Practical Book I

Laboratory: March 1st, March 8th
Report delivery: March 20th 2016

Introduction

Image matching is a fundamental aspect of many problems in computer vision including object and scene recognition. During the development of this practice we are going to analyze the process for extracting distinctive invariant features from images to perform reliable matching among different views of an object or scene. The extracted feature set should be highly distinctive in order to provide robust matching across a substantial range of affine distortions, changes in 3D viewpoint, addition of noise, and changes in illumination. The aim of the lab is to analyze the characteristics of image descriptors and the matching process for detecting a certain pattern. This practical block is expected to be performed in two sessions (March 1 and 8).

Practical session

The first session will introduce 2D image descriptors used in object recognition, namely SIFT (Scale-Invariant Feature Transform), see [1], and SURF (Speed Up Robust Features), see [2]. Both approaches describe a feature descriptor and a feature detector. The detector extracts from an image a set of significant points in a way that it is consistent with variations of illumination, viewpoint and other viewing conditions The descriptor associates to such points a signature or descriptor which identifies their appearance in a discriminative and compact way, see figure 1a.

The descriptor vectors extracted from different images can be then matched. The matching is often based on a distance between these vectors, e.g. the Euclidean or Mahalanobis distance. A simple way to proceed is to compute the distance between each descriptor in for each point. Then a matched pair from the best score (minimal distance), see figure 1b, is constructed. To increase robustness matches are rejected in relation to nearest neighbor heuristic approaches.

Although the distances approaches described above discard many of the false matches arising from background clutter, we will still may have (false) matches belonging to different objects. Therefore, it is necessary to exclude those points which do not maintain a geometric consistency. For planar scenes this is easy to manage: each possible transformation can be expressed with a homography. A homography (in 3D) is a mapping of a plane to another plane. Mathematically, a homography can be computed by establishing a correspondence between 4 points in each plane. Homographies between pairs of images can be estimated using RANSAC (RANdom SAmple Consensus), see [3]. In case of feature point matching, RANSAC removes outliers in an iterative procedure by randomly selecting a subset of the data and estimating the consensus score based on a classification hypothesis in the remaining set of data.

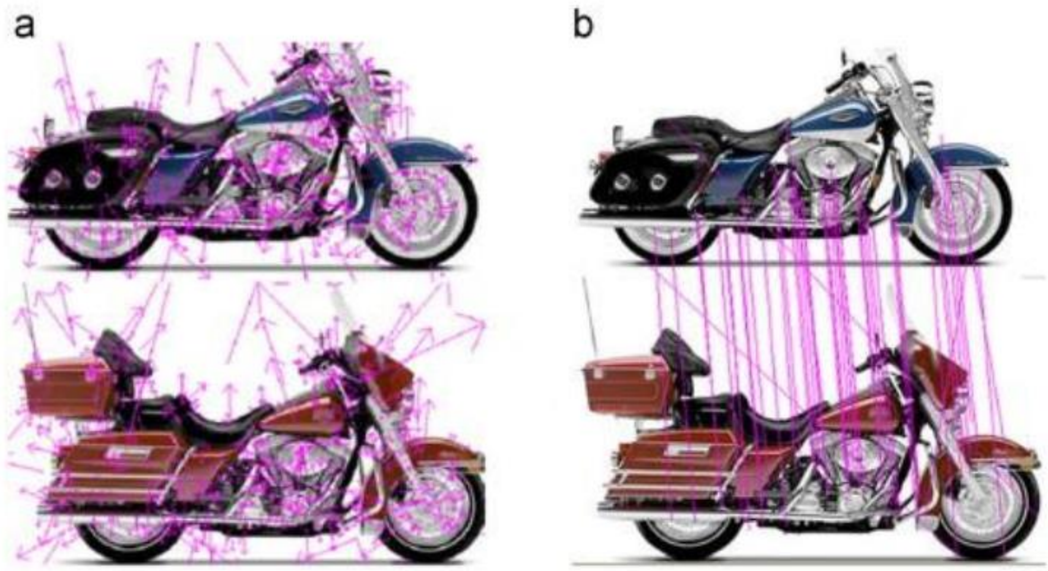

Figure 1: a) 2D detected keypoints and b) matched keypoints.

Work to perform

It is recommended to overview the three referenced papers at the end of this document before the first session. For the first session we will use the “Find Object” application which implements all the methods to test during the practical session (http://introlab.github.io/find-object/).

The report to deliver should address the following key issues:

- Discuss the effect of the main SIFT parameters (octave layers, edge/contrast threshold, sigma value).
- Discuss the effect of the main SURF parameters (octave layers, hessian threshold, upright flag).
- Comparison of two descriptors to scale, rotation, change in 3D viewpoint, change in illumination invariance and computational efficiency.
- Analysis of the heuristic approaches for matching points and the influence of RANSAC for object recognition.

A single report in PDF format of a maximum of three pages must be delivered for the first practical block. You are expected to use the two practical sessions (March 1 and 8) to experiment and write the report. The report has to be sent to xbaro@uoc.edu by March 20th latest (Subject: MAI PR1).

References:

[1] Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, November 2004.

[2] Bay, H. and Ess, A. and Tuytelaars, T. and Van Gool, L. Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding, June 2008.

[3] Fischler, M.A. and Bolles, R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM. June 1981.