Rotation–covariant visual concept detection using steerable Riesz wavelets and bags of visual words

Adrien Depeursinge[a,b,c], Antonio Foncubierta[c], Henning Müller[b,c], Dimitri Van de Ville[b,d]

[a]Stanford University, Stanford, California, USA;
[b]University and University Hospitals of Geneva, Medical Informatics, Geneva, Switzerland;
[c]University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland;
[d]Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland;

ABSTRACT

Distinct texture classes are often sharing several visual concepts. Texture instances from different classes are sharing regions in the feature hyperspace, which results in ill–defined classification configurations. In this work, we detect rotation–covariant visual concepts using steerable Riesz wavelets and bags of visual words. In a first step, K–means clustering is used to detect visual concepts in the hyperspace of the energies of steerable Riesz wavelets. The coordinates of the clusters are used to construct templates from linear combinations of the Riesz components that are corresponding to visual concepts. The visualization of these templates allows verifying the relevance of the concepts modeled. Then, the local orientations of each template are optimized to maximize their response, which is carried out analytically and can still be expressed as a linear combination of the initial steerable Riesz templates. The texture classes are learned in the feature space composed of the concatenation of the maximum responses of each visual concept using support vector machines. An experimental evaluation using the Outex_TC_00010 test suite allowed a classification accuracy of 97.5%, which demonstrates the feasibility of the proposed approach. An optimal number $K = 20$ of clusters is required to model the visual concepts, which was found to be fewer than the number of classes. This shows that higher–level classes are sharing low–level visual concepts. The importance of rotation–covariant visual concept modeling is highlighted by allowing an absolute gain of more than 30% in accuracy. The visual concepts are modeling the local organization of directions at various scales, which is in accordance with the bottom–up visual information processing sequence of the primal sketch in Marr's theory on vision.

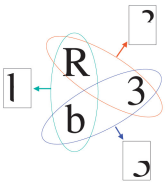
Keywords: Bags of visual words, texture classification, Riesz transform, steerability, rotation–covariance, wavelet analysis, visual concept detection.

1. INTRODUCTION

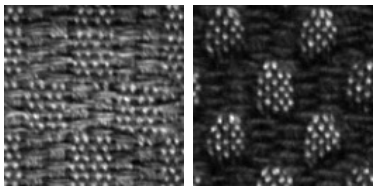
Low–level visual concept modeling is key to image understanding and categorization.[1,2] A large group of theories in visual processing support that the understanding of complex scenes is typically carried in a bottom–up process, in which information is processed sequentially with increasing complexities.[3] In[1] visual concepts are called *geons* and are simple forms including rectangles, circles, bricks and wedges. It is assumed that higher–level objects are constituted of geons and their relations, where the total number of geons does not exceed 40 elements. *Textons*[4] or *texture primitives*[5] are the counterpart of geons for texture understanding, and constitute elementary building blocks of higher–level texture classes. A more general definition of visual concepts is proposed by the *primal sketch* theory of Marr et al.[2] and the *sketchability* property introduced by Guo et al.,[6] where visual concepts are characterized by local organizations of directions at a fixed scale.

Distinct visual classes are often sharing several low–level visual concepts. This is illustrated in Fig. 1a, where the characters R, 3 and b are all sharing geometrical shapes. This is also the case for textured images as it can be observed in Fig. 1b, where both textures are containing patterns composed of tiny checkerboards of white dots. Biomedical textures resulting from the alteration of normal tissue are also by definition sharing visual primitives. This is illustrated in Fig. 1c, where micronodular patterns in computed tomography (CT) of the lungs are characterized be the superposition of micronodules and healthy tissue. When using computerized

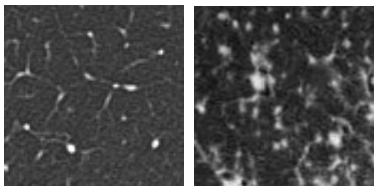
Further author information: (Send correspondence to Adrien Depeursinge, adrien.depeursinge@hevs.ch)



(a) Characters R, 3 and b are all sharing geometrical shapes (source[7]).



(b) Both textures are containing patterns composed of tiny checkerboards of white dots.



(c) Micronodular patterns (i.e., right image) in CT are characterized be the superposition of micronodules and healthy tissue (i.e., right image).

Figure 1: Higher–level classes are often sharing several low–level visual concepts.

approaches for the automatic categorization of image classes, classes that are sharing concepts are also sharing regions in the feature hyperspace, which results in ill–defined classification configurations.[7]

When shared visual concepts are known, specific detectors can be designed and used to decompose higher–level visual classes into simpler primitives, as suggested by vision scientists in the late 1970's.[3] Low–level visual concept detection has been a very active research field of the computer vision community during the past 30 years.[2,6] Basic visual concept detectors were initially proposed to model isotropic blobs, edges, ridges and corners.[2,6,8–11] Efforts for defining the hierarchical semantic vocabularies and the dependencies between visual concepts was proposed by the ImageNet initiative*.[12] This motivated several researchers to learn the low–level visual concepts of ImageNet using various approaches.[13,14] Automated visual concept annotation in large photo collections has been the center of interest of the ImageCLEF community†.[15] However, in all aforementioned approaches, the modeled visual concepts are not directly corresponding to the actual primitives in real datasets, which does not ensure to fully leverage the wealth of visual information present and results in limited categorization performance.

