for a stereo matching algorithm it is possible to reconstruct the 3D world. Since more computational resources will be needed for this task, hardware acceleration can give an impressive performance boost over pure software approach.

In this work I will explore the principal steps of a visual search pipeline composed by a 3D vision and a 3D description system. Both systems will take advantage of a parallelized architecture prototyped in RTL and implemented on an FPGA platform. This is a huge research field and in this work I will try to explain the reason for all the choices I made for my implementation, e.g. chosen algorithms, applied heuristics to accelerate the performance and selected device. In the first chapter we explain the Visual Search issues, showing the main components required by a Visual Search pipeline. Then I show the implemented architecture for a stereo vision system based on a Bio-informatics inspired approach, where the final system can process up to 30fps at 1024 × 768 pixels. After that a clever method for boosting the performance of 3D descriptor is presented and as last chapter the final architecture for the SHOT descriptor on FPGA will be presented.