Texton-based Texture Classification

Laurens van der Maaten [a] Eric Postma [a]

[a] *MICC, Maastricht University*
P.O. Box 616, 6200 MD Maastricht, The Netherlands

Abstract

Over the last decade, several studies on texture analysis propose to model texture as a probabilistic process that generates small texture patches. In these studies, texture is represented by means of a frequency histogram that measures how often texture patches from a codebook occur in the texture. In the codebook, the texture patches are represented by, e.g., a collection of filter bank responses. The resulting representations are called textons. A recent study claims that textons based on normalized grayvalues outperform textons based on filter responses (such as MR8 filter responses), despite the weaknesses of such image-based representations for image modelling. The paper investigates this claim by comparing image-based textons with textons obtained using the complex wavelet transform. The complex wavelet transform differs from the MR8 and similar filters in that it employs filters with relatively low support and in that it constructs image representations with less redundancy. Furthermore, the paper investigates to what extent image-based textons are susceptible to 2D rotations of the texture. It compares image-based textons with rotation-invariant textons based on spin images and polar Fourier features. The performance of the new types of textons is evaluated in classification experiments on the CUReT texture dataset. The results of our experiments with the complex wavelet transform support the claim that filter-based textons do not outperform their image-based counterparts. Furthermore, the results of our experiments reveal that image-based textons are very susceptible to 2D rotations of the texture, making image-based textons unapplicable to real-world texture classification problems. We show that strong rotation-invariant texton-based texture classifiers can be constructed by means of textons based on spin images and polar Fourier features.

1 Introduction

Texture analysis is a task in computer vision that aims at representing texture in a model that is invariant to changes in the visual appearance of the texture. The visual appearance of a single texture can change dramatically under the influence of, e.g., lighting changes and 3D rotations. Traditional approaches to texture analysis often model texture as a Markov Random Field [6], which is a model that is widely used in machine learning. Another very popular approach to texture modelling is to analyze statistics of the responses of large filter banks [7, 11]. A number of recent studies propose to model texture as a probabilistic process that generates small texture patches according to a probability distribution over all possible texture patches [2, 4, 10, 14, 15]. In these studies, texture classification is performed with the help of a texton codebook that is constructed by performing vector quantization on a set of randomly selected textons. Textons are the representations of small texture patches by, e.g., a collection of filter bank responses, and can be viewed upon as the textural counterpart of phonemes and graphemes. In texton-based texture classifiers, a texture is represented by means of a texton frequency histogram that measures the relative frequency by which textons from the codebook appear in the texture. Texture classification is performed by means of classifiers, such as nearest neighbor classifiers [14] or Support Vector Machines [2], that are trained on the texton frequency histograms. An application of such a texture classifier to defect detection is presented in [16].

In most studies, textons consist of a collection of filter bank responses [2, 4, 10, 14]. In [15], the authors claim that textons based on normalized grayvalues outperform textons based on filter bank responses. Their claim is supported by experiments in which such image-based textons are compared with textons based on responses of a filter bank specially designed for texture analysis (the so-called MR8 filter bank). An image-based texton is a concatenation of the normalized grayvalues in the texture patch. The results in [15] reveal the image-based textons to outperform textons based on the MR8 filter bank responses. This is a surprising result, because image-based representations are generally considered to be inappropriate for classification.

For instance, image-based representations are sensitive to noise and to changes in orientation and scale.
In the paper, we investigate the claim in [15] by comparing image-based textons with textons based on the complex wavelet transform [9]. We selected the complex wavelet transform because it has theoretical advantages over the MR8 filter bank that may mitigate the weaknesses of filter-based textons. Furthermore, we investigate to what extent image-based textons suffer from the presence of 2D rotations in the texture images and propose two new (rotation-invariant) textons based on spin images and on polar Fourier features. We investigate the performance of the three new textons in two experiments on the CUReT texture dataset. The results of our experiments support the claim that filter-based textons do not outperform their image-based counterparts, and show that our rotation-invariant textons are capable of overcoming the susceptibility to the presence of 2D rotations of texton-based texture classifiers.
The outline of the remainder of the paper is as follows. In section 2, we discuss texton-based texture classifiers in more detail. Section 3 presents the three new texton representations that we investigated. Our experiments with the new texton representations are discussed in section 4. The results of the experiments are discussed further in section 5. In section 6, we present conclusions and directions for future work.

