

# A projective invariant for textures

Yong Xu

Department of Computer Science  
South China University of Technology

Hui Ji and Cornelia Fermüller

Department of Computer Science  
University of Maryland, College Park

## Abstract

*Image texture analysis has received a lot of attention in the past years. Researchers have developed many texture signatures based on texture measurements, for the purpose of uniquely characterizing the texture. Existing texture signatures, in general, are not invariant to 3D transforms such as view-point changes and non-rigid deformations of the texture surface, which is a serious limitation for many applications. In this paper, we introduce a new texture signature, called the multifractal spectrum (MFS). It provides an efficient framework combining global spatial invariance and local robust measurements. The MFS is invariant under the bi-Lipschitz map, which includes view-point changes and non-rigid deformations of the texture surface, as well as local affine illumination changes. Experiments demonstrate that the MFS captures the essential structure of textures with quite low dimension.*

An ideal image texture descriptor for describing, classifying and recognizing textures should be able to capture essential perceptual properties of the texture structure, and it also should be invariant to environmental changes, such as changes in view-point, luminance and geometry of the underlying surface. In the Computer Vision literature, the search for invariances started in the nineties [17], when researcher dug up the mathematical literature on algebraic and geometric invariants. Great importance has been given to the study of quantities which are invariant to the viewpoint from which the image is taken. A number of projective invariants [13] have been found, which are defined on feature sets of points and lines and planar curves [17], and they have been used for object recognition and calibration. However, none of these descriptors provides a high-level characterization of textures.

Numerous texture descriptors have been proposed. Most of them are either statistics-based or filter-based ([1, 4, 5, 15]), which makes them sensitive to changes in viewpoint. Lazebnik, Schmid and Ponce ([6]) proposed a texture representation which is invariant to view-point changes in a weak form (it is *locally* invariant to *affine* transforms).

Here we introduce a novel texture descriptor called the MFS (multifractal spectrum vector), which is based on fractal geometry theory. The MFS is globally invariant under the *bi-Lipschitz transform*, a very general transform which includes perspective transforms (viewpoint changes) and smooth texture surface deformations. Furthermore, the MFS has extraordinary low dimension (26 using one feature and three times that much using three features as in our implementation here for a typical  $640 \times 480$  image texture). We demonstrate with a number of experiments on synthesized and real image textures that, first the MFS in practice is very robust to 3D transforms and non-rigid smooth transforms. Second, its performance is similar to the top methods in traditional texture retrieval and classification on standard texture datasets, but by using much lower dimensional feature vectors.

Fractal geometry has been used before in the description of textures ([2, 14]) and texture segmentation ([8, 9, 10]). However, the invariance of the fractal dimension to bi-Lipschitz maps has not been utilized in the vision community. Furthermore, existing approaches either simply compute a scalar number, which does not provide enough information for a good characterization of the texture; or they utilize the multi-fractal-spectrum for local feature description, in which there is no global invariance to bi-Lipschitz maps. Moreover, the fractal dimension was defined only on the image intensity, which makes it sensitive to changes in luminance. Intuitively, the MFS (multi-fractal-spectrum) is the extension of the fractal dimension. The MFS encodes the fractal dimension of every level set under a level set categorization of image points. Different categorizations of image points lead to different MFS vectors. The categorizations in our MFS approach are based on the *density function* defined on (functions of) the image intensity, and this makes it robust to luminance variations.

The paper is organized as follows. Section 2 gives a brief introduction to fractal geometry and multi-fractal spectrum theory. Section 3 presents our MFS texture descriptor and its invariance to spatial bi-Lipschitz transforms and local affine illumination changes. Section 4 provides experiments on texture retrieval and classification and a comparison to

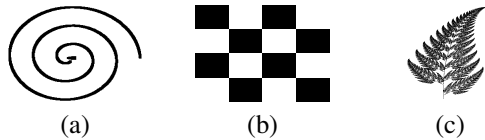


Figure 1. Fractal dimension  $D$  in 2d space. (a) Smooth spiral curve with  $D = 1$ . (b) The checkerboard with  $D = 2$ . (c) Fractal fern leaf with  $D \approx 1.7$ .

other methods. Section 5 presents the conclusions and future work.

