Invariant color descriptors for efficient object recognition

van de Sande, K.E.A.

Citation for published version (APA):
Chapter 6

Summary and Conclusions

6.1 Summary

In this thesis, we explore methods to recognize ‘what’ object is visible in an image and ‘where’ it is within the image. To recognize which object is visible in an image, a current successful approach is the bag-of-words model. To localise where the object is in the image, exhaustive search is currently most successful. In this thesis, we (1) analyze and (2) propose invariant color descriptors within the bag-of-words model, (3) improve the efficiency of the bag-of-words model by exploiting parallelism, and (4) propose a selective search strategy for object localisation. The results obtained in the thesis are discussed per chapter in the following paragraphs:

Chapter 2: Evaluating Color Descriptors for Object and Scene Recognition. In this chapter we have created a structured overview of color invariant descriptors in the context of image category recognition. So far, intensity-based descriptors have been widely used for feature extraction at salient points. To increase illumination invariance and discriminative power, color descriptors have been proposed. Because many different descriptors exist, we study the invariance properties and the distinctiveness of color descriptors. The analytical invariance properties of color descriptors are explored, using a taxonomy based on invariance properties with respect to photometric transformations, and tested experimentally using a dataset with known illumination conditions. In addition, the distinctiveness of color descriptors is assessed experimentally using two benchmarks, one from the image domain and one from the video domain. From the theoretical and experimental results, it can be derived that invariance to light intensity changes and light color changes affects category recognition. The results reveal further that, for light intensity shifts, the usefulness of invariance is category-specific. Overall, when choosing a single descriptor and no prior knowledge about the dataset and object and scene categories is available, the OpponentSIFT is recommended. Furthermore, a combined set of color descriptors outperforms intensity-based SIFT and improves category recognition on two benchmarks.

Chapter 3: Illumination-Invariant Descriptors for Discriminative Visual Object Categorization. Illumination-invariant color descriptors, as used in the previous chapter, are normally based on a limited set (usually 3) of predefined color channels. However, a predefined
set of 3 color channels may constrain the discriminative power of visual object categorization. Therefore, in this chapter, we aim to generate and select a general set of discriminative, illumination-invariant descriptors for object category recognition. First, we develop a class of new illumination-invariant descriptors based on uniform sampling of the RGB color space. The class of descriptors is proven to be invariant to light intensity changes and shifts under different normalizations. Then, this class of descriptors is used to increase discrimination of visual object categories through different selection strategies including multiple kernel learning and cross-validation. With a strategy to construct a descriptor based on a new color space, we find an optimum at a 6 channel space, which is similar to the (3 channel) opponent color space with 3 additional samplings in chromaticity plane. This new color descriptor performs better than OpponentSIFT (the recommended descriptor from the previous chapter) on both object classification and object localisation tasks.

Chapter 4: Empowering Visual Categorization with the GPU. The bag-of-words model has become the most powerful method for visual categorization of images and video. Despite its high accuracy, a severe drawback of this model is its high computational cost. In this chapter, we have identified two major bottlenecks in the bag-of-words model: the quantization step and the classification step. We address these two bottlenecks by proposing two efficient algorithms for quantization and classification by exploiting parallelism in newer CPU and GPU architectures. The algorithms are designed to (1) keep categorization accuracy intact, (2) decompose the problem and (3) give the same numerical results. In the experiments on large scale datasets it is shown that, by using a parallel implementation on the GPU, classifying unseen images is 4.8 times faster than a quad-core CPU version, while giving the exact same numerical results. In addition, we show how the algorithms can be generalized to other applications, such as text retrieval and video retrieval. Moreover, when the obtained speedup is used to process extra video frames in a video retrieval benchmark, the accuracy of visual categorization is substantially improved.

Chapter 5: Segmentation as Selective Search for Object Recognition. For object localisation, the current state-of-the-art is based on exhaustive search. However, in this chapter we propose a selective search strategy to enable the use of more expensive features and classifiers and thereby progress beyond the state-of-the-art. We adapt segmentation as a selective search by reconsidering segmentation: We propose to generate many approximate locations over few and precise object delineations because (1) an object whose location is never generated cannot be recognized and (2) appearance and immediate nearby context are most effective for object recognition. Our method is class-independent and is shown to cover 96.7% of all objects in the PASCAL VOC benchmark using only 1,500 locations per image. Replacing the exhaustive search strategy of the current state-of-the-art with our selective search reduces the accuracy by only 1%. Our selective search enables the use of the more expensive bag-of-words method which we use to substantially improve the state-of-the-art for 8 out of 20 classes on the PASCAL VOC 2010 object recognition challenge.
6.2 Conclusions

This thesis contributes to more accurate and more efficient object recognition. The first objective of this thesis is to analyze how viewpoint and illumination changes affect existing color descriptors and subsequently visual object classification. It was shown in chapter 2 that invariance to light intensity changes and light color changes affects object classification. Further, for light intensity shifts, the usefulness of invariance is category-specific. Overall, when choosing a single descriptor and no prior knowledge about the dataset and object categories is available, the OpponentSIFT is recommended. With a combined set of color descriptors, we obtain state-of-the-art results on the Pascal VOC object classification and the TRECVID concept classification tasks.

The second objective of this thesis is to design new color descriptors which improve object classification. In chapter 3, we developed a class of new illumination-invariant descriptors based on uniform sampling of the $RGB$ color space. The class of descriptors is proven to be invariant to light intensity changes and shifts under different normalizations. These were found to be the most important invariance properties in chapter 2. With this class of descriptors we increase discrimination of visual object categories. With a new descriptor in a color space with additional samplings in chromaticity plane, we improve over the best descriptor from chapter 2 in both object classification and object localisation tasks.

The third objective of this thesis is to exploit parallelism in CPU and GPU architectures to handle the computational cost of the bag-of-words model. In chapter 4, we have identified two major bottlenecks in the bag-of-words model: the quantization step and the classification step. We proposed two efficient algorithms for quantization and classification by exploiting parallelism, while (1) keeping categorization accuracy intact, (2) decomposing the problem and (3) giving the same numerical results. We have shown that, by using a parallel implementation on the GPU, classifying unseen images is 4.8 times faster than a quad-core CPU version. In addition, the algorithms can be generalized to other applications, such as text retrieval and video retrieval. Also, the accuracy of visual categorization is substantially improved when the obtained speedup is used to process extra video frames, as shown on the TRECVID video retrieval benchmark.

Finally, the fourth objective of this thesis is to create a selective search strategy which visits the locations in an image where it is probable that the object is actually present. We have adapted segmentation as a selective search by reconsidering segmentation: we generate many approximate locations over few and precise object delineations because (1) an object whose location is never generated cannot be recognized and (2) appearance and immediate nearby context are most effective for object recognition. The resulting strategy is class-independent and is shown to cover 96.7% of all objects in the Pascal VOC benchmark using only 1,500 locations per image. Selective search enables the use of the more expensive bag-of-words method which we use to substantially improve the state-of-the-art for 8 out of 20 classes on the Pascal VOC object recognition challenge.

To conclude, with illumination-invariant color descriptors, our exploitation of parallelism and our selective search strategy, we have contributed to object recognition in terms of accuracy.
and efficiency. By making corresponding software available, other researchers have been able to build on top of our work. Whereas object classification is becoming a tool to be used, we are confident that for object localisation, many new search strategies and representations have yet to be explored. Also, as the number of object categories to recognize increases, it will become necessary to organize them in a hierarchy and to exploit this structure.