2 Texton-based texture classifiers

Texton-based texture classifiers classify textures based on their texton frequency histogram. A schematic overview of a texton-based texture classifier is given in Figure 1. The construction of texton-based texture classifiers consists of three main stages: (1) construction of a texton codebook, (2) computation of a texton frequency histogram, and (3) training of the classifier based on the texture frequency histograms. The first two stages are described in more detail below. The third stage can be implemented by means of, e.g., a nearest neighbor classifier [14] or a Support Vector Machine [2].

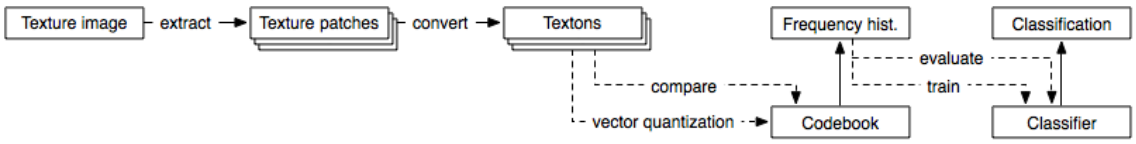

Figure 1: Schematic overview of the texton-based texture classifier.

2.1 Codebook construction

A texton codebook is a collection of textons that can be used to characterize texture images. In the construction of a texton codebook, small texture patches are extracted from random positions in the texture images. The pixel values of the texture patches are normalized in order to make them invariant to changes in lighting conditions. Textons are obtained by converting the texture patches to an appropriate image representation (such as a collection of filter bank responses or a concatenation of normalized pixel values). A part of a texton codebook is constructed by performing vector quantization (e.g., using k-means clustering or Kohonen maps) on a set of textons extracted from texture images that are assigned to a single texture class. A complete texton codebook is constructed by repeating this process for every texture class in the texture dataset. If the texture dataset is a proper subset of real-world textures, the texton codebook contains the most important textons that occur in real-world textures.

2.2 Texton frequency histogram

In texton-based texture classifiers, texture is viewed upon as a probabilistic generator of textons. The underlying probability distribution of the generator is estimated by means of a texton frequency histogram that measures the relative frequency of textons from the codebook in a texture image. A texton frequency histogram is constructed from a texture image by scanning over the texture image and extracting small texture patches. The small texture patches are converted to the image representation that is used in the codebook in order to obtain a collection of textons. Each extracted texton is compared to the textons in the codebook in order to identify the most similar texton from the codebook, and the texton frequency histogram bin corresponding to this texton is incremented. After normalization, the texton frequency histogram forms a feature vector that models the texture, and can be used in order to train a classifier.

2.3 Image-based textons

In most texton-based texture classifiers [2, 4, 10, 14], textons are represented by means of a collection of filter bank responses obtained from large filter banks such as the MR8 filter bank or the Leung-Malik filter bank[1]. However, in [15], it was shown that image-based textons outperform textons based on filter responses, leading to questions about the necessity of applying filter banks for the analysis of texture. Three main reasons for the relative strong performance of image-based textons were suggested by the authors. First, the use of filter banks reduces the number of textons that can be extracted from a texture image. This reduction is a consequence of the large support of filter banks; the number of patches that can be extracted from a, say, 200×200 pixel texture image is significantly reduced when this image is convolved with a 50×50 filter. The presence of a reduced number of textons affects the quality of the texton frequency histogram estimations, leading to inferior generalization performances. Second, the large support of filter banks leads to small errors in the localization of edges. Imprecise edge localization may significantly change the geometry of the textons, leading to errors in the estimation of the texton frequency histogram. Third, the application of most filters leads to some blurring on the texture images, which is the result of the Gaussian envelope in these filters. The blurring might remove local details in the textons, that are of interest in the classification of the texture.

The three main reasons why image-based textons are not expected to yield good classification performances are as follows. First, image-based textons do not contain information on the presence of different orientations in the texture. The measurement of edge orientations is known to be important in human vision [8]. Second, image-based textons are sensitive to the presence of noise in the image. Third, image-based textons are susceptible to the presence of 2D rotations in the texture images. In the next section, we propose three new types of textons that attempt to mitigate or resolve these weaknesses.

