Pattern Recognition Letters 30 (2009) 306–313

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Material-specific adaptation of color invariant features

Gertjan J. Burghouts [a,*], Jan-Mark Geusebroek [b]

[a] TNO Observation Systems, Electro Optics, Oude Waalsdorperweg 63, 2597 AK, The Hague, The Netherlands
[b] Intelligent Systems Lab Amsterdam, Informatics Institute, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 30 September 2007
Received in revised form 30 June 2008
Available online 18 October 2008

Communicated by H.H.S. Ip

Keywords:
Image modeling
Color
Codebook representation
Texture
Textons

ABSTRACT

For the modeling of materials, the mapping of image features onto a codebook of feature representatives receives extensive treatment. For reason of their generality and simplicity, filterbank outputs are commonly used as features. The MR8 filterbank of Varma and Zisserman is performing well in a recent evaluation. In this paper, we construct color invariant filter sets from the original MR8 filterbank. We evaluate several color invariant alternatives over more than 250 real-world materials recorded under a variety of imaging conditions including clutter. Our contribution is a material recognition framework that learns automatically for each material specifically the most discriminative filterbank combination and corresponding degree of color invariance. For a large set of materials each with different physical properties, we demonstrate the material-specific filterbank models to be preferred over models with fixed filterbanks.

1. Introduction

The appearance of materials change significantly under different imaging settings, depending on the settings themselves (Dana et al., 1999) and also on the physical properties of a material (Koenderink et al., 1999). Hence, materials-specific image representations may improve on the recognition performance, as they capture properties that are distinctive to the material and are balanced with the variation of imaging settings. For instance, for one material, the local intensity variation is a distinctive property, while the other is distinguished best from other materials based on its color properties. Fig. 1 depicts some materials from the ALOT dataset (Geusebroek and Smeulders, xxxx) and the testing conditions (Fig. 2). The first and second material are distinguished best when comparing their colors, more specifically, the red channel. For the third and fourth material, the most discriminative feature is the amount of intensity edges, while the fifth image in the first row and the second image in the third row are distinguished best when comparing the information in the green channel. These examples illustrate the advantage of material-specific representations. The objective in this chapter is to learn material-specific representations for more than 250 materials.

For material recognition (Varma and Zisserman, 2005; Leung and Malik, 2001) and classification (Hayman et al., 2004), but also for object and scene classification (Zhang et al., in press), the mapping of image features onto a codebook of feature representatives (Jurie and Triggs, 2005; Nowak et al., 2006) has received extensive treatment. Commonly used features are the class of SIFT-based features (Lowe, 2004; Mikolajczyk and Schmid, 2005; Burghouts and Geusebroek, 2008), see e.g. (Lazebnik et al., 2005). Alternatively, filterbank outputs are in use as features. Promising methods that use filterbanks to model object and scenes, have been proposed by Winn et al. (2005) and by Shotton et al. (2006).

In previous work by the authors Burghouts and Geusebroek (2006), image edges were filtered by a filterbank and subsequently annotated by their color improving the discriminative power of filterbanks further. The objective of Burghouts and Geusebroek (2006) was to extend a method that was originally proposed for grey-value images to include color information. The extension works well for images with many edges, and we do not expect it to work for more general images. Furthermore, the purpose of this paper is broader: we will integrate various ways of measuring color by filterbanks. We consider filterbanks for reason of their discriminative power, simplicity and generality.

To adapt the representation to a particular material, we consider various ways to represent an image. To that end, consider various intensity and color filterbanks. They are adapted from the MR8-filterbank which performs well in a recent evaluation (Varma and Zisserman, 2005). Each of the filterbanks measures different color channels, and each achieves a different degree of photometric invariance. We adopt techniques from the literature on invariant feature design, see e.g. (Finlayson et al., 1994; Gevers and Smeulders, 1999; Geusebroek et al., 2001). The general scheme to construct a representation of a filtered image, typically a histogram,

* Corresponding author. Tel.: +31 64 545 82 93; fax: +31 70 374 06 53.
E-mail addresses: gertjan.burghouts@tno.nl (G.J. Burghouts), mark@science.uva.nl (J.-M. Geusebroek).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.10.005

Fig. 1. Example materials from the ALOT dataset (Geusebroek and Smeulders, xxxx).

Fig. 2. Test images for the ALOT material depicted above on third row, first column.

is to first establish representatives of the filter outputs, or textons (Leung and Malik, 2001). A standard solution that aims to minimize the average reconstruction error is the *k*-means algorithm, employed originally by Sivic and Zisserman (2003) and Csurka et al. (2004). Alternatively, Winn et al. (2005) employed an information–maximization approach. For any of these approaches to establish textons, the problem is how to arrive at a representation that is specific to the material at hand.

A recent method proposed by Perronnin et al. (2006) establishes class-specific textons for each of N classes. As the authors point out, the straightforward accumulation of all textons into one large codebook is not feasible, as the learning of materials will be hampered as a result of the large histograms representing the images (curse of dimensionality). To avoid this problem, they suggest to use the N sets of class-specific textons to create respectively N codebooks. Each image is subsequently represented by the N codebooks resulting in N histograms. Elegantly, for each image the N histograms are fed to N class-specific classifiers. Classification of the image is based on the N thus obtained posterior probabilities. In (Perronnin et al., 2006), high performance is reported for the classification of 7–10 categories. However, for the classification of more than 250 materials, the method in (Perronnin et al., 2006) will be hampered by the creation of more than 250 histograms for each image. With 24 images per class, over $24 \cdot 250 \cdot 250$ histograms need to be constructed, which is not feasible in practice. Rather, we will propose a scalable alternative to construct material-specific representations, by representing the image by $M \ll N$ histograms. The M histograms are obtained from M color invariant codebooks, each learned from one filterbank with specific color and invariant properties. As a result, the class-specificity of codebooks is not in the learned textons, but in their color and invariance properties.