1.1. The bags of visual words approach

The aforementioned limitation motivated researchers to develop approaches that can learn the visual concepts shared by higher–level classes. The ensemble of visual concepts is often called the *visual vocabulary*. A notable example is the bags of visual words (BOVW) approach.[16] The central idea of BOVW comes from the text processing community, where a textual document is described as the histogram of occurrences of words present in the collection of documents. BOVW counts the occurrences of visual words (VW) appearing in different subregions (e.g., square blocks or patches) of an image. VWs are typically defined as cluster centers in a given feature space populated by image patches. The BOVW approach was used in a wide range of applications including video indexing,[16] medical image retrieval[17–20] and general image annotation.[21] Several attempts were made for visualizing the VWs, aiming at interpreting the visual semantics modeled. In[17,20] color image overlays are used to mark the local presence of VWs in image examples. in[18,19,21] prototype image blocks that are the closest to the respective VWs are displayed to visualize the information modeled. Unfortunately, VWs are often very difficult to interpret, especially when texture information is considered.

1.2. Rotation–covariant texture concepts

Although humans are able to distinguish texture concepts with high precision, the semantic vocabulary for texture description is scarce. The primal sketch theory highlights the importance of the local organization of directions at a fixed scale in human visual interpretation. Leveraging this property calls for the design of multi–directional and multi–scale operators. The multiresolution theory of the wavelet transform provides an elegant solution to the locality problem for scale characterization.[22] Steerable filterbanks allow continuous characterizations of the directions from linear combinations of the basis filters, where the linear weights can be determined analytically.[23,24] Steerable wavelets (e.g., the steerable pyramid[25,26]) are combining the two frameworks, enabling multi–scale and multi–directional analysis.

*`http://www.image-net.org/`, as of 30 July 2013.

†`http://imageclef.org/2013/photo/`, as of 30 July 2013.

In some cases, it can be desirable to detect visual concepts independently from their orientation. Whereas isotropic operators (e.g., Laplacian of Gaussian) are providing identical output for rotated versions of visual concepts,[20] they do not allow for characterizing local directions. These operators are providing rotation–invariant analysis. Rotation–covariant operators allow keeping local directional information while normalizing the operators' outputs over the instances. A simple approach for obtaining rotation–covariant operators is to compute the response of directional operators at equally sampled directions, and concatenate the response to create multi–directional feature vectors. Unfortunately, the latter requires to optimize the trade–off between angular precision and dimensionality of the feature space by choosing the number of directions. A more elegant approach to obtain rotation–covariance with infinitesimal angular precision is to use steerability to locally align the operators.[27]

Several researchers proposed to learn filters from data using linear combinations of multi–scale and/or steerable filterbanks, aiming at modeling local organizations of scales and directions.[28–34] Most approaches[28,29,31,32] use singular value decomposition (SVD) and principal component analysis (PCA) to estimate the importance of every basis template (i.e., the linear weights). In previous work,[33,34] we used support vector machines (SVM) to learn class–wise optimally discriminant combinations of steerable Riesz wavelets. We observed that when the classes are sharing several low–level concepts, the direct discrimination between the classes leads to ill–defined classification boundaries, because they are sharing regions in the feature space. This observation motivated the use of the BOVW approach to decompose higher–level texture classes into subsets of low–level visual concepts. The steerability property of the Riesz wavelets is leveraged to maximize the local responses of the learned visual concepts analytically. This allows for rotation–covariant visual concept detection.

2. MATERIAL AND METHODS

This section describe our approach for rotation–covariant visual concept detection using BOVW and steerable Riesz wavelets. The definition of Nth–order Riesz filterbanks and their properties are detailed in Section 2.1. The approach for modeling rotation–covariant visual concepts is described in Sections 2.2 and 2.3. The dataset and experimental setup used to evaluate the proposed approach are explained in Section 2.4.

2.1. Steerable Riesz filterbanks

Multi–scale and multi–directional image representations are obtained using steerable Riesz filterbanks.[35] The Nth–order Riesz $\boldsymbol{\mathcal{R}}^N$ transform of a 2–D signal $f(x)$ yields $N+1$ components as:

$$\boldsymbol{\mathcal{R}}^N\{f\}(\boldsymbol{x}) = \begin{pmatrix} \mathcal{R}^{(0,N)}\{f\}(\boldsymbol{x}) \\ \vdots \\ \mathcal{R}^{(n,N-n)}\{f\}(\boldsymbol{x}) \\ \vdots \\ \mathcal{R}^{(N,0)}\{f\}(\boldsymbol{x}) \end{pmatrix}, \tag{1}$$

with $n = 0, 1, \ldots, N$. A singular component $\mathcal{R}^{(n,N-n)}\{f\}(\boldsymbol{x})$ is defined in the Fourier domain as:

$$\mathcal{R}^{(n,N-n)}\{f\}(\boldsymbol{x}) \overset{\mathcal{F}}{\longleftrightarrow} \widehat{\mathcal{R}^{(n,N-n)}}\{f\}(\boldsymbol{\omega}),$$

where

$$\widehat{\mathcal{R}^{(n,N-n)}}\{f\}(\boldsymbol{\omega}) = \sqrt{\frac{N}{n!(N-n)!}}\frac{(-j\omega_1)^n(-j\omega_2)^{N-n}}{||\boldsymbol{\omega}||^N}\hat{f}(\boldsymbol{\omega}), \tag{2}$$

with $\omega_{1,2}$ corresponding to the frequencies along the vertical and horizontal directions $x_{1,2}$. The multiplication with $j\omega_{1,2}$ in the numerator corresponds to partial derivatives of f and the division by the norm of $\boldsymbol{\omega}$ in the denominator makes that only phase information is retained. Therefore, $\boldsymbol{\mathcal{R}}^N$ yields allpass[‡] filterbanks with directional (singular) components $\mathcal{R}^{(n,N-n)}$.[35] The angular coverage of the Riesz components is determined by the partial derivatives in Eq. (2). Therefore, the angular selectivity of the components is controlled by the order

[‡]Except for the DC component.