## 1. Fractal Theory and the MFS

We first give a brief introduction to fractal theory. In the later eighties it has been realized that irregular objects provide a much better representation of many natural phenomena than do the regular objects of classical geometry. Fractal geometry was developed, which provides a general framework for studying irregular sets as well as regular sets.

The fractal dimension is the key quantity in the study of fractal objects. Fundamental to the fractal dimension is the concept of "measurement at scale  $\delta$ ". For each  $\delta$ , we measure an object in a way that ignores irregularity of size less than  $\delta$ , and we see how these measurements behave as  $\delta$  goes to 0. Scientists found that most of the natural phenomena satisfy the power law, which states that the the estimated quantity (for example the length of a coastline [7]), is proportional to  $(\frac{1}{\delta})^D$  for some  $D$ . For most natural objects,  $D$  is almost the same for small scales  $\delta$ . Thus we can compute its limit, which is called the *fractal dimension*. For the case of an irregular point set defined on  $R^2$ , the fractal dimension of the set  $E$  is defined as

$$\dim(E) = \lim_{\delta \rightarrow 0} \frac{\log N(\delta, E)}{-\log \delta}, \quad (1)$$

where  $N(\delta, E)$  is the smallest number of sets of diameter less than  $\delta$  that cover  $E$ .

Intuitively, the fractal dimension gives a global description of how complex or how irregular a geometric object is. The fractal dimension  $D$  of any object in 2D space is in the range of  $1 \leq D \leq 2$ . Any smooth curve has a fractal dimension of 1, and a completely filled rectangle has a fractal dimension of 2, which are the same as their integer topological dimensions. Irregular sets have a fractional dimension between 1 and 2. Indeed, most man-made geometric objects have an integer fractal dimension  $D$ , while most objects in nature have a fractional fractal dimension. An illustration is given in Fig. 1.

The question arises then whether natural textures are objects with fractional dimension information. Let's check some typical texture first. A typical grass texture image is shown in Fig. 2 (a). If we represent the image as a 3D-surface in Fig. 2 (b) by taking the image intensity as the

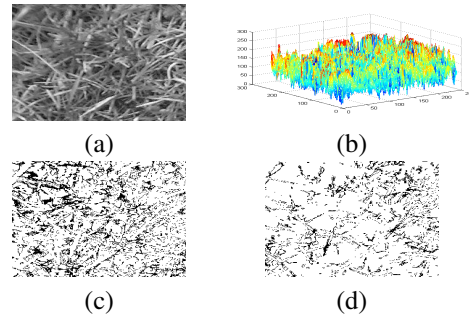


Figure 2. (a) Original texture image. (b) 3D surface visualization of the image with  $D = 2.79$ . (c) Black-white image obtained by thresholding intensity values between 100 and 120, its fractal dimension  $D = 1.67$ . (d) Black-white image obtained by thresholding intensity values between 80 and 100, its fractal dimension  $D = 1.49$ .

height, its box fractal dimension in 3D is 2.79.

So, the answer would be yes: nature textures do encode fractal dimension information. However the fractal dimension alone, as defined above, does not provide a rich description. Physicists and Applied Mathematicians have developed multifractal analysis as an extension to classical fractal analysis for more powerful descriptions. The concept of the fractal dimension is extended to the concept of the MFS (Multifractal spectrum) as follows: First define a point categorization on the object function according to some criteria. Then the fractal dimensions are calculated for every point set from this categorization. A such defined MFS vector gives a rich description of the inherent texture structure.

To give a demonstration, we take a simplistic categorization of image points based on the gray values of the pixels. The image pixels are categorized by their intensities. Then for every pixel set, we obtain a black-and-white image by assigning 0 to pixels in this set and 1 to all other pixels. The collection of the fractal dimensions for all the black-and-white images is called an MFS vector. Two examples of such obtained images are shown in Fig. 2 (c) and Fig. 2 (d). Clearly these are fractal objects with fractional fractal dimensions.