The paper is organized as follows: In Section 2, the MR8 filterbank and its color invariant versions are introduced. We propose the framework to learn material-specific color information and invariance in Section 3. In Section 4, we evaluate first the performance of the intensity and color filterbanks on the CURET (Dana et al., 1999) and ALOT datasets based on discriminative power, invariance to image settings and clutter. Second, we evaluate the framework to adapt the use of filterbanks to the material. Conclusions are drawn in Section 5.

2. Color invariant filterbanks

In this section, we introduce the MR8 filterbank and several extensions to color. The MR8 filterbank is shown in Fig. 3a. Typically, before the image is convolved with the MR8 filterbank, the image is normalized to zero mean and unit variance, to achieve invariance to imaging conditions, see e.g. (Varma and Zisserman, 2005). In the following subsections, we extend the MR8 filterbank to incorporate color information, and we consider various transformations to achieve color invariance from literature.

2.1. MR8-NC

In a first modification of the MR8 filterbank to extend it to use color information, we apply the filterbank to the image's color channels directly. This is a straightforward extension that is also employed by Winn et al. (2005), who have applied the MR8 filterbank to *Lab* color values. We largely follow (Winn et al., 2005) here. However, we restrain to a linear subspace of RGB, and apply the filterbank to the three opponent color channels of the image. Opponent colors have the advantage that the color channels are largely decorrelated. Here, we consider the Gaussian opponent color model, which is computed from RGB values directly by Geusebroek et al. (2001)

$$\begin{bmatrix} \widehat{E}(x,y) \\ \widehat{E}_\lambda(x,y) \\ \widehat{E}_{\lambda\lambda}(x,y) \end{bmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{pmatrix} \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix}, \quad (1)$$

where $\widehat{E}$, $\widehat{E}_\lambda$ and $\widehat{E}_{\lambda\lambda}$ denote the intensity, blue–yellow and green–red channel.

Likewise the usage of the MR8 filterbank in the VZ algorithm (Varma and Zisserman, 2005), we normalize each of the color channels $\widehat{E}$, $\widehat{E}_\lambda$ and $\widehat{E}_{\lambda\lambda}$, to zero mean and unit variance

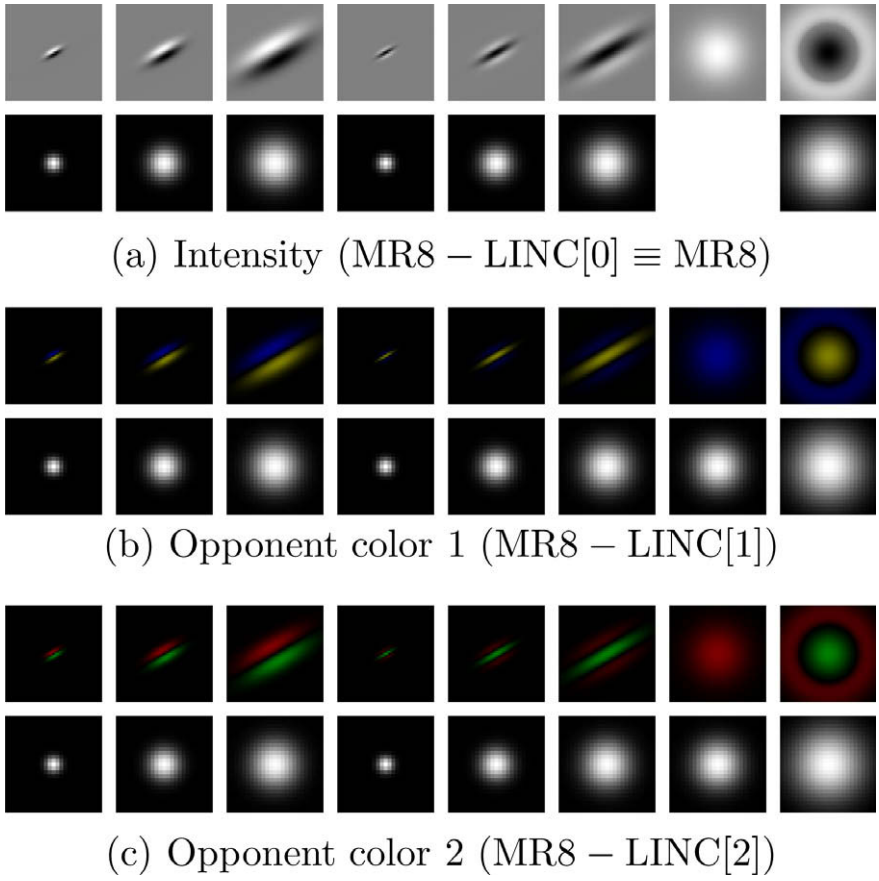


Fig. 3. MR8-LINC: a color invariant filterbank. The original MR8-filterbank (a – top row) is convolved with each of the image's opponent colors channels (a–c – upper rows), to yield 24 responses per pixel. Each of the 24 filter outputs is normalized by the local intensity as is measured by a Gaussian kernel of the same size of the MR8 filter (a-c – lower rows). The only MR8 filter that is not normalized is the Gaussian kernel that measures intensity (otherwise it would yield a constant output). The normalization achieves invariance to local intensity changes. (The figure is best viewed in color.)