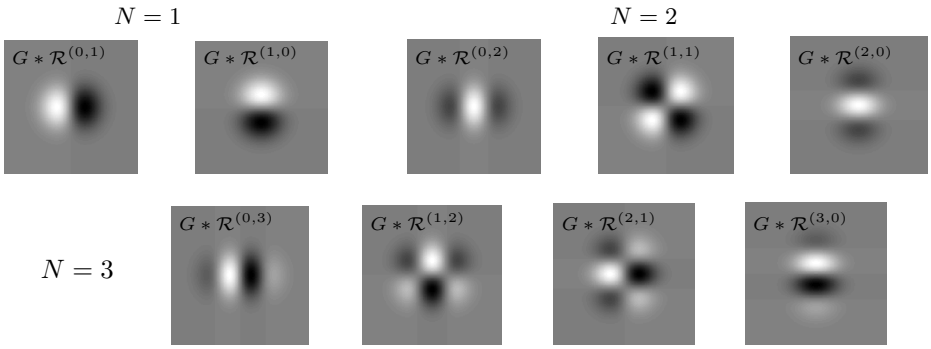


Figure 2: Templates corresponding to the Riesz kernels convolved with a Gaussian smoother for N=1,2,3.

N of the transform. The higher–order versions as specified in (2) are obtained by regrouping the 2^N Riesz filters into $N+1$ components by commutativity of convolution (e.g., $\partial^2/\partial x \partial y$ is equivalent to $\partial^2/\partial y \partial x$). The Riesz kernels $\mathcal{R}^{(n,N-n)}$ convolved with Gaussian kernels for N=1,2,3 are depicted in Fig. 2. The Riesz components are forming steerable filterbanks, which means that the local response of each component $\mathcal{R}^{(n,N-n)}$ of an image $f(\boldsymbol{x})$ rotated by an arbitrary angle θ can be derived analytically from a linear combination of the responses of all components of the filterbank using a steering matrix $\boldsymbol{A}^\theta$ as follows:[34]

$$\boldsymbol{\mathcal{R}}^N \left\{ f^\theta \right\} (\mathbf{0}) = \boldsymbol{A}^\theta \boldsymbol{\mathcal{R}}^N \left\{ f \right\} (\mathbf{0}). \tag{3}$$

We obtain multi–scale versions of these filterbanks by coupling the Riesz transform with Simoncelli's multi–resolution framework based on isotropic band–limited wavelets.[35]

2.2. Visual concept modeling

The multi–scale and multi–directional properties of the Riesz filterbanks are leveraged to build visual concepts in the sense of Marr's primal sketch, where the local organization of directions is characterized for various scales. We define K visual concepts Γ_k^N as linear combinations of multi–scale Riesz components as:[33,34]

$$\Gamma_k^N = w_1 \left(\mathcal{R}^{(0,N)} \right)_{s_1} + w_2 \left(\mathcal{R}^{(1,N-1)} \right)_{s_1} + \cdots + w_{J(N+1)} \left(\mathcal{R}^{(N,0)} \right)_{s_J}, \tag{4}$$

where $\boldsymbol{w}_k$ contains the weights of the respective Riesz components and s_j, $j = 1, \ldots, J$ is the scale index. Scale–wise visual concepts $\Gamma_{k,j}^N$ can be obtained when using only weights and corresponding Riesz templates at the scale j. An example of the construction of Γ_k^8 for a texture containing the visual concept k is illustrated in Figure 3. The sets of weights $\boldsymbol{w}_k$ are obtained as the coordinates of K clusters in the hyperspace of the energies $E\left(\mathcal{R}^{(n,N-n)}\left\{f\right\}(\boldsymbol{x})\right)$ of multi–scale N–th order Riesz wavelets. This is similar to the definition of visual words in the BOVW approach. K–means clustering with a l_2–norm Euclidean distance is used in the feature space spanned by the normalized energies E of $J(N+1)$ Riesz components to determine the cluster coordinates.

2.3. Rotation–covariant visual concepts: steering Γ_k^N

By combining Eqs. (3) and (4), the response of Γ_k^N rotated by an arbitrary angle θ can be derived analytically as:[34]

$$\Gamma_k^{N,\theta} = \boldsymbol{w}_k^{\mathrm{T}} \boldsymbol{A}^\theta \boldsymbol{\mathcal{R}}^N. \tag{5}$$

It can be observed that the expression of $\Gamma_k^{N,\theta}$ can still be expressed as a linear combination of the initial Riesz templates $\boldsymbol{\mathcal{R}}^N$.

To obtain rotation–covariant representations of the visual concepts, Γ_k^N are locally steered to maximize their response over θ. The dominant orientation θ_{dom} of $\Gamma_{k,j}^N$ at the position $\boldsymbol{x}_p$ is

$$\theta_{\mathrm{dom}}(\boldsymbol{x}_p) = \underset{\theta \in [0,\pi]}{\arg\max} \left(\boldsymbol{w}_{k,j}^{\mathrm{T}} \boldsymbol{A}^\theta \boldsymbol{\mathcal{R}}^N \left\{ f \right\} \right) (\boldsymbol{x}_p). \tag{6}$$

