A NOVEL FEATURE DESCRIPTOR BASED ON THE SHEARLET TRANSFORM

William Robson Schwartz[1], *Ricardo Dutra da Silva*[1], *Larry S. Davis*[2], *Helio Pedrini*[1]

[1]Institute of Computing
University of Campinas
Campinas, SP, 13084-971, Brazil

[2]Department of Computer Science
University of Maryland
College Park, MD, 20740, USA

ABSTRACT

Problems such as image classification, object detection and recognition rely on low-level feature descriptors to represent visual information. Several feature extraction methods have been proposed, including the Histograms of Oriented Gradients (HOG), which captures edge information by analyzing the distribution of intensity gradients and their directions. In addition to directions, the analysis of edge at different scales provides valuable information. Shearlet transforms provide a general framework for analyzing and representing data with anisotropic information at multiple scales. As a consequence, signal singularities, such as edges, can be precisely detected and located in images. Based on the idea of employing histograms to estimate the distribution of edge orientations and on the accurate multi-scale analysis provided by shearlet transforms, we propose a feature descriptor called Histograms of Shearlet Coefficients (HSC). Experimental results comparing HOG with HSC show that HSC provides significantly better results for the problems of texture classification and face identification.

***Index Terms*—** Feature Descriptors, Shearlet Transform, Histograms of Oriented Gradients, Histograms of Shearlet Coefficients, Histogram-based Feature Descriptor.

1. INTRODUCTION

Visual information contained in images is usually represented by low-level feature descriptors focusing on different types of information, such as color, texture, and shape. An adequate feature descriptor is able to discriminate between regions with different characteristics and it allows similar regions to be grouped together even when captured under noisy conditions. However, it is usually difficult to have a single feature descriptor adequate for many application domains; this has motivated researchers to develop a variety of feature extraction methods, such as histograms of oriented gradients [1], local binary patterns [2], geometric blur [3], co-occurrence matrices [4], and Gabor jets [5].

HOG, especially, has been applied successfully in a wide variety of applications. It captures shape information by the distribution of gradient magnitudes and directions. Regions of the image are divided into small connected cells from which histograms of gradient directions are computed. These histograms are then combined and used as feature descriptors. Among the main advantages of HOG are its robustness to local variations and its invariance to rotations smaller than the orientation bin interval.

Several descriptors closely related to HOG have been proposed in the literature [6–8]. Chuang et al. [6] proposed an augmented HOG applied to human detection by adding human shape properties. Also focused on human detection, Wang et al. [7] described an augmented version of HOG based on local binary patterns. Bosch et al. [8] proposed a pyramid of HOG, which captures the spatial layout by tiling the image into cells at multiple resolutions.

The amount of visual information captured by HOG might be increased by considering edge responses at multiple scales simultaneously, instead of a single scale. However, there is a need for better ways of estimating and locating edges than methods based on local gradient estimation due to instabilities at fine scales and for noisy images [9, 10].

Traditional methods for multi-scale signal analysis and time-frequency representation, such as wavelet and Gabor transforms, have been extensively used in image processing applications. However, these methods are not capable of efficiently capturing directional features from data due to their limited directional sensitivity. On the other hand, shearlet transforms can analyze signals defined not only at various orientations but also at multiple scales; they provide efficient mathematical and computational methods to address singularities such as image edges [11].

Based on the accurate multi-scale analysis provided by shearlet transforms and on the use of histograms to estimate the distribution of edge orientations, we propose a histogram-based feature descriptor called Histograms of Shearlet Coefficients (HSC). This method analyzes edges at multiple scales and orientations, differently from HOG. At each decomposition level provided by the shearlet transform, HSC estimates a histogram based on edge responses considering multiple orientations. Then, all histograms are concatenated and normalized to be used as feature descriptor.

The estimation of edge orientation based on histograms retains similar advantages to HOG, such as the invariance within the rotation angle. However, in contrast to HOG, HSC analyzes the image at multiple scales, which provides a more comprehensive description of the edges present in the image. In addition, the shearlet transform is robust to noise and is accurate at fine scales, in contrast to methods based on local gradient estimation [9, 10], such as HOG. Experimental results will show that HSC outperforms HOG for problems of texture classification and face identification.

2. SHEARLET TRANSFORMS

Signal representations in both the time and frequency domains, such as wavelet and Gabor transforms, have received significant attention in the last decades. However, a disadvantage of these methods is their limited directional sensitivity - their limited capability of representing directional features.

