**Extracting interesting points from images**

Kornel Bertók, Gergő Pólyi

*Faculty of Informatics, University of Debrecen, Debrecen, Hungary*

[bertok.kornel@inf.unideb.hu](mailto:bertok.kornel@inf.unideb.hu), [polyigergo@freemail.hu](mailto:polyigergo@freemail.hu)

# Introduction

The purpose of this work is to create a mobile application which will allow a user to discover information about a given building or landmark. After determining in which building the user is interested, specific information can be returned about this building and/or the surrounding area. In particular, this technology can be used as part of a tourist application based in a city space, which will provide both historical and current information about the area.

So in this paper, we study the problem of building recognition. As an instance of the recognition problem, this domain is interesting since the class of buildings posses many similarities, while at the same time calls for techniques which are capable of fine discrimination between different instances of the class.

One of the central issues pertinent to the recognition problem is the choice of suitable representation of the class and its scalability to a large number of exemplars. There is a large amount of literature on general object recognition. The existing methods exploit geometric and/or appearance information, consider part based models or more holistic representations. In the context of the presented work we will review some related works which obtain the desired representations from image appearance, computing both global and local image features.

Global approaches typically consider the entire image as a point in the high-dimensional space and model the changes in the appearance as a function of viewpoint using subspace methods. Given the subspace representation the pose of the camera can be obtained by spline interpolation method, exploiting the continuity of the mapping between the object appearance and continuously changing viewpoint.

Alternative global representations proposed in the past include responses to banks of filters, multidimensional receptive field histograms and color histograms. These representations do not encode spatial information inherent in the image. Although quite robust to changes in viewpoint and/or scale, they are often not very discriminative. Partial means of encoding the spatial information can be obtained by computing the global descriptors over different image regions separately.

Alternative representations in terms of local features have become very effective in the context of different object/category recognition problems. In this case the descriptors are computed only over local image regions, whose location is first determined using various saliency measures. These representations perform favorably in the presence of large amount of clutter and changes in viewpoint. The representatives of local image descriptors include scale invariant features and their associated descriptors, which are robust with respect to affine transformations.

## Outline

This study concentrate only for extracting interesting points from images, the feature matching and indexing is not the part of it. In this paper, we propose the base of a building recognition problem by a two stage hierarchical scheme. The first stage is comprised to extract simultaneously efficient local image descriptors to get a small number of bestcandidate images from the database. These candidates are chosen for the second recognition stage, which rectifies the images along the local image descriptors. Then global image descriptors will be extracted simultaneously from the normalized images to make the final decision.

# Extracting Local Image Features

Most object recognition systems tend to use either global image features, which describe an image as a whole, or local features, which represent image patches. Local features are computed at multiple points in the image and are consequently more robust to occlusion and clutter. However, they may require specialized classification algorithms to handle cases in which there are a variable number of feature vectors per image.

Most local features represent texture in an image patch. For example, SIFT features use histograms of gradient orientations. One advantage of using local features is that they may be used to recognize the object despite significant clutter and occlusion. They also do not require a segmentation of the object from the background, unlike many texture features, or representations of the object’s boundary (shape features).

## SIFT: Scale Invariant Feature Transform

Scale Invariant Feature Transform is an image descriptor for image-based matching. This descriptor is used for a large number of purposes in computer vision related to point matching between different views of a 3D scene and view-based object recognition. The SIFT descriptor is invariant to translations, rotations and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations. It is very useful in practice for image matching and object recognition.

## SURF: Speeded Up Robust Feature

SURF is a robust image detector and descriptor that can be used in computer vision tasks like object recognition or 3D reconstruction. The standard version of SURF is several times faster than SIFT. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images.

## FAST: Features from Accelerated Segment Test

FAST is a corner detection method, which can be used to extract feature points and later used to track and map objects in many computer vision tasks. The advantage of FAST corner detector is its computational efficiency. FAST corner detector is very suitable for real-time video processing application because of high-speed performance.

## ORB: Oriented FAST and Rotated BRIEF

ORB a very fast binary descriptor based on BRIEF, which is rotation invariant and resistant to noise. This feature builds on the well-known FAST keypoint detector and BRIEF descriptor.

## MSER: Maximally Stable Extremal Regions

MSER is used as a method of blob detection in images. This technique was proposed to find correspondences between image elements from two images with different viewpoints. This method of extracting a comprehensive number of corresponding image elements contributes to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms.

## STAR:

# Extracting Global Image Features

Despite the robustness advantages of local features, global features are still useful in applications where a rough segmentation of the object of interest is available. Global features have the ability to generalize an entire object with a single vector. Consequently, their use in standard classification techniques is straightforward. Global features include contour representations, shape descriptors, and texture features.

Global texture features and local features provide different information about the image because the support over which texture is computed varies. We expect classifiers that use global features will commit errors that differ from those of classifiers based on local features. A disadvantage of global features is that they are sensitive to clutter and occlusion. As a result it is either assumed that an image only contains a single object, or that a good segmentation of the object from the background is available.

## LBP: Local Binary Patterns

LBP is a type of feature used for classification. It has been found to be a powerful feature for texture classification. It has further been determined that when LBP is combined with the Histogram of oriented gradients (HOG) classifier, it improves the detection performance considerably on some datasets.

## LRF: Local Rank Functions

## Gabor filters

A Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image.

## HoG: Histogram of Oriented Gradients

HoG are feature descriptors for the purpose of object detection. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

# References