

COLOR TEXTONS FOR BUILDING DETECTION

Ronny Hänsch and Olaf Hellwich

Technische Universität Berlin
Computer Vision and Remote Sensing; FR3-1
Franklinstr. 28/29; 10587 Berlin, Germany
r.haensch@tu-berlin.de

KEY WORDS: feature extraction, textons, building detection, color information, filter

ABSTRACT:

Textons are known to be powerful operators in capturing textural properties of image regions. This paper proposes a new method to consistently combine structural cues as well as color information in an unified framework of color textons. They are used as features to detect buildings from optical imagery. Despite the simple classification algorithm, presented results are promising and show the usefulness of the proposed feature operator in remote sensing applications.

1 INTRODUCTION

The authors of (Julesz, 1981) define the term “textons” as “putative units of pre-attentive human texture perception”. During the last decades this biological concept has been adapted, extensively investigated, and improved by the computer vision community as for instance in (Malik et al., 1999) and (Zhu et al., 2005).

Nowadays, textons are used to solve various computer vision tasks like texture recognition (Burghouts and Geusebroek, 2006) or scene categorization (Battiatto et al., 2008).

Despite the success in computer vision applications, textons are only seldom applied in remote sensing. One of the few examples is the work of (Garcia-Pineda et al., 2008) that uses textons and an artificial neural network classifier to detect oil slicks in synthetic aperture radar (SAR) images. Another example of using gray-scale textons in remote sensing applications can be found in (He et al., 2008), where gray-value texton histograms serve as one of several features to detect buildings in polarimetric SAR images. The authors of (Sun and He, 2009) even propose a segmentation framework based on normalized cuts and color textons.

Since textons were shown to be very useful for recognition tasks from gray-valued images, there have been several attempts to exploit not only spatial intensity variations, but color information as well. However, previous methods are mostly based on extensions, that are not fully consistent with the original texton algorithm. This work introduces a novel approach to consistently combine color and structural information within a unified framework for color textons.

Building detection is important for various remote sensing applications as for example in urban change detection, city modelling and planning. Therefore, robust as well as accurate algorithms are needed, which are able to detect buildings of various size, form, and color in different environments and lighting conditions. Although buildings mostly follow a clear geometrical structure, the inner-class variability is very high, which complicates the recognition task.

The main purpose of this paper is the introduction of a novel color texton approach and the evaluation of its general usability in building detection tasks from optical imagery. That is why a standard classifier is applied to use color textons as features in or-

der to distinguish between the two classes of ‘building’ and ‘non-building’. Sophisticated classification algorithms would take several additional other cues into account like shape, context information, and prior knowledge. In this work Naïve Bayes is used as a very basic classifier to provide a baseline performance estimation and to be able to better assess the advantages and limitations of the proposed feature operator.

The results are compared to standard gray-level textons as well as another color texton method in order to evaluate the gain in performance achieved by this novel approach. The detection rate is improved by more than 11%, while the false-positive rate is slightly decreased by 3.5% at the same time.

2 GRAY-SCALE TEXTONS

A detailed description of textons is provided for instance in (Zhu et al., 2005). This section gives a brief introduction of standard gray-scale textons in order to familiarize the reader with the key-points of this concept. Color textons as proposed by this work and explained in the following section 3 are a consistent extension of the here described method.

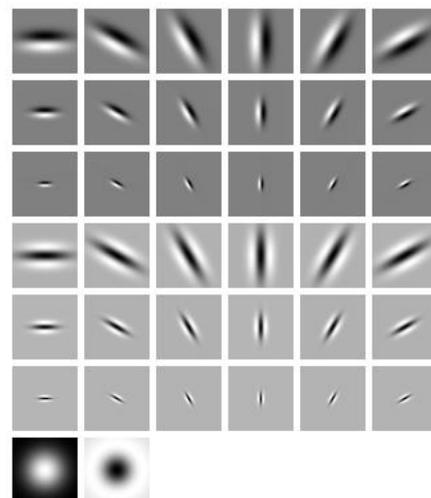


Figure 1: MR8 filter-bank

The first step is the convolution of a set of (gray-scale) images with a certain filter-bank. An often used example is the MR8-filter-bank (Varma and Zisserman, 2005) illustrated in figure 1. This specific filter-bank consists of 38 linear filters of different order, scale, and orientation. Only the maximum response of different orientations of each orientation-dependent filter is used in order to derive rotation invariance as well as to reduce the number of output dimensions. Therefore, the output of the application of this filter-bank to a gray-scale image is one eight-dimensional vector of filter responses for each pixel.

