

# Bag-of-Words based Image Classification

Rémi de Zoeten and Anouk Visser

December 27, 2013

## 1 Introduction

In this report we will discuss the implementation and results of a system for image classification. The system can tell if there is an object of one of four given classes (motorbikes, cars, faces and airplanes) in an image. The classification system is based on a Bag-of-Words approach, meaning that the Support Vector Machine that performs the classification uses histograms of words for training and classification of an image. The words are obtained by extracting descriptors from a set of training images and cluster them using the *kmeans* algorithm.

## 2 Implementation

We have extracted key point SIFT features and dense SIFT features. In addition to this we have also trained and evaluated our system using RGBSIFT, rgbSIFT and opponentSIFT, these are just key point- or dense SIFT features applies to all three of the color channels of the specific colorspace, resulting in descriptors of dimension  $128 \times 3$ . For dense SIFT sampling we used a bin size of 10 pixels and a step size of 5.

To build a visual vocabulary, we use the extracted descriptors from a number of training images (of all classes) and use this as the data for the kmeans clustering algorithm. The number of clusters represents the number of words in the vocabulary. Because of our limited computational resources the largest vocabulary we have built is a vocabulary of 100 words based on 5 training images. Our system saves a lot of data to disk, like the different extracted SIFT descriptors and the resulting clusters (i.e. words).

## 3 Evaluation

1. Keypoint SIFT vs. dense sampling SIFT
2. Different SIFTs (i.e. grey, RGB, rgb, Opponent)
3. Different Vocabulary sizes

4. Number of training samples

5. kernel choice for SVM

Please, make a simple document with four ranked lists of test images as discussed in Section 2.6. This document should also contain all your settings (size of visual vocabulary, number of positive and negative samples, and so on), Average Precision per class, and Mean Average Precision.

1. Different sifts over training size

2. Different sift over vocabulary size

3. Different kernel choices for best sift over vocabulary size