Some variations of wavelet transforms [12] have been proposed to overcome these limitations, such as directional wavelets, bandelets, brushlets, contourlets, phaselets, directionlets, ridgelets and curvelets. Shearlets [13] possess important properties, for instance, they are optimal in approximating 2D smooth functions with discontinuities along C^2-curves and they form an affine system.

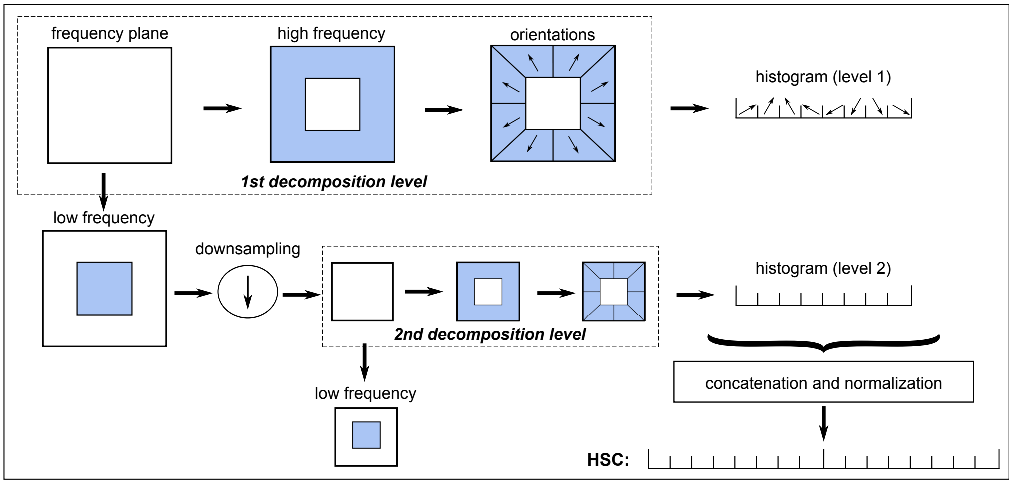


Fig. 1: Feature extraction for a two level shearlet decomposition considering eight orientations. A histogram containing the same number of bins as the number of orientations considered by the shearlet transform is estimated for each level of decomposition (the spatial frequencies histogrammed at each level are marked in blue). These histograms describe the distribution of edge orientations at the given scales. The entry for each histogram bin is obtained by adding up the absolute value of the shearlet coefficients at the corresponding orientation. The histograms of shearlet coefficients (HSC) is obtained by the concatenation of the histograms estimated for each decomposition level, followed by $L2$-norm normalization.

Shearlet transforms provide a general framework for analyzing and representing data with anisotropic information at multiple scales. As a consequence, signal singularities, such as edges, can be precisely detected and located in images. The continuous shearlet transform for an image f is defined as the mapping

$$f \to \mathcal{SH}_\psi f(a, s, t) = \langle f, \psi_{a,s,t} \rangle \tag{1}$$

where ψ is a generating function, $a > 0$ is the scale parameter, $s \in \mathbb{R}$ is the shear parameter, $t \in \mathbb{R}^2$ is the translation parameter, and the analyzing elements $\psi_{a,s,t}$ (shearlet basis functions) are given by

$$\psi_{a,s,t}(x) = |\det M_{a,s}|^{-\frac{1}{2}} \psi(M_{a,s}^{-1}(x - t)) \tag{2}$$

where $M_{a,s} = \begin{bmatrix} a & s\sqrt{a} \\ 0 & \sqrt{a} \end{bmatrix}$

Each element $\psi_{a,s,t}$ has frequency support on a pair of trapezoids at several scales, symmetric with respect to the origin, and oriented along a line of slope s. Therefore, the shearlets $\psi_{a,s,t}$ form a collection of well-localized waveforms at various scales a, orientations s and locations t. Figure 2 illustrates the frequency support of the shearlets. The continuous shearlet transform can be discretized by properly sampling the scale, shear and translation parameters. An efficient discrete shearlet transform scheme, based on a Laplacian pyramid combined with appropriate shearing filters, was proposed in [14]. This scheme is used in our method.

3. HISTOGRAMS OF SHEARLET COEFFICIENTS

Our proposed Histograms of Shearlet Coefficients (HSC) performs a multi-scale decomposition of the image obtained by applying the shearlet transform to capture visual information provided by edges detected not only at different orientations but also at multiple scales.