One of the basic assumptions of the texton framework is, that similar textures within the image will result in similar signatures of filter responses. The k-means algorithm finds a pre-defined number of dominant signatures in a subsequent clustering step. The cluster centers serve as signature prototypes and are named textons. Depending on the number of clusters it is assumed that this codebook is able to represent all dominant textures within the actual image data.

Since the filter response differs significantly at different positions within a textured area, a single cluster cannot represent a specific texture in general. Rather, one cluster corresponds to a certain part of a texture, which may or may not be shared with other textures. However, the collective similarity to several cluster centers is able to accurately describe the whole texture. Different textures will correspond to different sets of similarities to the learnt cluster centers.

Therefore, a local texton histogram is calculated over a small neighborhood around each pixel. It represents the relative occurrence of different filter signatures and is able to serve as a powerful texture descriptor. This histogram is exploited by further analysis, for instance classification algorithms.

3 COLOR TEXTONS

Texture is able to serve as a powerful feature for recognition tasks. Nevertheless, some object categories are far easier to recognize by use of radiometric information or a combination of texture and color.

In (Sun and He, 2009) color is included into the texton framework at the level of filter responses by concatenating the filter response vector with the RGB-color vector. The distinction between color and textural information neglects the potential different structural signature within and between the different color channels. One approach to overcome this problem is to independently apply the filter-bank to each color-channel. The authors of (Burghouts and Geusebroek, 2006) investigate two further possibilities to exploit color information. The first method includes color directly at texton level. Color invariants are used to handle shading and shadow. The second method uses the standard gray-value texton approach and applies a color-dependent weighting as post-processing step.

All those methods suffer from the fact, that they do not analyse the whole spectrum of possible color-structure combinations.

Gray-scale textons have their origin in a biological concept. A remarkable fact about the filters of the MR8 filter-bank (see figure 1) is their analogy to the behavior of simple-cells in the visual cortex of mammals.

The receptive field of simple-cells consists of clearly distinct excitatory as well as inhibitory areas, which result in zero activation under diffuse lighting (Nicholls, 2001). Figure 2(a) illustrates an

exemplary receptive field of a simple-cell. An excitatory “on”-part at the center is surrounded by inhibitory “off”-areas. The optimal stimulus for such a cell is a small bright bar surrounded by dark areas. It corresponds to the filters of order two in the MR8 filter-bank. There are simple-cells with receptive fields which match other types of filters, too.

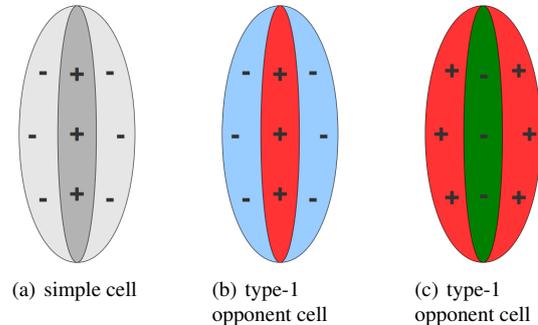


Figure 2: Illustration of receptive fields of cells within the visual cortex

This work follows the idea of biologically inspired textons to exploit color information. There are different kinds of cells in the visual cortex, which are sensitive to color information. One of them is named type-1 opponent cell and has an excitatory or inhibitory center surrounded by antagonistic areas (Nicholls, 2001) similar to color-insensitive simple-cells. However, both areas are sensitive to potentially different colors as illustrated exemplarily in figure 2(b) and 2(c).

For that reason, this work proposes the usage of a color-dependent filter-bank. Basically, the same filters as in the standard gray-scale texton framework are used, namely the MR8 filter-bank. Besides the Gaussian filter, all kernels of this filter-bank have a positive (excitatory) as well as a negative (inhibitory) part. The positive and negative areas of each filter kernel are applied to different color channels i and j of the image \mathbf{I} . The summation of both individual outputs leads to the final filter response f_{ij} .

$$f_{ij}(\mathbf{x}) = \sum_{\mathbf{z} \in D} (\mathbf{I}_i(\mathbf{x})k^+(\mathbf{x} - \mathbf{z}) + \mathbf{I}_j(\mathbf{x})k^-(\mathbf{x} - \mathbf{z})) \quad (1)$$

$$k^+(\mathbf{x}) = \begin{cases} k(\mathbf{x}) & \text{if } k(\mathbf{x}) > 0, \\ 0 & \text{else.} \end{cases} \quad (2)$$

$$k^-(\mathbf{x}) = \begin{cases} k(\mathbf{x}) & \text{if } k(\mathbf{x}) < 0, \\ 0 & \text{else.} \end{cases} \quad (3)$$

where $\mathbf{I}_i(\cdot)$ is the intensity of the i -th color channel, and $k(\cdot)$ is one of the kernels from the filter-bank with domain D .