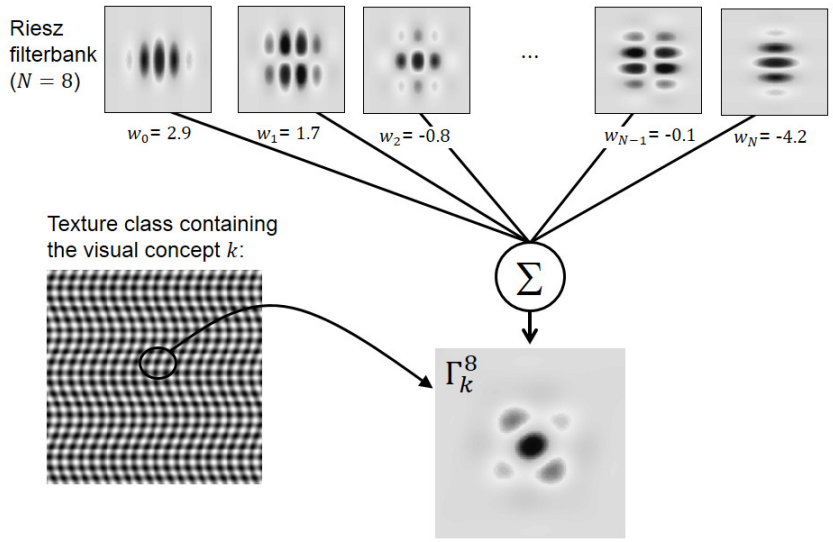


Figure 3: Example of the construction of a visual concept Γ_k^8 using a linear combination of the Riesz templates $\mathcal{R}^{(n,N-n)}$ with $N = 8$ and $J = 1$.

Eq. (6) can be efficiently solved by finding the roots of the derivative of Eq. (5) with respect to θ. A matrix $\Theta(\boldsymbol{x})$ of all angles is obtained for all positions $\boldsymbol{x}_p$. Riesz templates from all scales are steered together using one unique multi–scale angle matrix $\Theta_j(\boldsymbol{x})$, which contains local angle values from scale j having a maximum magnitude of $\Gamma_{k,j}^N$.

2.4. Outex_TC_00010 test suite and experimental setup

The proposed framework is evaluated using the Outex database[§].[36] This is a publicly available set of real textures photographed with controlled illumination conditions for the experimental evaluation of texture classification algorithms. The Outex_TC_00010 test suite has recently been used by several studies on texture recognition to focus on the rotation–invariant properties of the approaches.[37–52] It consists of the 24 texture classes with pronounced directional structures. For each class, the underlying texture patterns are roughly uniform over the whole initial images of size 538×746, although gray–scale variations caused by color variations of the photograph exist. Each texture sample is captured using nine rotation angles (0°, 5°, 10°, 15°, 30°, 45°, 60°, 75°, and 90°). The full images are divided into 128×128 non–overlapping blocks, leading to 20 texture instances per class. A total of 4320 ($24 \times 20 \times 9$) image instances are used to evaluate the proposed approach. The training set consists of the 480 (24×20) non–rotated images and the remaining 3840 ($24 \times 20 \times 8$) images from 8 orientations are constituting the test set. Texture instances for each class are depicted in Figure 4.

Every image is expressed in the feature space spanned by the concatenation of the energies of the K visual concepts. The dimensionality of the feature space is $K \cdot J \cdot (N + 1)$. It is important to note that although $\Gamma_{k,j}^N$ satisfies the wavelet admissibility condition (i.e., zero mean), the feature space obtained is not equivalent to the convolution of the $\Gamma_{k,j}^N$ with the signal, because every image instance is still expressed in terms of the normalized energies E of each locally steered individual Riesz component $\mathcal{R}^{(n,N-n)}$. The visual concepts from all scales are steered together using the angle matrix $\Theta_1(\boldsymbol{x})$ from the finest scale $j = 1$. Within this feature space, SVMs with a Gaussian kernel are trained using the 480 non–rotated images and tested on the remaining 3840 rotated images. The cost C of the SVMs and the width σ of the Gaussian kernel are optimized as $C = 10^0, \ldots, 10^8$ and $\sigma = 10^{-5}, \ldots, 10^5$.

[§] `http://www.outex.oulu.fi/`, as of 30 July 2013.