3 Three new texton representations

In this section, we propose three new texton representations: a representation based on the complex wavelet transform (subsection 3.1), a representation based on spin images (subsection 3.2), and a representation based on polar Fourier features (subsection 3.3). The texton representation based on the complex wavelet transform is proposed in order to further investigate the claim that image-based textons outperform textons based on filter bank responses. Because of its relatively low redundancy and small support, the complex wavelet transform may overcome the drawbacks of MR8 filter banks. The texton representations based on spin images and polar Fourier features are proposed in order to construct texton-based texture classifiers that are invariant to changes in the orientation of the texture images.

3.1 Complex wavelet transform

The wavelet transform expands a signal into a collection of frequency components (similar to the Fourier transform). Unlike the Fourier transform, the wavelet transform does so by using a collection of localized basis functions in order to resolve the Gibbs effect from which the Fourier transform suffers. In practice, the wavelet transform is implemented as a dyadic filter tree in which a low-pass filter g and a high-pass filter h are employed. Both filters are applied on the signal, the low-pass filter response is downsampled, both filters are applied on the result, and this process is iterated. If both filters meet certain requirements (such as orthogonality of the filters), the responses of the high-pass filters provide the wavelet coefficients. An extensive introduction on wavelet theory can be found in [3].

The complex wavelet transform (CWT) is capable of capturing more phase information than the traditional wavelet transform by the use of complex filters, and thereby, it provides approximate shift invariance to the wavelet transform [9]. The CWT is implemented by means of a dual dyadic filter tree, of which a one-dimensional version is shown schematically in Figure 2. In the figure, square boxes indicate a filtering with either the high-pass filter h_i or the low-pass filter g_i, and $\downarrow 2$ indicates a downsampling of the signal by 2.

In addition to the restrictions on the filters in the traditional wavelet transform, the filters in the two branches of the filter tree should form Hilbert pairs. In other words, filter g_1 should be the Hilbert transform of filter

[1]The MR8 filter bank consists of six orientation-sensitive filters at three different scales. The responses of the MR8 filter bank are given by the maximum responses of the complete filter bank over all scales. In addition to the orientation-sensitive filters, the MR8 filter bank contains a Gaussian and a Laplacian of Gaussian filter. The Leung-Malik filter bank consists of six orientation-sensitive filters, a Gaussian filter, and a Laplacian of Gaussian filter.

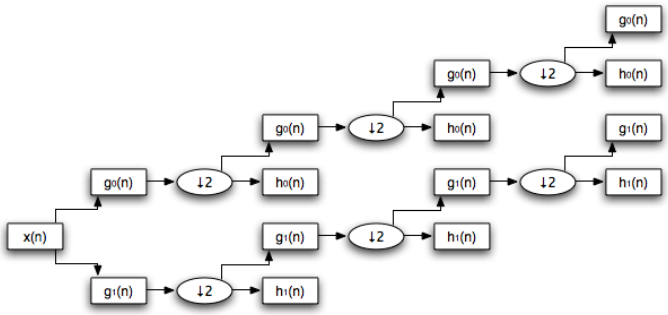

Figure 2: Complex wavelet transform filter tree.

g_0, and filter h_1 should be the Hilbert transform of filter h_0. If this requirement is met, the responses of the filters can be shown to complement each other, leading to a lower susceptibility to shifts in the signal. In the 2D case, the wavelets in the CWT show great resemblance to orientation-sensitive filters such as Gabor filters, as is illustrated in Figure 3. However, the CWT has three important advantages over Gabor (and related) filters. First, CWT coefficients are less redundant than Gabor wavelet coefficients (the 2D CWT is only four times complete), leading to an image representation of lower dimensionality that can be computed more efficiently. Second, the support of the filters that are used in the CWT is generally small, allowing for a better estimation of the texton generation distribution and for an additional computational advantage. Third, the CWT has an inverse transform, which implies that the original image can be reconstructed from the wavelet coefficients for, e.g., visualization purposes. In our texton-based texture classifiers that employ the CWT, we perform the complex wavelet transform on the normalized texture patches using the filters described in [1].

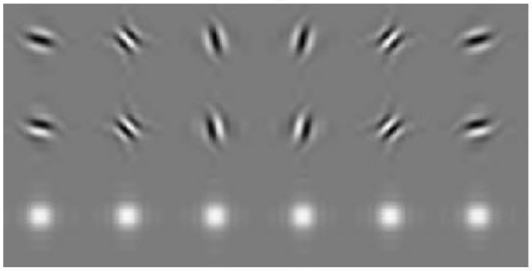

Figure 3: Wavelets corresponding to the complex wavelet transform. The upper row represents the real parts of the six wavelets, whereas the middle row represents the imaginary parts of the wavelets. The magnitude of the filters is depicted in the bottom row, revealing that the real and imaginary parts of the wavelets are complementary. The wavelets were obtained using the filters described in [1].