The result of the application of the MR8 filter-bank in the standard texton framework are 38 filter outputs, which are reduced to 8 dimensions by using only the maximum response of all oriented versions of the same filter. The set of convolutions defined by equation 1 leads to $37 \cdot 9 + 3 = 336$ filter responses for all 9 possible color-channel combinations. Even if only the maximum response over different orientations is used, there are still 66 remaining filter responses. Such a high-dimensional feature space might cause severe problems during the subsequent clustering step.

This work uses principle component analysis (PCA) as dimensionality reduction technique instead of the maximum response approach. Only the first N principle components which explain 95% of the variance are kept for further analysis. During the experiments described in section 4, N was always found to be

less than 20. This shows that PCA is able to significantly reduce the dimensionality of this filter response space, while using only statistically important information instead of an arbitrary subset. Additionally, the orientation dependency of the oriented filters is lost due to the transformation into the eigenvector space.

Due to the application of PCA linear correlations within the data are removed and the covariance matrix is diagonalized, which is advantageous for the subsequent clustering process. As in the standard texton approach k-means is used as clustering algorithm.

Afterwards a texton histogram is estimated from a small neighborhood around each pixel and used as descriptor, which unified captures texture as well as color information.

4 BUILDING DETECTION

Like most man-made objects buildings consist of clear defined geometrical structures, which are often the most dominant visual features. This is particularly true for orthorectified air-born images, where the roof is the dominant visible building part. Parallel and orthogonal lines, rectangles and other simple geometrical objects occur far more often in images of buildings than for example in images of forests. Even other man-made structures as for example streets consist of other arrangements of those geometrical primitives. That is why, textural properties represent a well suited cue for building detection. Another important characteristic of roof tops is color, since the variability of roof colors is rather low and distinct from other object classes.

The main purpose of this paper is the introduction of color textons as defined by the previous section and to evaluate their principle usability in building detection tasks. Therefore, instead of highly sophisticated classification algorithms which would take several other cues (like shape, context information, prior knowledge, etc.) into account, only a very basic classifier is used.

Naïve Bayes assumes the statistical independence of the individual dimensions of the feature space. This assumption does not hold in general, but is mostly clearly wrong. Nevertheless, it still often leads to successful classifications and even outperforms more complicated models. Due to its simple implementation and clear statistical interpretability it serves as a standard baseline approach.

Given the feature vector \mathbf{f} , the class c^* with the largest posterior probability $P(c^*|\mathbf{f})$ is defined as the final classification result. According to Bayes' Theorem $P(c|\mathbf{f})$ can be modelled as

$$P(c|\mathbf{f}) = \frac{p(\mathbf{f}|c) \cdot P(c)}{p(\mathbf{f})} \propto p(\mathbf{f}|c) \cdot P(c) \quad (4)$$

The optimal class decision c^* is estimated under the assumption of a uniform prior as well as independent features by

$$c^* = \arg \max_c P(c|\mathbf{f}) = \arg \max_c p(\mathbf{f}|c) = \arg \max_c \prod_{i=1}^F p(f_i|c) \quad (5)$$

where F is the number of features.

Figure 3(a) shows an exemplarily part of the used data set. The image shows a small area of the city of Dorsten, Germany. In figure 3(c) the final classification result is visualized.

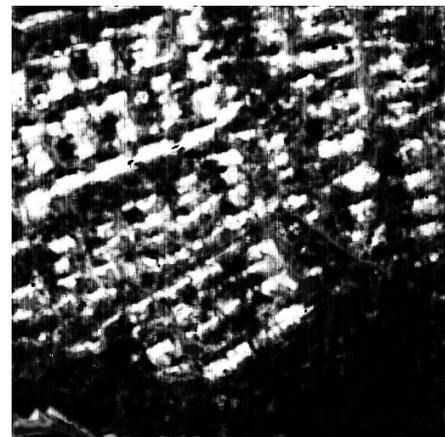
From the 336-component filter response vector, only the 18 first principle components are kept and used as input for k-means. The



(a) Dorsten
 ©"Geobasisdaten: Land NRW, Bonn, 2111/2009"



(b) Reference Data
 ©"Geobasisdaten: Land NRW, Bonn, 2111/2009"



(c) Classification Result

Figure 3: Image and reference data, as well as classification result

number of clusters was fixed to 50. To calculate the texton histogram window size was set to 10. Codebook generation as well as the training of the classifier were performed using a different part of the dataset.