$$\widehat{E}' = \frac{\widehat{E} - \mu_{\widehat{E}}}{\sigma_{\widehat{E}}}, \quad \widehat{E}'_{\lambda} = \frac{\widehat{E}_{\lambda} - \mu_{\widehat{E}_{\lambda}}}{\sigma_{\widehat{E}_{\lambda}}}, \quad \widehat{E}'_{\lambda\lambda} = \frac{\widehat{E}_{\lambda\lambda} - \mu_{\widehat{E}_{\lambda\lambda}}}{\sigma_{\widehat{E}_{\lambda\lambda}}}, \tag{2}$$

where $\mu_{\widehat{E}}$, $\sigma_{\widehat{E}}$ denote the mean and standard deviation of the intensity channel, respectively; $\mu_{\widehat{E}_{\lambda}}$, $\sigma_{\widehat{E}_{\lambda}}$ denote the mean and standard deviation of the blue-yellow opponent color channel, respectively; and, equivalently, $\mu_{\widehat{E}_{\lambda\lambda}}$, $\sigma_{\widehat{E}_{\lambda\lambda}}$ denote the mean and standard deviation of the green–red opponent color channel, respectively.

Next, each of the normalized color channels $\widehat{E}'$, $\widehat{E}'_{\lambda}$, and $\widehat{E}'_{\lambda\lambda}$ is convolved with the MR8 filterbank, yielding 24 filter outputs per pixel. This first extension of the MR8 filterbank is termed MR8 with normalized colors, or MR8-NC.

2.2. MR8-INC

In a second modification, we normalize the color channels such that they maintain more color information than is the case with MR8-NC. With MR8-NC, the means of the yellow–blue and red–green channels are normalized to zero, effectively discarding the actual chromaticity in the image, and only considering the variation. The color channels will be affected mainly by the lighting direction relative to the object and to the camera (Suen and Healey, 2000), which are mostly characterized by intensity fluctuations. Hence, we propose to normalize the three opponent color channels only by the standard deviation of the intensity. Normalizing the intensity channel by the standard deviation of intensity

$$\widehat{E}' = \frac{\widehat{E} - \mu_{\widehat{E}}}{\sigma_{\widehat{E}}}, \tag{3}$$

sets the variance of this channel to unity. Here, $\mu_{\widehat{E}}$ and $\sigma_{\widehat{E}}$ indicate the mean and standard deviation of the intensity channel as before. Normalizing the yellow–blue and red–green channels also by the intensity standard deviation

$$\widehat{E}'_{\lambda} = \frac{\widehat{E}_{\lambda}}{\sigma_{\widehat{E}}}, \quad \widehat{E}'_{\lambda\lambda} = \frac{\widehat{E}_{\lambda\lambda}}{\sigma_{\widehat{E}}}, \tag{4}$$

yields a more stable responses when the intensity variation fluctuates as a consequence of lighting or viewpoint changes. At the same time, it maintains information about the chromaticity in the image. Likewise MR8-NC, each of the normalized color channels is convolved with the MR8 filterbank, yielding 24 filter outputs per pixel. We refer to this filterbank as MR8 with intensity-normalized colors, or MR8-INC.

2.3. MR8-LINC

In a third modification, we modify the MR8-filterbank to achieve invariance to local intensity changes by a local color normalization rather than a global one. We follow closely the invariant Gaussian features developed in (Geusebroek et al., 2001). In (Geusebroek et al., 2001), each of the local image measurements is normalized by the intensity in a small neighborhood. This achieves invariance to the local intensity level.

We propose to filter for each pixel the non-normalized opponent color values using the MR8-filterbank, to obtain 24 filter outputs per pixel. Also, for each pixel, we measure the local intensity with a Gaussian kernel at the same scale as the MR8 filter under consideration. Per pixel, we normalize each output of the MR8 filterbank by the local intensity as measured by that Gaussian filter, yielding the transformed filter responses MR8′

$$\mathrm{MR8}'\left(\widehat{E}\right)=\frac{\mathrm{MR8}(\widehat{E})}{\widehat{E}^{\sigma}},\quad \mathrm{MR8}'\left(\widehat{E_{\lambda}}\right)=\frac{\mathrm{MR8}(\widehat{E_{\lambda}})}{\widehat{E}^{\sigma}},$$
$$\mathrm{MR8}'\left(\widehat{E_{\lambda\lambda}}\right)=\frac{\mathrm{MR8}(\widehat{E_{\lambda\lambda}})}{\widehat{E}^{\sigma}}, \tag{5}$$

where $\mathrm{MR8}(\cdot)$ indicates the successive application of a filter from the filterbank, and $\widehat{E}^{\sigma}$ represents the intensity image smoothed at the same spatial scale as the filter of MR8 under consideration, see Fig. 3. Obviously, the zeroth order Gaussian filter from the MR8-filterbank is not normalized by the local intensity, otherwise its output would be constant. We refer to this color filterbank as MR8 with local intensity-normalized colors, or `MR8-LINC`.