The use of shearlet transforms is motivated by the fact that they provide a multi-scale analysis of the image especially adapted to capture the geometry of edges, in contrast to methods based on local gradient operators, such as HOG descriptors, which have only limited capability to detect the orientations of singularities, affecting the accuracy of edge detection, especially in the presence of noise [9]. In practical numerical computations, the gradient operator is approximated by a local finite difference. This is a significant source of inaccuracies in the estimation of the edge orientation, especially at fine scales [10].

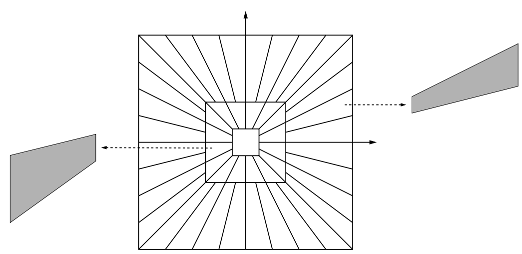


Fig. 2: Support of shearlets $\psi_{a,s,t}$ in the frequency domain, which is partitioned into trapezoidal shaped tiles.

Easley et al. [11] have shown that shearlet coefficients of large magnitude will come from edges. More precisely, given an edge point x with orientation s, the shearlet coefficient corresponding to x is negligible unless its orientation corresponds to s. Therefore, shearlet coefficients provide edge responses at a given scale and orientation. Our proposed HSC is based on the magnitude of shearlet coefficients at different scales and orientations.

For each decomposition level, we estimate a histogram with the same number of bins as the number of orientations considered by the shearlet transform and the entry at each bin is computed as the sum of the absolute values of the shearlet coefficients

$$H_l(s) = \sum |\mathcal{SH}_\psi f(a, s, t)| \tag{3}$$

where $H_l(s)$ denotes the s-th bin of the histogram computed for the l-th decomposition level (the s-th bin corresponds to edge responses at a specific orientation). Finally, the histograms computed for all levels are concatenated and $L2$-norm normalization is employed, resulting in the HSC, which is used to describe the image. This process, considering only two levels of decompositions and eight orientations, is illustrated in Figure 1.

Regarding computational cost, similarly to HOG, integral images [15] can be employed to compute HSC. For a given image, shearlet coefficients are extracted for all decomposition levels and orientations. Thus, for each pair of decomposition level and orientation, an integral image is computed. That supports efficient computation of the absolute values of the shearlet coefficients.

4. EXPERIMENTAL RESULTS

In this section, we evaluate HSC using the applications of texture classification and face identification. We evaluate the performance obtained by HSC with respect to both the number of decomposition levels and number of orientations[1]. Results are compared to HOG.

For texture classification, we used the UMD texture dataset [16], composed of samples of 1280×960 pixels distributed into 25 classes with 40 samples each. A maximum of 6 levels of decomposition are used. A linear SVM is employed to perform classification, where 10 samples for each class are considered for training and the remaining for classification, chosen randomly. The experiment is repeated 50 times and the average classification rate is reported.

For face identification, we considered the FERET dataset [17] and apply the method proposed by Schwartz et al. [18], using either HOG or HSC instead of the feature descriptors considered in that work. For this experiment, the samples are cropped and rescaled to 110×110 pixels and at most 5 levels of decomposition are applied. We followed the protocol described in [18] using set *fa* as gallery, composed of 1196 subjects, and set *fb*, with 1195 images, as probe.

Since the original HOG descriptors are obtained at a single scale and descriptors obtained by HSC consider multiple scales (through multiple decomposition levels performed by the shearlet transform), we adapt the HOG descriptors so that we are able to perform a fair comparison between both descriptors. This descriptor is obtained by extracting HOG from low frequency images in the Laplacian pyramid decomposition, using the same number of decomposition levels used by HSC. Once extracted for each level, HOG descriptors are concatenated and the feature vector is normalized using $L2$-norm. Therefore, we have both the original HOG (when only one level of decomposition is considered) and its extension, which is obtained from multiple scales of the image.