Figure 4 shows the ROC statistics of the proposed color-textons (green), the standard gray-value textons (red), as well as another approach (blue) to include color information into the texton framework by simply applying the filter-bank to all channels individually. The according error rates are estimated by 4-fold cross-validation.

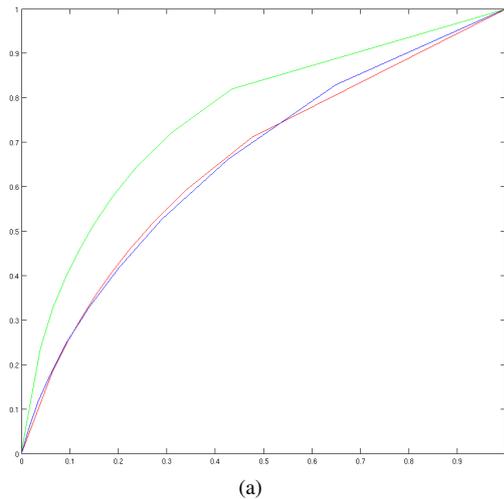


Figure 4: ROC curves under usage of different textons
green: proposed method; blue: filters independently applied to all color-channels; red: grey-scale textons

Grey-valued textons and the naive approach of including color by simply applying the filter-bank to each channel independently show no significant difference. However, the proposed method is significantly better than both of them.

Even under usage of a very simple classification framework promising results are achieved. The actual detection rate increased from 40.1% for grey-valued textons to 51.7% correctly classified building pixels for the proposed color textons. At the same time the false-positive rate slightly dropped from 18.6% to 15.1%.

It should be noted that no context or prior knowledge was used. Spatial relations were only analysed in terms of spatially extended filters and local pixel neighborhoods to calculate texton histograms.

Although a rather large percentage of building pixels is not correctly labelled, the larger part of most buildings is. The majority of false-positives are very close to actual buildings or occur between two buildings which are close to each other. Those facts indicate that the proposed color textons are indeed capable of capturing radiometric and textural information relevant to detect buildings from optical imagery.

5 CONCLUSION AND FUTURE WORK

Despite the simple classification method the results are already very promising. It was shown, that they outperform not only standard grey-valued textons, but naive approaches to include color, too.

Future work will investigate the performance of the proposed color textons for further computer vision tasks. The implications of PCA with respect to the usage of a specific filter-bank will be addressed, as will be other clustering techniques.

One critical point of the proposed algorithm are the computational costs. The number of necessary convolutions during code-book generation, PCA, and the number of iterations during the clustering process lead to rather long processing times. Future research will concentrate on ways how to reduce the amount of time needed to extract color textons at least during the application phase of the proposed framework.

REFERENCES

- Battiato, S., Farinella, G., Gallo, G. and Ravi, D., 2008. Scene categorization using bag of textons on spatial hierarchy. In: 15th IEEE International Conference on Image Processing, pp. 2536–2539.
- Burghouts, G. and Geusebroek, J., 2006. Color textons for texture recognition. In: British Machine Vision Conference, Vol. III, pp. 1099–1108.
- Garcia-Pineda, O., MacDonald, I. and Zimmer, B., 2008. Synthetic aperture radar image processing using the supervised textural-neural network classification algorithm. In: IEEE International Geoscience and Remote Sensing Symposium, Vol. IV, pp. 439–442.
- He, W., Reigber, A. and Hellwich, O., 2008. Building extraction from polarimetric sar data using mean shift and conditional random fields. In: Proceedings of EUSAR 2008, Vol. 3, pp. 1265–1268.
- Julesz, B., 1981. Textons, the elements of texture perception and their interactions. *Nature* 290, pp. 91–97.
- Malik, J., Belongie, S., Shi, J. and Leung, T., 1999. Textons, contours and regions: Cue integration in image segmentation. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2, IEEE, pp. 918–925.
- Nicholls, J., 2001. *From Neuron to Brain*. Palgrave Macmillan.
- Sun, F. and He, J., 2009. The remote-sensing image segmentation using textons in the normalized cuts framework. In: International Conference on Mechatronics and Automation, pp. 1877–1881.
- Varma, M. and Zisserman, A., 2005. A statistical approach to texture classification from single images. *International Journal of Computer Vision* 62, pp. 61–81.
- Zhu, S.-C., Guo, C.-E., Wang, Y. and Xu, Z., 2005. Textons, the elements of texture perception and their interactions. *International Journal of Computer Vision - Special Issue on Texture Analysis and Synthesis* 62, pp. 121–143.