2.4. MR8-SLINC

Finally, we construct a shadow and shading invariant filterbank, termed `MR8-SLINC`. Similar to `MR8-LINC`, the invariance is achieved locally. With `MR8-LINC`, first the filterbank outputs are computed before normalization by the local intensity. Alternatively, the color values $\widehat{E}_{\lambda}(x,y)$ and $\widehat{E}_{\lambda\lambda}(x,y)$ can be normalized locally first before filtering the thus obtained images

$$\mathrm{MR8}'(\widehat{E})=\frac{\mathrm{MR8}(\widehat{E})}{\widehat{E}^{\sigma}},\quad \mathrm{MR8}'(\widehat{E_{\lambda}})=\mathrm{MR8}\left(\frac{\widehat{E_{\lambda}}}{\widehat{E}}\right),$$
$$\mathrm{MR8}'(\widehat{E_{\lambda\lambda}})=\mathrm{MR8}\left(\frac{\widehat{E_{\lambda\lambda}}}{\widehat{E}}\right). \tag{6}$$

Under Lambertian reflection, the normalization of color values by the local intensity results in color values independent of the intensity distribution. Hence, the filterbank outputs of `MR8-SLINC` are invariant to shadow and shading.

2.5. Filterbank properties

Similar to `MR8`, the color-based filterbanks `MR8-NC` and `MR8-INC` involve a global color normalization. In other words, the normalization is dependent on the contents of the image. Hence, clutter will affect the normalization. This makes the output of `MR8-NC` and `MR8-INC` *scene-dependent*. In contrast, the local normalizations that are employed in `MR8-LINC` and `MR8-SLINC` are not scene-dependent, but only *locally dependent* on the actual color values.

Further, the filterbanks can be ordered by their degree of invariance. `MR8-SLINC` is most invariant as its color channels aim to discard intensity variation. `MR8` and `MR8-NC` retain respectively the intensity and color variation, but they discard their mean and variance. `MR8-LINC` retains more of the intensity and color variations, as it discards locally the variance due to intensity fluctuations. Finally, `MR8-INC` is less invariant than `MR8-LINC`, as it discards only the global variance due to intensity fluctuations.

3. Color invariant codebooks and material-specific adaptation

In this section, we consider the construction of color invariant codebooks from the several filterbanks, and the methodology to apply the codebooks in a material-specific setting. First, we formalize the color invariant filterbanks as follows: `MR8-X` = {`MR8-X[0]`, `MR8-X[1]`, `MR8-X[2]`}, where `X` ∈ {`NC`, `INC`, `LINC`, `SLINC`}. To avoid the joint learning of color channels, we learn one codebook for each color channel `MR8-X`[i], with $i \in \{0,1,2\}$. For codebook construction, we follow the common scheme of learning textons by k-means clustering of filterbank outputs (Leung and Malik, 2001, 2003, 2004, 2005). We consider a single set of 20 images randomly drawn from the learning set of material images. Each is filtered by one of the filterbanks `MR8-X`[i], and from each filtered image we store 10 cluster centers. As a result, for each filterbank `MR8-X`[i], we obtain a codebook of 200 textons. For the filterbank `MR8-X`, we have obtained 3 codebooks of length 200. For fair comparison with the single-channel `MR8` filterbank, the length of the `MR8` codebook is increased to 600 by storing 30 instead of 10 cluster centers per learning image.

To represent an image in terms of codebooks, it is filtered by each of the color channel filterbanks `MR8-X`[i] first, before mapping the filter outputs onto the corresponding codebook and counting the most similar occurrences. For each `MR8-X`[i], a histogram of length 200 is obtained; hence for `MR8-X` three histograms are obtained. After concatenation of the histograms per color channel, a histogram of length 600 is obtained that corresponds to the filterbank `MR8-X`. The codebook representation is outlined in Fig. 4.

3.1. Material-specific adaptation

The limitation of the color codebook representation as proposed above, is that the discriminative power of the color channels is averaged by using a single histogram comparison measure. For instance, the intensity information may be less distinctive for a given material than is the color information. The averaging of the information in the color channels may lead to incorrect classification of materials. The misclassification of an image of the blueish material, mistakenly considered to be more similar to the pink material, is illustrated in Fig. 5a.

To overcome the limited resolving power of the direct combination of the three color channels, we start with classification of a material at the level of individual color channels and to give preference to a distinctive combination thereof. Fig. 5b illustrates that the blueish material is well separated from the pink material using the information in the third color channel.

We propose to train one classifier per color channel per filterbank to discriminate one material from all other materials. Hence, with I filterbanks, $F_{1,\ldots,I}$, and J color channels, $c_{1,\ldots,J}$, we obtain $I \times J$

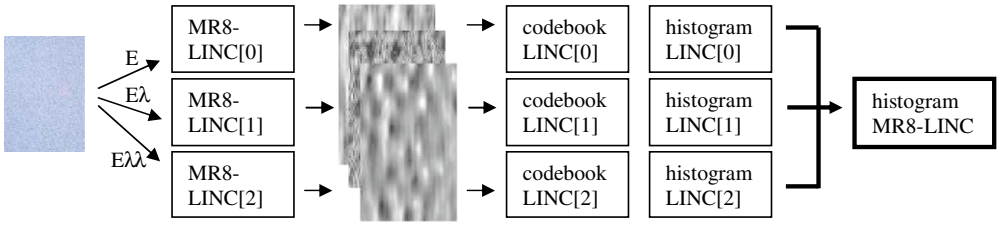
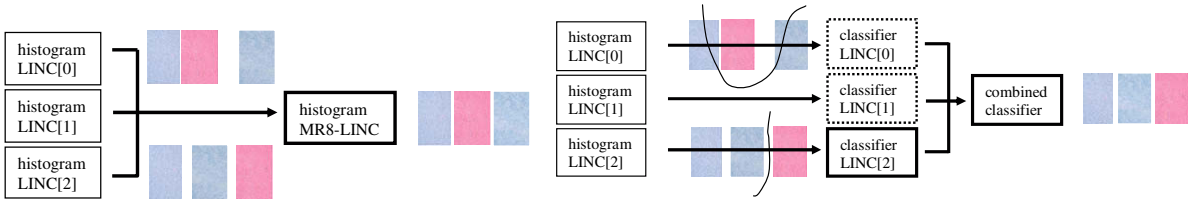


Fig. 4. Color codebook approach where the three color channels are separately filtered and represented by a histogram. Subsequently, the histograms are combined into one.