3.2 Spin images

Spin images estimate the joint intensity-radius distribution of an image in a coarse histogram [12]. In the construction of a spin image, the distance of every pixel to the center of the image (i.e., the radius) is computed. The radiuses and the corresponding pixel values are quantized and binned in a joint histogram. The construction of spin images is illustrated in Figure 4. The main advantage of the use of spin images is that they are invariant to changes in the orientation of the image.

In our texton-based texture classifiers, we construct spin images with 8 intensity bins from the normalized texture patches. The number of radius bins is set to the width (or height) of the texture patches in pixels. We estimate the texton frequency histograms using an overcomplete texture patch basis (i.e., there is overlap in the texture patches).

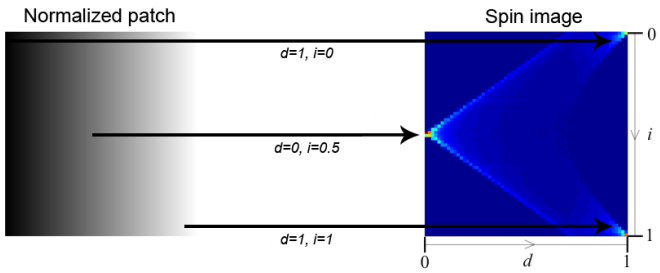

Figure 4: Example of the construction of a spin image.

3.3 Polar Fourier features

Polar Fourier features are based on the observation that the magnitude of the Fourier transform of a histogram is invariant under circular shifts, because all phase information is in the sign of the Fourier coefficients [13]. In the computation of polar Fourier features, the image is converted to polar space. In the polar space, one axis represents the distance to the center of the image, whereas the other axis represents the angle from the baseline (which is the horizontal line through the center of the image). As a result, a rotation of the original image leads to a circular shift in the 'distance-bands' of the polar image. The polar image is made rotation-invariant by computing the magnitude of the Fourier transform of every 'distance-band' in the polar image. The construction of polar Fourier features is illustrated in Figure 5.
In our texton-based texture classifiers using polar Fourier features, the texton frequency histograms are constructed using an overcomplete basis of normalized texture patches (i.e., there is overlap in the texture patches).

Figure 5: Example of the construction of polar Fourier features.

4 Experiments

In the previous section, we presented three new texton representations that may overcome some of the drawbacks of image-based textons. This section investigates the performance of the new texton representations on two texture classification tasks: (1) a task in which there are no 2D rotations on the textures, and (2) a task in which the classifier has to deal with 2D rotations of the textures. The setup of our experiments is described in subsection 4.1. Subsection 4.2 presents the results of our experiments.

4.1 Experimental setup

In order to evaluate the quality of the textons presented in the previous section, we performed texture classification experiments on the CUReT dataset [5]. The CUReT dataset contains images of 61 different materials that were photographed under 205 different viewpoints. The differences in viewpoints lead to a large variability in the visual appearance of the same material, as is illustrated in Figure 6. From the 205 images, we selected the 123 images that allow for the extraction of a texture image of 200×200 pixels. The extracted texture images are converted to grayscale images, since we are only interested in the textural information in the images. The texton codebooks were constructed by performing k-means clustering on 5,000 textons, that were obtained by random selection from the trainingset. In our experiments, we used a value of $k = 10$, leading to texton codebooks consisting of 610 textons. In the experiments with image-based textons and textons based on spin images, we performed experiments with texture patches of size 3×3 to 8×8 pixels.

The experiments with textons based on CWT features and polar Fourier features were performed only with texture patches of 4×4 and 8×8 pixels, because wavelet and Fourier transforms require a signal length that is a power of 2. In our experiments, the classification is performed by a 1-nearest neighbor classifier. The generalization performance of the classifiers is evaluated using 10-fold cross validation. Our experimental setup is roughly similar to the setup employed in [15].
We evaluated the performance of the trained classifiers in two settings: (1) a setting in which there are no changes in the orientations of the texture images and (2) a setting in which the texture classifier has to deal with 2D rotations of the texture images. In the latter setting, we train the classifiers on normal texture images, and test the classifiers on texture images that are rotated by 90 degrees in clockwise direction. We opt for a rotation of 90 degrees, because a rotation of 90 degrees does not introduce artefacts into the texture images.