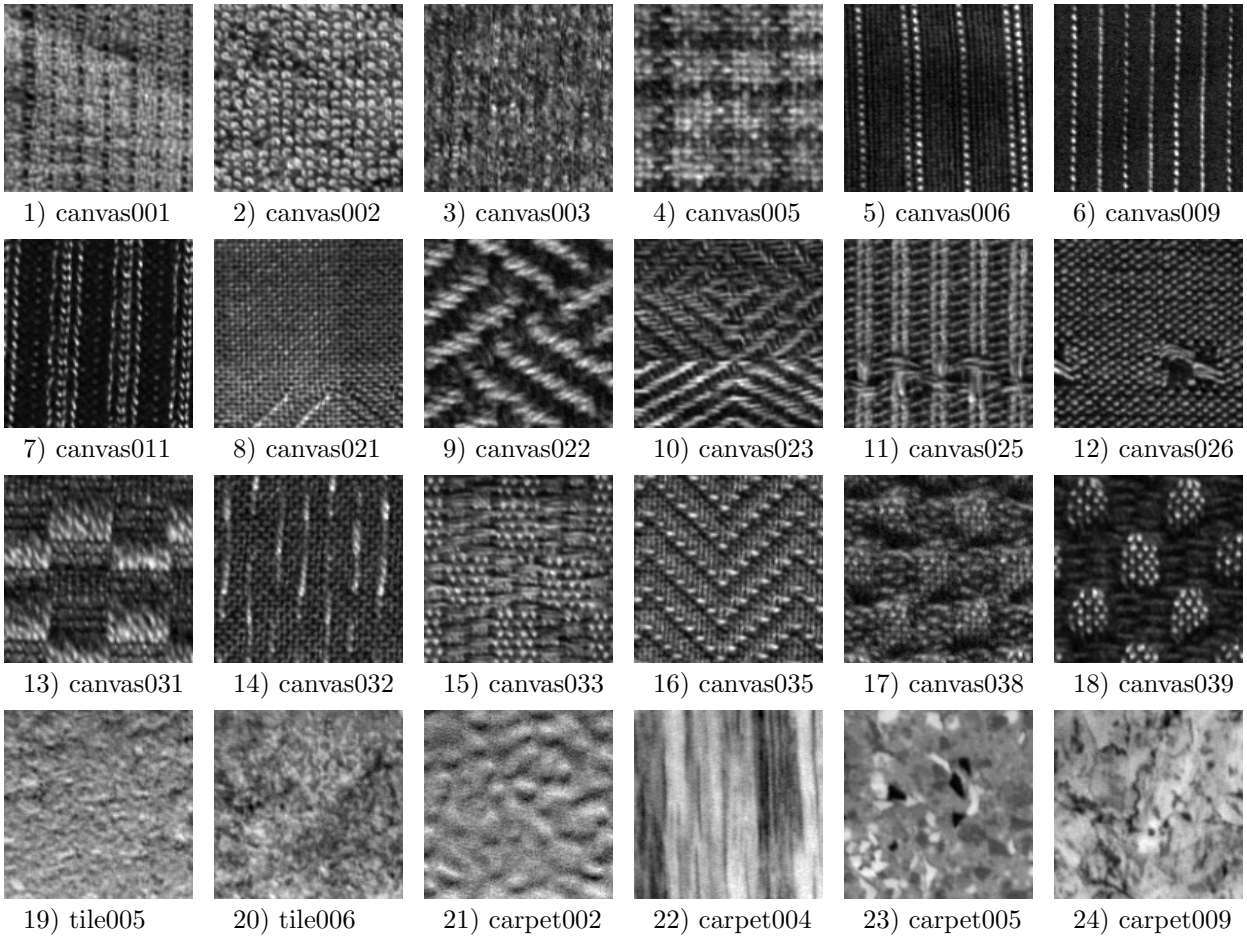


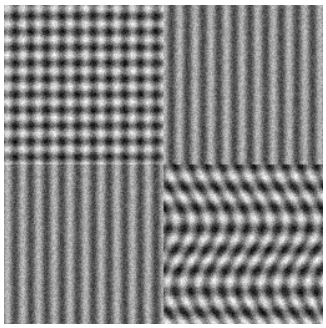
Figure 4: 128×128 blocks from the 24 texture classes of the Outex_TC_00010 test suite.

3. RESULTS

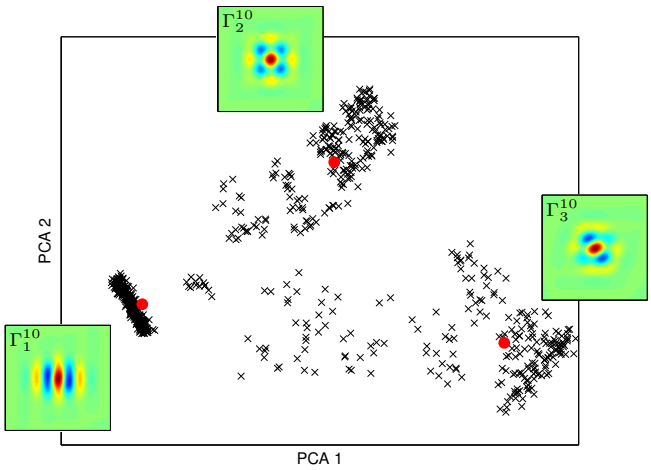
A qualitative interpretation of the information modeled by the visual concepts Γ_k^N is proposed in Fig. 5 using a synthetic image composed of three well–defined visual concepts along with uniformly distributed noise (see Fig. 5a). The classification performance obtained with the Outex_TC_00010 test suite is shown in Fig. 6 for $N = 1, \ldots, 8$ and $K = 5, \ldots, 30$. The best accuracy of 0.975 is obtained with $N = 4$ and $K = 20$. The corresponding distribution of image instances and visual concepts is depicted in Fig. 7. The performance gain allowed by the local steering of the templates Γ_k^N when compared to using SVM classification from the energies of the initial Riesz components is illustrated in Fig. 8 for $K = 20$ and $N = 1, \ldots, 8$.

4. DISCUSSIONS AND CONCLUSIONS

We developed an approach for the detection of rotation–covariant visual concepts by combining the BOVW framework with steerable Riesz wavelets. The visual concepts are modeling the local organization of directions at various scales, which is in accordance with the bottom–up visual information processing sequence of the primal sketch in Marr's theory on vision.[2,3] The relevance of the modeled visual concepts can be verified by visualizing the shapes of templates built from linear combinations of steerable Riesz components. The templates Γ_k^{10} displayed in Fig. 5 are corresponding to the actual visual concepts contained in the synthetic image (see Fig. 5a) for the scale $j = 3$. Qualitatively, Γ_1^{10} corresponds to a line detector, whereas Γ_2^{10} and Γ_3^{10} are implementing straight and wiggled checkerboard detectors, respectively. This experiment demonstrates the ability of the framework to extract distinct visual concepts in an unsupervised manner, similarly to the BOVW



(a) Synthetic image containing 3 visual concepts:
1) vertical lines (quadrants I and III),
2) checkerboard (quadrant II),
3) wiggled checkerboard (quadrant IV).



(b) PCA visualization of 32×32 overlapping blocks and clusters from the left image ($N = 10$, $J = 4$, $K = 3$). The templates Γ_k^{10} corresponding to the respective visual concepts are dislayed for scale $j = 3$.

Figure 5: Qualitative evaluation of the visual concepts Γ_k^{10} found using K–means in the feature space spanned by the energies of the multi–scale Riesz components.