(a) Fixed combination scheme; resulting ordering on the right

(b) Adaptive combination scheme; resulting ordering on the right

Fig. 5. Separation of two images of the same material from one image of an other material. The fixed representation in (a) is not able to distinguish correctly between the two, while the material-specific representation is able to distinguish between the two (third color channel). (The figure is best viewed in color.)

classifiers. With N materials, each classifier outputs N posterior probabilities. With this procedure, $I \times J \times N$ values are produced by the first classifier stage.

In the combination stage, one classifier is trained using the $I \times J \times N$ values obtained for each material image. This one versus all classifier learns per material the discriminant function from the posterior probabilities assigned to each material by the individual classifiers. As a result, the combined classifier learns *implicitly* the filterbank and color channel that is most distinctive for the specific material. To infer *explicitly* from the material-specific discriminant function provides information which filterbank and color representation combination is most distinctive for a given material, we determine for each material which of the individual classifier's outputs approximates the normal to the discriminant function of the combining classifier best. This measure indicates the importance of a particular filterbank for the classification of the given material.

4. Experiments

In the experiments, we evaluate the color filterbanks and their combination. We take two datasets into account to cover a wide range of real-world materials and imaging conditions under which they can be viewed. First, we consider the well-known CURET dataset (Dana et al., 1999). This dataset enables one to test the robustness under varying imaging conditions, i.e. changes of the illumination direction and of the camera viewpoint. For color-based methods, a critical issue is whether the method is robust to color transformations in the image as a consequence of varying illumination color. Second, we consider the ALOT dataset (Geusebroek and Smeulders, xxxx) to also include variations of the illumination color. Additionally, this dataset contains more color and 3D variation. Some of the materials that are included in the ALOT dataset are illustrated in Fig. 1, while some test images are shown in Fig. 2. In total, we evaluate the filterbanks on 61 textures of the CURET dataset and on 200 textures of the ALOT dataset. In the experiments we use in total 5612 CURET images and 7200 ALOT images, respectively. For CURET, the train, test and texton learn sets are mentioned in (Varma and Zisserman, 2005); for ALOT the sets are publicly available on the website of the ALOT database.

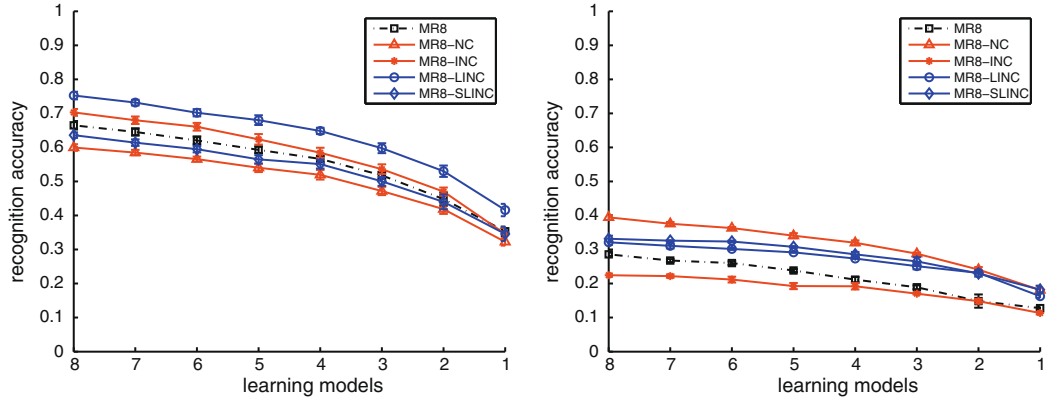
In the experiments, the number of textons is always set to 200 (as shown in (Varma and Zisserman, 2005) this parameter does not affect the results significantly). For the individual and combined classifiers, we prefer respectively the nearest mean classifier (Euclidean distance) and the linear Bayes-normal classifier (Duda et al., 2000), as these are performing best.

4.1. Color invariant codebooks

4.1.1. Random images

We start the performance evaluation by establishing the classification accuracy when selecting randomly the learning images. This experiment gives an indication of the discriminative power and robustness of each of the color filterbanks. We include the original `MR8` as a baseline comparison. We consider the mean and standard deviation of classification accuracy over 1000 repetitions (random selections).

Fig. 6a and b shows the recognition results for the CURET and ALOT datasets, respectively. First, we discuss the results for the CURET dataset. The filterbanks with most invariant properties, `MR8`, `MR8-NC` and `MR8-SLINC` filterbanks perform less than the less



(a) CURET

(b) ALOT

Fig. 6. Accuracy of material recognition for various filterbanks with randomly selected images of (a) the CURET dataset and (b) the ALOT dataset. The vertical bars indicate standard deviation over 1000 repetitions (best viewed in color).

invariant MR8-INC and MR8-LINC filterbanks. MR8 performs somewhat better than MR8-NC and MR8-SLINC, as it's nearest mean classifier puts all emphasis on the intensity information. With MR8-NC and MR8-SLINC, emphasis of the nearest mean classifier is also put on the color channels, of which almost all information is lost due to the normalization of the mean and variance. The MR8-LINC filterbank performs better than does MR8-INC, as it provides a better approximation of the changing intensity effects by doing so locally.