Table 1 compares HOG with HSC for texture classification and face identification varying the number of decomposition levels and orientations. In this experiment, for a given decomposition level, the best result based on various numbers of orientations is reported. For instance, the result shown in the first row of the table for texture classification performed by HOG (0.638), reports the best performance obtained considering each of the 4, 8, 16, or 32 orientations with 1 level of decomposition. An analogous procedure is employed for the orientations. HSC achieves superior results in all tests. Furthermore, the original HOG, which considers only a single scale, obtains a classification rate of 0.638 for texture classification and a recognition rate of 0.094 for face identification. Such values are much lower than the best results obtained by the proposed HSC or even HOG with multiple levels of decomposition, as shown in the last row of Table 1.

Figure 3 displays contour plots with isolines indicating the same classification and recognition rates for texture classification and face identification, respectively. In these plots, the x-axis shows the decomposition levels and y-axis shows the number of orientations. A

[1]The source code for the proposed HSC descriptor is available at http://www.liv.ic.unicamp.br/~wschwartz/softwares.html .

Table 1: Results obtained for texture classification and face identification using HOG and HSC considering several levels of decompositions and orientations.

		Texture Classification		Face Identification	
		HOG	HSC	HOG	HSC
levels	1	0.638	0.653	0.094	0.236
	2	0.698	0.717	0.282	0.461
	3	0.720	0.753	0.365	0.636
	4	0.725	0.786	0.455	0.739
	5	0.711	0.803	0.552	0.764
	6	0.707	0.810	-	-
orientations	4	0.542	0.793	0.468	0.559
	8	0.701	0.810	0.545	0.682
	16	0.724	0.810	0.552	0.747
	32	0.725	0.808	0.505	0.764
best results		0.725	**0.810**	0.552	**0.764**

dashed arrow is added to each plot indicating the direction in the plane whose results improve the most. Results obtained with HSC tend to be improved when both the number of orientations and decomposition levels increase. However, results obtained with HOG might decrease with these parameters. This behavior is consistent with the discussion in Section 3, that is, HOG descriptors have diminished ability to detect edge orientations in the presence of noise, especially at fine scales.

Results reported in Table 1 and Figure 3 were obtained using only one descriptor of HOG or HSC extracted from the whole image. Similarly to previous research [1, 18, 19], results are expected to be further improved by first decomposing the input image into multiple cells and then using the concatenated features extracted from each cell to describe the images. For example, by splitting the input image into four cells and extracting features from them, the classification rates by HSC for texture classification improved from 0.810 to 0.856 and the recognition rates for face identification increased from 0.764 to 0.827.

5. CONCLUSIONS

In this paper, we introduced a novel descriptor called histograms of shearlet coefficients. It estimates a histogram based on edge responses in several orientations for each level of decomposition provided by the shearlet transform computed over an image region.

Experimental results shown that HSC provides better results for face identification and texture classification when compared with results obtained by HOG. This is due to the multi-scale analysis and the accurate edge localization and orientation provided by the shearlet transform.