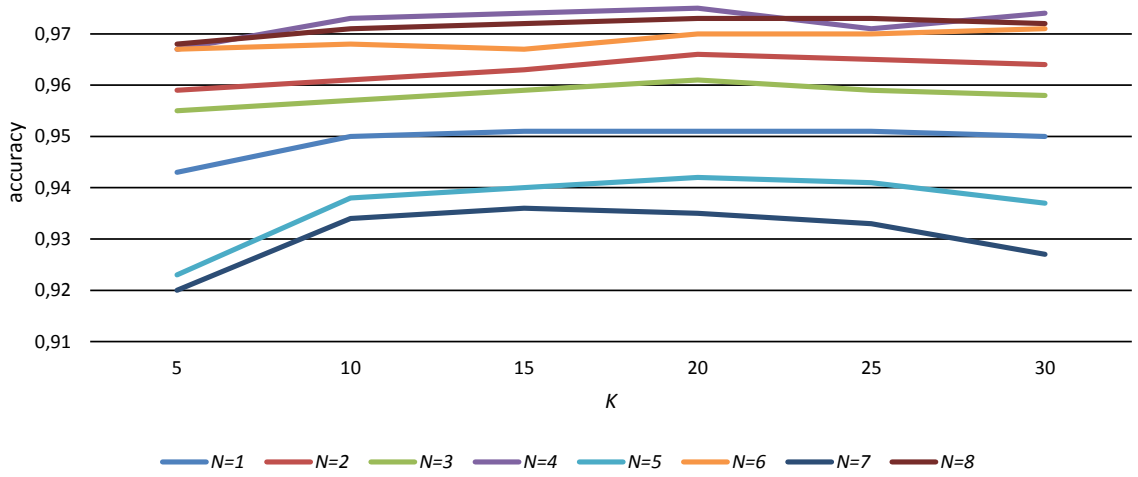


Figure 6: Classification accuracy with the Outex_TC_00010 test suite. An optimal number of visual concepts $K = 20$ and order $N = 4$ allowed an accuracy of 97.5%.

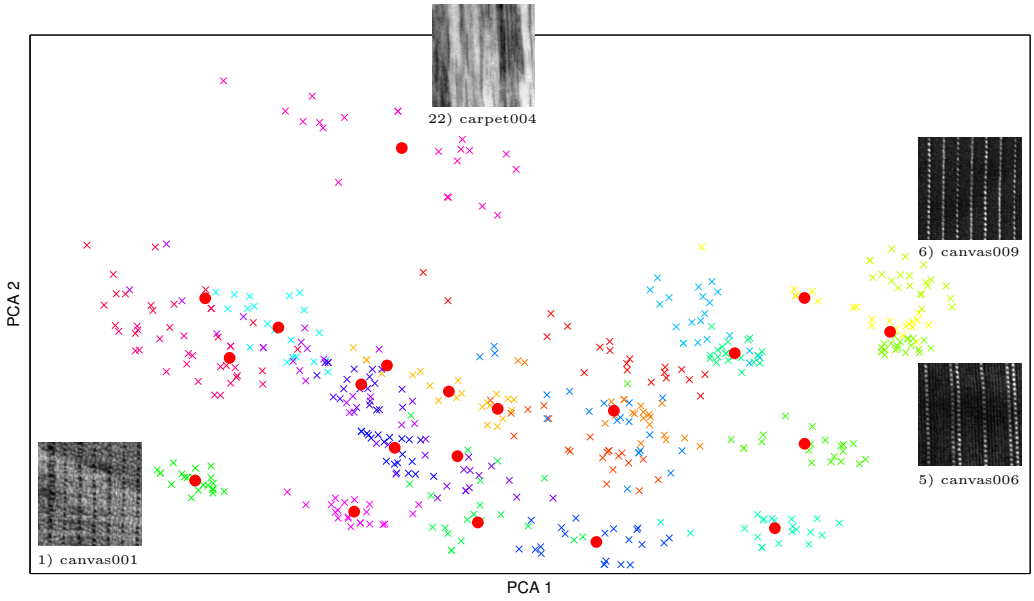


Figure 7: Distribution of the 24 classes of the Outex_TC_00010 test suite and associated visual concepts for the best configuration ($N = 4$, $K = 20$, classification accuracy=0.975).

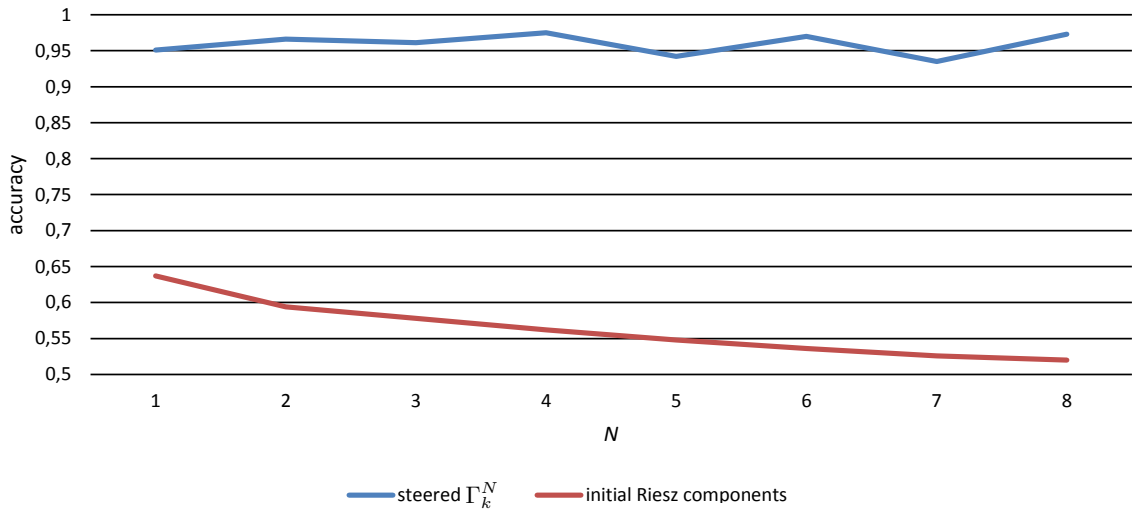


Figure 8: Influence of the local steering of Γ_k^N ($K = 20$) on the classification accuracy when compared to using the energies of the initial Riesz components for the classification.