As expected, for ALOT the performance of the filterbanks is different, as this dataset contains more color and 3D variation. The severe 3D variations causes the intensity to change in such a way that it cannot be approximated well globally. This explains the low performance of the MR8-INC filterbank. At the same time, with much more colorful materials, the global normalization of image colors makes sense: local color variations in the image are now kept albeit relative to each other. Also, the severe 3D variations across materials causes their appearance to change significantly with different illumination. Keeping color variations while being very invariant, explains the good performance of the MR8-NC filterbank. The MR8-INC and MR8-LINC filterbanks are less invariant, hence they perform somewhat less than MR8-NC. The distinctive color information maintained by MR8-INC and MR8-LINC explains their better performance compared to the MR8 filterbank.

4.1.2. Cluttered images

Robustness to clutter is of importance for image modelling where the image frame is not fixed, or/and where no image segmentation is available. The setup of the previous experiments involves images that contain no clutter, as the image frames are fixed and each image captures one material only. In this experiment, we evaluate the sensitivity of the color-based filterbanks MR8, MR8-NC, MR8-INC and MR8-LINC to clutter.

First, we randomly select one learning image for each texture. Second, we simulate clutter by concatenating the learning image with a randomly selected image of another texture. For the first cluttered test image, the percentage of original vs. clutter is 90% vs. 10%. To simulate various degrees of clutter, we increase the clutter percentage, up to 40% (note: with 50%, the classification would become chance). The cluttered images are publicly available on the website of the ALOT (Geusebroek and Smeulders, xxxx) database. Obviously, for generalization purposes, we use the texton dictionary from the previous experiment (i.e. we do not learn new textons from cluttered images).

Fig. 7a and b shows the results for increasingly cluttered images of the CURET and ALOT datasets, respectively. The MR8-LINC filterbank performs significantly better than the other filterbanks, MR8, MR8-NC, and MR8-INC, over various degrees of clutter. The low performance of MR8, MR8-NC, and MR8-INC is due to the global normalization schemes that they employ. A global normalization is distorted by clutter, so the filterbank input is different when dealing with variations of clutter. The local normalization employed in MR8-LINC is not distorted by clutter. The small performance drop here is due to ambiguity in the images themselves as a result of the cluttering. However, even with 40% clutter, the MR8-LINC filterbank achieves a classification accuracy of 75.5% on the ALOT dataset, while the runner-up (MR8-LINC) has an accuracy of 39.0% only.

The results of individual filterbanks are summarized as follows. From the previous two experiments, we conclude that the locally-invariant MR8-LINC and MR8-SLINC filterbanks are very robust to clutter, and that they perform well on different datasets. The MR8-LINC is performing best on the CURET dataset (limited 3D variation), whereas MR8-SLINC performs second-best on the ALOT dataset (severe 3D variation).

4.2. Adaptive color invariant codebooks

Since MR8-LINC and MR8-SLINC perform well but on different datasets, and given that the datasets contain very different types of materials, we establish in this experiment whether the tuning of each of the filterbanks to a particular material is beneficial.

As expected, Fig. 8a and c indicate that the classification accuracy is increased by combining the MR8-LINC and MR8-SLINC filterbanks. While the classification accuracy of MR8-LINC is almost saturated for the CURET dataset, 0.96, the combination achieves a marginal improvement, 2%. For the ALOT dataset, the performance is increased from 0.35 to 0.42 achieving an improvement of 19.8%.

Indeed, as laid down in Fig. 8b and d, the most distinctive filterbank per material varies significantly across the datasets, and also across the individual materials. The CURET dataset contains many materials of which the structure is similar. Hence, the intensity variation, although very discriminative (see previous experiments), is not *most* discriminative. Rather, color information is most discriminative, as the color channels of the filterbanks are often most distinctive. The information in the filterbanks that are not invariant to shadow and shading, MR8-LINC, is in 56% most distinctive. Most CURET materials are uni-colored, hence the color information is distinctive. With uni-colored materials, too much information is lost when discarding shadow and shading variation. Hence, the shadow and shading invariant filterbank MR8-SLINC is in less cases, 27%, most distinctive.

For the ALOT dataset, the performance improvement due to filterbank tuning is significant. As this dataset contains more variation