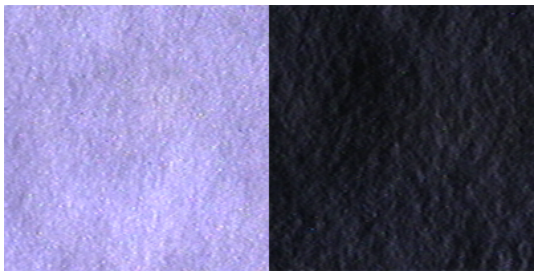

Figure 6: Visual appearance of a texture photographed under different lighting conditions.

4.2 Results

In the previous subsection, we described the setup of our experiments. This subsection presents the results of the experiments. The results are divided into two parts: results of experiments in which there are no changes in the texture orientations (subsubsection 4.2.1) and results of experiments in which the texture images in the test set have a different orientation (subsubsection 4.2.2).

4.2.1 Fixed-orientation texture classification

In Table 1, we present the generalization errors of 1-nearest neighbor classifiers that were trained on texton frequency histograms using the four texton representations that we discussed. The table presents generalization errors of classifiers trained on image-based textons, textons based on the complex wavelet transform, textons based on spin images, and textons based on polar Fourier features. In the experiments, the texture images were presented in their original orientation.
From the results presented in the table, we can make three observations. First, the results in the tables reveal that image-based textons perform very strong. Our results with image-based textons are comparable to those presented in [15]. The performance of CWT-based textons is comparable to those of image-based textons, and appears to be better than the performance of the textons based on MR8 filter bank responses presented in [14]. Second, we observe that the use of textons based on spin images degrades the performance of texton-based texture classifiers with approximately 3%. Third, we observe that textons based on polar Fourier features perform better than those based on spin images. Textons based on polar Fourier features almost perform comparable to image-based textons (especially for larger patch sizes).

Patch size	*Image*	*CWT*	*Spin image*	*Polar Fourier*
3×3	0.0264 ± 0.0053	–	0.0649 ± 0.0100	–
4×4	0.0206 ± 0.0064	0.0260 ± 0.0056	0.0546 ± 0.0073	0.0466 ± 0.0072
5×5	0.0204 ± 0.0062	–	0.0509 ± 0.0121	–
6×6	0.0177 ± 0.0044	–	0.0530 ± 0.0123	–
7×7	0.0195 ± 0.0057	–	0.0516 ± 0.0102	–
8×8	0.0187 ± 0.0051	0.0179 ± 0.0038	0.0530 ± 0.0085	0.0243 ± 0.0070

Table 1: Generalization errors of texton-based texture classifiers (fixed orientation).

Patch size	*Image*	*CWT*	*Spin image*	*Polar Fourier*
3×3	0.5369 ± 0.0143	–	0.0622 ± 0.0099	–
4×4	0.5597 ± 0.0119	0.6330 ± 0.0151	0.0541 ± 0.0082	0.0630 ± 0.0121
5×5	0.6104 ± 0.0153	–	0.0540 ± 0.0061	–
6×6	0.6512 ± 0.0050	–	0.0543 ± 0.0093	–
7×7	0.6778 ± 0.0144	–	0.0523 ± 0.0062	–
8×8	0.6971 ± 0.0179	0.6552 ± 0.0216	0.0520 ± 0.0107	0.0253 ± 0.0062

Table 2: Generalization errors of texton-based texture classifiers (variable orientation).

4.2.2 Variable-orientation texture classification

In Table 2, we present the results of 1-nearest neighbor classifiers that were trained on normal texture images, but tested on texture images that were rotated 90 degrees in a clockwise direction. The table presents the generalization errors of classifiers based on the four texton types.
From the results in Table 2, we observe that the use of textons that are not rotation-invariant strongly degrades the performance of texton-based texture classifiers when 2D rotations are present in the texture images. The presence of 2D rotations in the texture images causes an increase of the generalization error by 50% to 70%. The degradation in the performance of the classifier increases with the size of textons that are used. In constrast to image-based and CWT-based textons, the rotation-invariant textons do not suffer from the presence of 2D rotations in the test data.