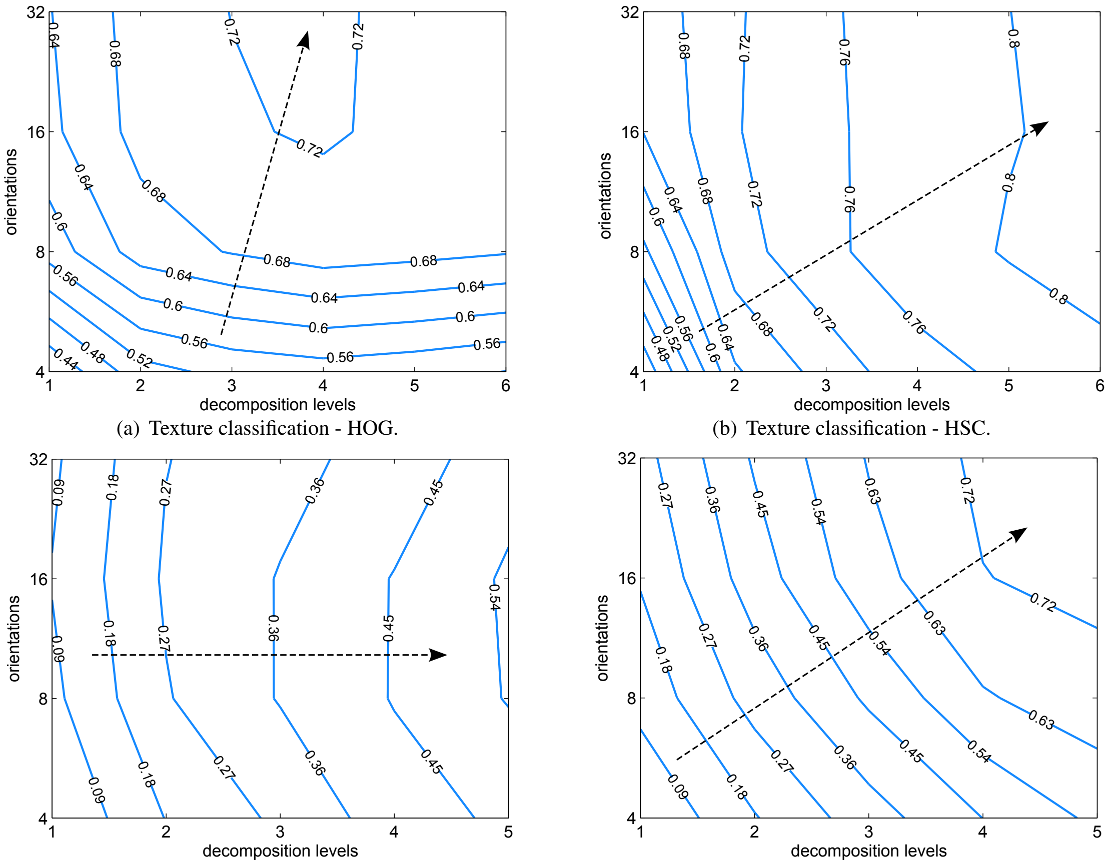
6. ACKNOWLEDGEMENTS

The authors are thankful to FAPESP, CNPq and CAPES for the financial support. This research was partially supported by ONR grant N000140910044.

7. REFERENCES

[1] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *CVPR*, 2005.

[2] T. Ojala, M. Pietikinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Featured Distributions," *Pattern Recognition*, 1996.



(a) Texture classification - HOG.

(b) Texture classification - HSC.

(c) Face identification - HOG.

(d) Face identification - HSC.

Fig. 3: Contour plot displaying isolines with results achieved using HOG and HSC for texture classification and face identification as functions of decomposition levels and number of orientations. The dashed arrows indicate in which directions (according to the plane defined by decomposition levels and orientations) the results improve the most.

[3] A.C. Berg and J. Malik, "Geometric Blur for Template Matching," in *CVPR*, 2001.

[4] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, 1973.

[5] L. Wiskott, J. Fellous, N. Kuiger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *PAMI*, 1997.

[6] C.-H. Chuang, S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao, "Monocular Multi-human Detection using Augmented Histograms of Oriented Gradients," in *ICPR*, 2008.

[7] X. Wang, T.X. Han, and S. Yan, "An HOG-LBP Human Detector with Partial Occlusion Handling," in *ICCV*, 2009.

[8] A. Bosch, A. Zisserman, and X. Munoz, "Representing Shape with a Spatial Pyramid Kernel," in *ACM International Conference on Image and Video Retrieval*, 2007.

[9] K. Guo, D. Labate, and W.Q. Lim, "Edge Analysis and Identification using the Continuous Shearlet Transform," *Applied and Computational Harmonic Analysis*, 2009.

[10] S. Yi, D. Labate, G.R. Easley, and H. Krim, "Edge Detection and Processing using Shearlets," in *ICIP*, 2008.

[11] G.R. Easley, D. Labate, and F. Colonna, "Shearlet-Based Total Variation Diffusion for Denoising," *TIP*, 2009.

[12] J.-L. Starck, F. Murtagh, and J.M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, May 2010.

[13] W.-Q. Lim, "The Discrete Shearlet Transform: A New Directional Transform and Compactly Supported Shearlet Frames," *TIP*, 2010.

[14] G. Easley, D. Labate, and W.-Q. Lim, "Sparse Directional Image Representations using the Discrete Shearlet Transform," *Applied and Computational Harmonic Analysis*, vol. 25, pp. 25–46, 2008.

[15] P. Viola and M. Jones, "Robust Real-time Object Detection," *IJCV*, 2002.

[16] Y. Xu, S.-B. Huang, H. J., and C. Fermuller, "Combining Powerful Local and Global Statistics for Texture Description," in *CVPR*, 2009.

[17] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *PAMI*, 2000.

[18] W.R. Schwartz, H.M Guo, and L.S. Davis, "A Robust and Scalable Approach to Face Identification," in *ECCV*, 2010.

[19] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human Detection Using Partial Least Squares Analysis," in *ICCV*, 2009.