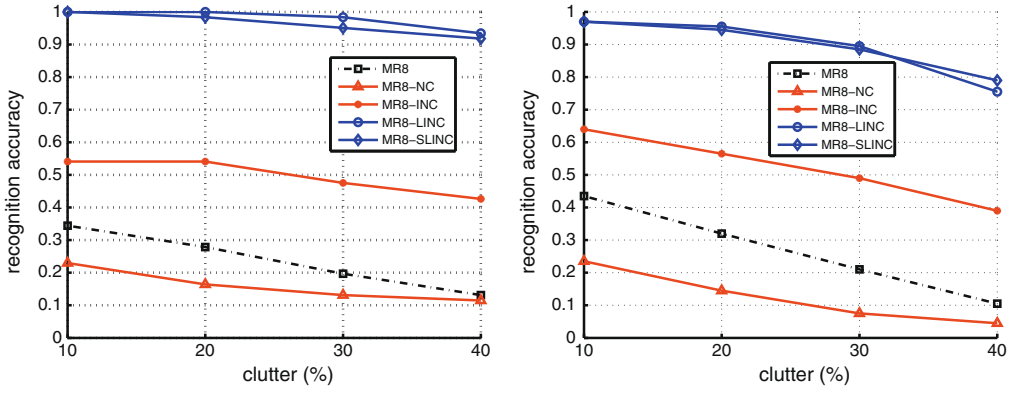
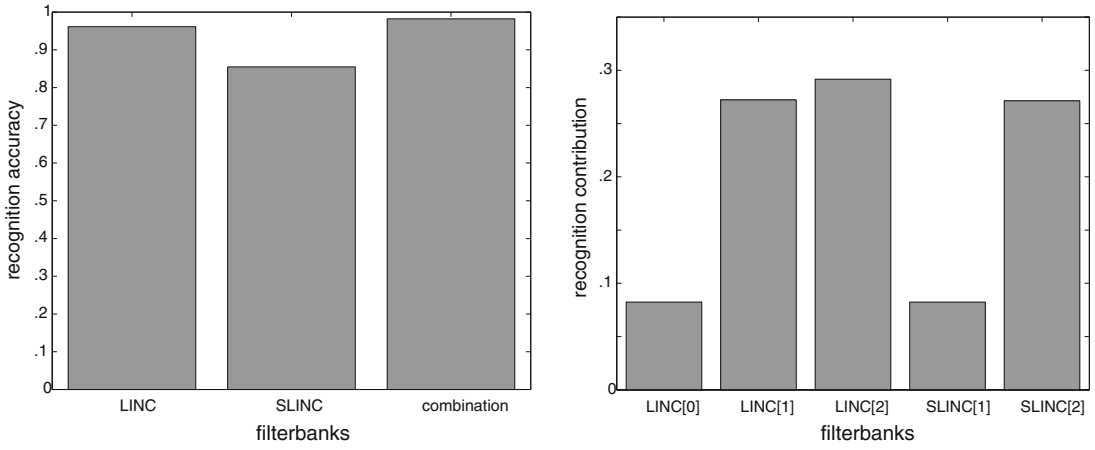
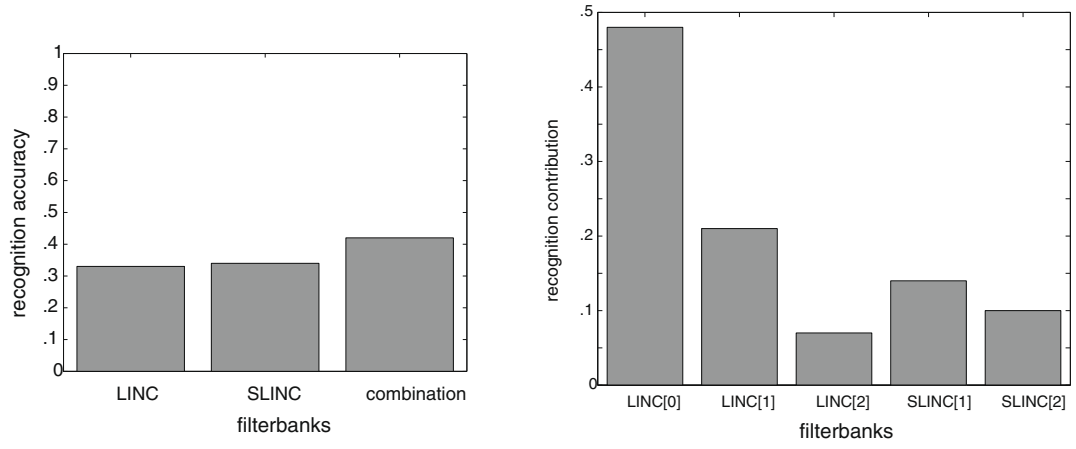


Fig. 7. Accuracy of material recognition for various filterbanks with increasingly cluttered images of (a) the CURET dataset and (b) the ALOT dataset.



(a) CURET: performance (b) CURET: filterbanks



(c) ALOT: performance (d) ALOT: filterbanks

Fig. 8. Accuracy of material recognition for the best performing filterbanks `MR8-LINC` and `MR8-SLINC` and their combination for the CURET dataset (a) and the ALOT dataset (c). Percentages indicate how often a particular color channel of a filterbank is most distinctive (b and d). Both filterbanks have three color channels, but note that `MR8-SLINC[0]` ≡ `MR8-LINC[0]`, so five results are shown.

Table 1
Summary of the performance of the `MR8` filterbank and proposed color modifications.

	Performance using 4 examples		Performance with 20% clutter		Merit	
	CURET (%)	ALOT (%)	CURET (%)	ALOT (%)	CURET (%)	A LOT (%)
`MR8`	58	23	29	32		
`MR8-NC`	54	36	17	15		
`MR8-INC`	60	22	54	57		
`MR8-LINC`	67	30	99	96		
`MR8-SLINC`	57	31	97	94		
Adaptive					2	20

of the material properties, and because more materials are included, the results generalize better. For ALOT the most distinctive filterbanks corresponds to intensity information. This can be explained from the fact that intensity variation rather than color variation is the dominating factor in material appearance (Koenderink et al., 1999). The information in the filterbanks that are not invariant to shadow and shading, `MR8-LINC`, is in 28% most distinctive. The shadow and shading invariant filterbank `MR8-SLINC` is in 25% most distinctive. We conclude that `MR8-LINC` and `MR8-SLINC` are discriminative for large but different sets of materials, respectively.

Finally, we stress that the recognition of materials from the ALOT dataset is obviously a far from solved problem. Here, we have demonstrated the merit of automatically tuning filterbanks with different invariant properties to individual materials with different physical properties.