approach. The best classification accuracy of 97.5% obtained with the Outex_TC_00010 test suite is among the best performing approaches based on filters and wavelets in the literature.[37–44] The optimal number of visual concepts K is 20 and is inferior to the number of classes (i.e., 24), which shows that several low–level concepts are shared among the higher–level classes (see Fig. 6). The distribution of the visual concepts and classes is shown in Fig. 7. It can be observed that classes *1) canvas001* and *22) carpet004* are not sharing much visual content with others whereas classes *5) canvas006* and *6) canvas009* are both containing vertical lines of small dots. The optimal order of the Riesz transform $N = 4$ constitutes an excellent trade–off between the dimensionality of the feature space and the wealth of the filterbank. The importance of rotation–covariance of the visual concepts is demonstrated in Fig. 8 where the classification performance based on the initial Riesz components is very low.

The results obtained are encouraging and call for future work to further push the classification accuracy. Our previous work using SVMs to learn class–wise templates $\Gamma_{c,j}^{N}$ allowed a classification accuracy of 98.4% for $N = 8$,[34] which indicates that further optimization of the current approach is required. Other clustering algorithms will be investigated. The images from the Outex dataset will be divided into smaller blocks (e.g., 32×32) to avoid measuring the energies of the coefficients over the whole image, which mixes the responses of distinct visual concepts. Further modeling of the contextual relationships between visual concepts will also be investigated (e.g., co–occurrences of visual concepts[53]).

Acknowledgments

This work was supported by the CIBM, the Swiss National Science Foundation (under grants 205320–141300/1, PBGEP2_142283 and PP00P2–123438), and the EU 7th Framework Program in the context of the Khresmoi project (FP7–257528).

REFERENCES

1. I. Biedermann, "Recognition–by–components: A theory of human image understanding," *Psychological Review* **94**, pp. 115–147, April 1987.
2. D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings Royal Society of London* **207**, pp. 187–217, February 1980.
3. D. C. Marr, "Early processing of visual information," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **275**, pp. 483–519, October 1976.
4. B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature* **290**, pp. 91–97, March 1981.
5. R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE* **67**, pp. 786–804, May 1979.
6. C.-E. Guo, S.-C. Zhu, and Y. N. Wu, "Towards a mathematical theory of primal sketch and sketchability," in *Ninth IEEE International Conference on Computer Vision*, **2**, pp. 1228–1235, 2003.
7. A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(5), pp. 854–869, 2007.
8. K. I. Laws, "Rapid texture identification," in *24th Annual Technical Symposium*, **238**, pp. 376–381, SPIE, 1980.
9. L. Kitchen and A. Rosenfeld, "Gray–level corner detection," *Pattern Recognition Letters* **1**(2), pp. 95–102, 1982.
10. J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6), pp. 679–698, 1986.
11. T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision* **30**(2), pp. 117–156, 1998.
12. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large–scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 248–255, 2009.
13. T. Deselaers and V. Ferrari, "Visual and semantic similarity in ImageNet," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011*, pp. 1777–1784, 2011.

14. O. Russakovsky and L. Fei-Fei, "Attribute learning in large–scale datasets," in *Trends and Topics in Computer Vision*, K. N. Kutulakos, ed., *Lecture Notes in Computer Science* **6553**, pp. 1–14, Springer Berlin Heidelberg, 2012.
15. S. Nowak and P. Dunker, "Overview of the CLEF 2009 large–scale visual concept detection and annotation task," in *Multilingual Information Access Evaluation II. Multimedia Experiments*, C. Peters, B. Caputo, J. Gonzalo, G. J. F. Jones, J. Kalpathy-Cramer, H. Müller, and T. Tsikrika, eds., *Lecture Notes in Computer Science* **6242**, pp. 94–109, Springer Berlin Heidelberg, 2010.
16. J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, *ICCV '03*, pp. 1470–1477, IEEE Computer Society, (Washington, DC, USA), 2003.
17. B. André, T. Vercauteren, A. M. Buchner, M. B. Wallace, and N. Ayache, "Endomicroscopic video retrieval using mosaicing and visual words," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, *ISBI 2010*, pp. 1419–1422, 2010.
18. U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X–ray categorization and retrieval on the organ and pathology level, using patch–based visual words.," *IEEE Transactions on Medical Imaging* **30**(3), pp. 733–746, 2011.
19. A. Burner, R. Donner, M. Mayerhoefer, M. Holzer, F. Kainberger, and G. Langs, "Texture bags: Anomaly retrieval in medical images based on local 3D–texture similarity," in *Medical Content-based Retrieval for Clinical Decision Support*, H. Greenspan, H. Müller, and T. Syeda-Mahmood, eds., *MCBR-CDS 2011* **7075**, pp. 116–127, Lecture Notes in Computer Sciences (LNCS), September 2012.
20. A. Foncubierta-Rodríguez, A. Depeursinge, and H. Müller, "Using multiscale visual words for lung texture classification and retrieval," in *Medical Content–based Retrieval for Clinical Decision Support*, H. Greenspan, H. Müller, and T. Syeda Mahmood, eds., *MCBR–CDS 2011* **7075**, pp. 69–79, Lecture Notes in Computer Sciences (LNCS), September 2012.
21. Y. Wang, T. Mei, S. Gong, and X.-S. Hua, "Combining global, regional and contextual features for automatic image annotation," *Pattern Recognition* **42**(2), pp. 259–266, 2009.
22. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 674–693, July 1989.
23. W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, pp. 891–906, September 1991.
24. E. P. Simoncelli and H. Farid, "Steerable wedge filters for local orientation analysis," *IEEE Transactions on Image Processing* **5**, pp. 1377–1382, September 1996.
25. H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 222–228, June 1994.
26. E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi–scale derivative computation," in *Proceedings of International Conference on Image Processing, 1995.*, **3**, pp. 444–447, October 1995.
27. A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, "Lung texture classification using locally-oriented Riesz components," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2011*, G. Fichtinger, A. Martel, and T. Peters, eds., *Lecture Notes in Computer Science* **6893**, pp. 231–238, Springer Berlin / Heidelberg, September 2011.
28. P. Perona, "Deformable kernels for early vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**, pp. 488–499, May 1995.
29. G. Sommer, M. Michaelis, and R. Herpers, "The SVD approach for steerable filter design," in *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems*, *ISCAS 1998* **5**, pp. 349–353, 1998.
30. G. Gonzalez, F. Fleuret, and P. Fua, "Learning rotational features for filament detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, *CVPR 2009*, pp. 1582–1589, June 2009.
31. Z. Puspoki, C. Vonesch, and M. Unser, "Detection of symmetric junctions in biological images using 2–D steerable wavelet transforms," in *IEEE 10th International Symposium on Biomedical Imaging*, *ISBI 2013*, pp. 1496–1499, 2013.

