# 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 classi-
fication. 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.
We will then use the words to quantize training and test images. We do this by extracting
descriptors from the image and find the descriptor from the vocabulary with the smallest
euclidean distance with respect to the descriptor from the image. We will keep track of
which words (i.e. descriptors from the vocabulary) were selected and how many times this
happened in a histogram. When every descriptor has been 'matched' to a word from the
vocabulary, we normalize the resulting histogram by dividing every bin by the total number
of descriptors extracted from the image. We perform this quantization on both the images
for training and testing.

The histograms of the training images are then used to train a Support Vector Machine. We train one SVM for every class, for the 'motorbikes' we will use the histograms of training images of a motorbike as positive examples and histograms of the other images as negative examples. After we have trained the classifiers we can use these to make predictions about the test histograms.

As an evaluation measure we have used the Mean Average Precision. We provide every classifier with all the test images and rank the results based on the probability that the test image is off the class the classifier is trained on. For every classifier we will thus have a ranked list of test images, if everything works well the top of the list should be dominated by images of the class the classifier can predict. In the next section we will show the results of our system and provide an analysis of these results.

## 3   Evaluation