5. Conclusion

In this paper, we have proposed a framework to learn for each material specifically the most distinctive filterbank from a set of intensity and color invariant filterbanks. The considered filterbanks are adopted from the distinctive `MR8` filterbank of Varma and Zisserman, from which color invariant filterbanks are constructed using techniques from literature. First we have established the distinctiveness, and the robustness to image settings and clutter, for individual filterbanks for the classification of more than 250 materials from the CURET and ALOT datasets, recorded under various illumination directions, viewpoints, and illumination colors. `MR8-NC` is the straightforward extension of `MR8` to color, and likewise it normalizes the mean and variance per color channel. We have shown that this proves to be a good strategy if multiple colors are apparent. `MR8-INC` normalizes each color channel by the variation of the intensity channel. This is a good strategy if the 3D variation of materials is limited. Two color filterbanks normalize locally the filterbank outputs. `MR8-LINC` normalizes locally by the intensity level to counteract intensity fluctuations, whereas

MR8-SLINC aims at shadow and shading invariant filterbank output. The locally-invariant filterbanks perform on average best, where MR8-LINC (MR8-SLINC) distinguishes better between materials with limited (significant) 3D variation. Additionally, we have demonstrated that the locally-invariant filterbanks are significantly more robust to image clutter than are filterbanks that involve global normalization. The results are summarized in Table 1.

Second, we have considered the performance of adapting filterbank combinations to each material specifically. This allows to tune for each material the color channel(s) and invariant properties that discriminates it best from other materials. We have proposed a scheme to do so by learning automatically the best discriminant function in joint filterbank space. Indeed, we have shown that the most distinctive filterbank differs across the CURET and ALOT datasets and across their individual materials. We have demonstrated that this automated tuning of color information and invariance to individual materials results in performance improvements of up to 20%. This result illustrates the merit of tuning a set of invariants to instances that have different physical properties.

Acknowledgements

This work is partly sponsored by the EU funded NEST project PERCEPT and the Dutch BSIK project Multimedian.

References

Burghouts, G.J., Geusebroek, J.M., 2006. Color textons for texture recognition. In: Proc. British Machine Vision Conf., vol. 3, pp. 1099–1108.
Burghouts, G.J., Geusebroek, J.M., 2008. Performance evaluation of local colour invariants. Comput. Vis. Image Und., in press.
Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints. In: Proc. European Conf. on Computer Vision.
Dana, K.J., Ginneken, B., Nayar, S.K., Koenderink, J.J., 1999. Reflectance and texture of real world surfaces. ACM Trans. Graphics 18 (1), 1–34.
Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. Wiley, New York.
Finlayson, G.D., Drew, M.S., Funt, B., 1994. Color constancy: Generalized diagonal transforms suffice. J. Optical Soc. Amer. A 11 (11), 3011–3019.
Geusebroek, J.M., Smeulders, A.W.M. Amsterdam Library of Textures (ALOT). Available from: <http://www.science.uva.nl/~mark/ALOT>.
Geusebroek, J.M., Boomgaard, R., Smeulders, A.W.M., Geerts, H., 2001. Color invariance. IEEE Trans. Pattern Anal. Machine Intell. 23 (12), 1338–1350.
Gevers, T., Smeulders, A.W.M., 1999. Color based object recognition. Pattern Recognition 32, 453–464.
Hayman, E., Caputo, B., Fritz, M., Eklundh, J.O., 2004. On the significance of real-world conditions for material classification. In: Proc. European Conf. on Computer Vision, No. 3, Springer-Verlag, pp. 253–266.
Jurie, F., Triggs, B., 2005. Creating efficient codebooks for visual recognition. In: Proc. Internat. Conf. on Computer Vision.
Koenderink, J.J., Doorn, A., Dana, K.J., Nayar, S., 1999. Bidirectional reflection distribution function of thoroughly pitted surfaces. Internat. J. Comput. Vision 31, 129–144.
Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Machine Intell. 27 (8), 1265–1278.
Leung, T., Malik, J., 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. Internat. J. Comput. Vision 43 (1), 29–44.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vision 60 (2), 91–110.
Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. 27 (10), 1615–1630.
Nowak, E., Jurie, F., Triggs, B., 2006. Sampling strategies for bag-of-features image classification. In: Proc. European Conf. on Computer Vision, Springer-Verlag.
Perronnin, F., Dance, C., Csurka, G., Bressan, M., 2006. Adapted vocabularies for generic visual categorization. In: Proc. European Conf. on Computer Vision, Springer-Verlag.
Shotton, J., Winn, J., Rother, C., Criminisi, A., 2006. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Proc. European Conf. on Computer Vision, Springer-Verlag.
Sivic, J., Zisserman, A., 2003. Video Google: A text retrieval approach to object matching in videos. In: Proc. Internat. Conf. on Computer Vision.
Suen, P., Healey, G., 2000. The analysis and recognition of real-world textures in three dimensions. IEEE Trans. Pattern Anal. Machine Intell. 22 (5), 491–503.
Varma, M., Zisserman, A., 2005. A statistical approach to texture classification from single images. Internat. J. Comput. Vision 62 (1–2), 61–81.
Winn, J., Criminisi, A., Minka, T., 2005. Object categorization by learned universal visual dictionary. In: Proc. Internat. Conf. on Computer Vision, IEEE Computer Society, pp. 1800–1807.
Zhang, J., Marsza, M., Lazebnik, S., Schmid, C., in press. Local features and kernels for classification of texture and object categories: A comprehensive study. Internat. J. Comput. Vision.