32. C. Vonesch, F. Stauber, and M. Unser, "Design of steerable filters for the detection of micro–particles," in *IEEE 10th International Symposium on Biomedical Imaging, ISBI 2013*, pp. 934–937, 2013.
33. A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, "Multiscale lung texture signature learning using the Riesz transform," in *Medical Image Computing and Computer–Assisted Intervention MICCAI 2012, Lecture Notes in Computer Science* **7512**, pp. 517–524, Springer Berlin / Heidelberg, October 2012.
34. A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, "Rotation–covariant feature learning using steerable Riesz wavelets," *IEEE Transactions on Image Processing* , submitted.
35. M. Unser and D. Van De Ville, "Wavelet steerability and the higher–order Riesz transform," *IEEE Transactions on Image Processing* **19**, pp. 636–652, March 2010.
36. T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen, "Outex – new framework for empirical evaluation of texture analysis algorithms," in *16th International Conference on Pattern Recognition, ICPR* **1**, pp. 701–706, IEEE Computer Society, August 2002.
37. K. Jafari-Khouzani and H. Soltanian-Zadeh, "Radon transform orientation estimation for rotation invariant texture analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, pp. 1004–1008, June 2005.
38. P. Janney and Z. Yu, "Invariant features of local textures—a rotation invariant local texture descriptor," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7, 2007.
39. P. Southam and R. Harvey, "Towards texture classification in real scenes," in *Proceedings of the British Machine Vision Conference*, pp. 240–250, 2005.
40. T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray–scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, pp. 971–987, July 2002.
41. R. Porter and N. Canagarajah, "Robust rotation–invariant texture classification: wavelet, Gabor filter and GMRF based schemes," *IEE Proceedings on Vision, Image and Signal Processing* **144**, pp. 180–188, jun 1997.
42. U. Kandaswamy, S. A. Schuckers, and D. Adjeroh, "Comparison of texture analysis schemes under nonideal conditions," *IEEE Transactions on Image Processing* **20**, pp. 2260–2275, August 2011.
43. X. Qian, X.-S. Hua, P. Chen, and L. Ke, "PLBP: An effective local binary patterns texture descriptor with pyramid representation," *Pattern Recognition* **44**(10–11), pp. 2502–2515, 2011.
44. Y. Ben Salem and S. Nasri, "Rotation invariant texture classification using support vector machines," in *IEEE International Conference on Communications, Computing and Control Applications, CCCA 2011*, pp. 1–6, IEEE, 2011.
45. O. Ghita, D. Ilea, A. Fernandez, and P. Whelan, "Local binary patterns versus signal processing texture analysis: a study from a performance evaluation perspective," *Sensor Review* **32**(2), pp. 149–162, 2012.
46. Z. Guo, Q. Li, J. You, D. Zhang, and W. Liu, "Local directional derivative pattern for rotation invariant texture classification," *Neural Computing and Applications* **21**(8), pp. 1893–1904, 2012.
47. N. Doshi and G. Schaefer, "A comparative analysis of local binary pattern texture classification," in *Visual Communications and Image Processing (VCIP)*, pp. 1–6, 2012.
48. Y. He, N. Sang, and R. Huang, "Local binary pattern histogram based texton learning for texture classification," in *IEEE International Conference on Image Processing, ICIP 2011*, pp. 841–844, 2011.
49. F. M. Khellah, "Texture classification using dominant neighborhood structure," *IEEE Transactions on Image Processing* **20**(11), pp. 3270–3279, 2011.
50. Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing* **19**, pp. 1657–1663, June 2010.
51. L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification," *Image and Vision Computing* **30**(2), pp. 86–99, 2012.
52. Y. He, N. Sang, and C. Gao, "Pyramid–based multi–structure local binary pattern for texture classification," in *Computer Vision ACCV 2010*, R. Kimmel, R. Klette, and A. Sugimoto, eds., *Lecture Notes in Computer Science* **6494**, pp. 133–144, Springer Berlin Heidelberg, 2011.
53. N. Rasiwasia and N. Vasconcelos, "Holistic context modeling using semantic co–occurrences," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 1889–1895, 2009.