5 Discussion

In the previous section, we presented the results of our experiments with three new texton representations on the CUReT dataset. In this section, we discuss the two main observations that can be made from the results of our experiments.
First, the performance of textons based on the complex wavelet transform supports the claim in [15] that the popular use of filter responses in texture classification is debatable. We surmise that the main disadvantage of the filter-based textons in [15] is the reduction of the number of textons (due to the large support of the filters). Since our textons based on the complex wavelet transform do not suffer from this weakness, they outperform the MR8-based textons. An additional advantage of our CWT-based textons is the low redundancy of the complex wavelet transform, which reduces the dimensionality of the textons. Despite the advantages of the complex wavelet transform, CWT-based textons do not outperform image-based textons, which indicates that imprecise edge localization is an important problem in filter-based textons.
Second, the results of our experiments show that current texton-based texture classifiers are very sensitive to the presence of 2D rotations in the texture images. The results show that our rotation-invariant textons might degrade the accuracy of the classifier somewhat, but that this accuracy degradation is very limited. Polar Fourier features almost perform comparable to their image-based counterparts that are sensitive to changes in the orientation of the texture. A disadvantage of our rotation-invariant texture classifiers is that they model each 'distance band' separately in a rotation-invariant model. As a result, information on the alignment of the distance bands is lost. Our rotation-invariant texture classifiers may be improved by incorporating alignment information in the rotation-invariant features. Most likely, our texture classifiers are also very sensitive to the presence of scale changes in the texture images[2].

6 Conclusions

Texton-based texture classifiers form a new alternative to traditional texture classification approaches such as Markov Random Fields or filter bank models. Recent work on texton-based texture classifiers suggests the use of image-based textons over filter-based textons. We investigated this claim by performing experiments with textons based on the complex wavelet transform, which has a number of theoretical advantages over

[2]The reader should note that the texton-based texture classifiers are invariant to the presence of 3D rotations, because the training data contains textures that are photographed under a large number of 3D rotations.

large filter banks (such as the MR8 filter bank). The results of our experiments support the claim that the popular use of filter bank responses in texture classification is debatable, although the image-based textons did not outperform our CWT-based textons.

One of the main disadvantages of current texton-based texture classifiers is that they are very sensitive to changes in the orientation of the texture. We empirically evaluated this sensitivity, and encountered performance drops of more than 50% under the influence of 2D rotations in the texture images. In order to resolve this problem, we proposed the use of rotation-invariant textons based on spin images and polar Fourier features. Our experiments show that rotation-invariant texture classifiers based on spin images and polar Fourier features perform strongly. The results show that the performance of textons based on polar Fourier features is almost similar to the performance of their image-based counterparts.

Future work focuses on the development of texton-based texture classifiers that are not only invariant to changes in orientation, but also to changes in scale. Invariance to scale changes could be obtained by constructing texton frequency histogram for a number of scales in the scale space, and aligning the resulting models by maximizing the frequency histogram similarities. The incorporation of scale invariance into our texton-based texture models would lead to texture models that are invariant to all main variations in texture images. In addition, texture classification results could be improved by employing mixture models instead of a texton codebook, or by the use of multi-scale textons.

References

[1] A.F. Abdelnour and I.W. Selesnick. Design of 2-band orthogonal near-symmetric CQF. In *Proceedings of the IEEE International Conference Acoustic, Speech, and Signal Processing*, volume 6, pages 3693–3696, 2001.

[2] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1597–1604, 2005.

[3] C.K. Chui. *An Introduction to Wavelets*. Elsevier, 1992.

[4] O.G. Cula and K.J. Dana. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1):33–60, 2004.

[5] K.J. Dana, B. van Ginneken, S.K. Nayar, and J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, 1999.

[6] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1039–1046, 1999.

[7] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 14–19, 1990.

[8] J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[9] N.G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Journal of Applied and Computational Harmonic Analysis*, 10(3):234–253, 2001.

[10] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[11] J. Portilla and E.P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–71, 2000.

[12] C. Schmid, S. Lazebnik, and J. Ponce. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278, 2004.

[13] T. Szoplik and H.H. Arsenault. Rotation-variant optical data processing using the 2D nonsymmetric Fourier transform. *Applied Optics*, 24(2):168–172, 1985.

[14] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the* 7^{th} *European Conference on Computer Vision*, volume 3, pages 255–271, 2002.

[15] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 691–698, 2003.

[16] X. Xie and M. Mirmehdi. TEXEMS: Texture exemplars for defect detection on random textured surfaces. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 29(8):1454–1464, 2007.