As can be easily understood, different criteria for defining the point categorization lead to different fractal spectra. The simplistic approach demonstrated above has serious drawbacks in practice. A categorization based on image intensity is numerically unstable in its computation, and it is sensitive to illumination changes. Therefore, we will use another categorization method which is based on the idea of the "density function". The density function can be defined on different functions of the image intensity. Intuitively, the density function of a quantity defines the change of that quantity over scale. Our categorization leads to an MFS which is robust to illumination changes, and there also ex-

ists an efficient and robust algorithm for computing it [3].

## 2. MFS for Textures

The MFS is the vector of the fractal dimensions of some categorization of the image. Using the idea of the local density function in the categorization has the advantage that we can compute the fractal spectra in a robust and efficient way.

**General model of the MFS for texture** Let  $\mu$  be a finite Borel regular measure on  $R^2$ . (The details of  $\mu$  for texture images will be discussed in the practical algorithm.) For  $x \in R^2$ , denote  $B(x, r)$  the closed disc with center  $x$  and radius  $r > 0$ . It is easy to see that  $\mu(B(x, r)) \rightarrow 0$  and  $\log \mu(B(x, r)) \rightarrow \infty$  when  $r \rightarrow 0$ . Then the *local density function* of  $x$  is defined as

$$D(x) = \lim_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r}. \quad (2)$$

For  $\alpha \in R$ , define

$$E_\alpha = \{x \in R^2 : D(x) = \lim_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r} = \alpha\}. \quad (3)$$

That is,  $E_\alpha$  is the set of all image points  $x$  with local density  $\alpha$ . Usually this set is irregular and has a fractional fractal dimension  $dim(E_\alpha)$ . Thus we obtain a point categorization  $\{E_\alpha : \alpha \in R\}$  of the image with an MFS denoted as

$$f(\alpha) = \{dim(E_\alpha) : \alpha \in R\}. \quad (4)$$

The MFS  $f(\alpha)$  defined by (4) is the natural extension of the concept of fractal dimension, but difficult to estimate in practice. Here we adopt another equivalent expression for  $f(\alpha)$  based on generalized moment measurements. (See [3] for more details.) The new expression of the MFS is both efficient and robust in practical computations.

**Practical algorithm for computing the MFS.** First, we partition  $R^2$  into  $r$ -mesh squares (non-overlapping boxes of length  $r$ ). Then we define the  $q$ -th moment of a measure  $\mu$  as

$$M_r(q) = \sum \mu(B(x, r))^q, \quad (5)$$

where the sum is over the  $r$ -mesh squares  $\{B(x, r)\}$  for which  $\mu(B(x, r)) > 0$ . For a series of  $r = 1, 2, \dots, n$ , we have corresponding measurements  $M_r(q)$ . Then the power law behavior of  $M_r(q)$  is identified by the number  $\beta(q)$  which is defined as:

$$\beta(q) = \text{Slope of } \log(M_r(q)) \text{ vs } \log(r). \quad (6)$$

In other word,  $\beta(q)$  is the slope of the line which best fits the data sets  $\{\log(M_r(q)), \log(r)\}$ . It is shown in [3] that the MFS and  $\beta(q)$  are related to each other by a Legendre transform as

$$f(\alpha(q)) = q\alpha(q) - \beta(q), \quad (7)$$

where

$$\alpha(q) = \frac{d\beta(q)}{dq}. \quad (8)$$

Using equations (7) and (8) we estimate the MFS. The reliability of  $f(\alpha(q))$  is measured by the residual of fitting  $\beta(q)$  in (6). This residual is used in our weighting scheme in texture retrieval and classification.

Intentionally we haven't given the definition of the measurement function  $\mu(B(x, r))$  yet. The first approach is to work directly on the intensity domain. Define  $\mu(B(x, r))$  as:

$$\mu(B(x, r)) = \int_{B(x, r)} (G_r * I) dx \quad (9)$$

where  $"*"$  is the 2D convolution operator and  $G$  is a Gaussian smoothing kernel with variance  $r$ . In other words  $\mu$  is the average intensity value inside the disc  $B(x, r)$ . This results in the definition of the density of the intensity function, which describes how the intensity at a point changes over scale (as we change the size of the neighborhood). Finally, the corresponding MFS vector encodes the fractal dimension for multiple values of the density of the intensity. The algorithm is summarized below.

---

### Algorithm 1 Computation of the MFS.

---

- 1) Compute  $M_r(q)$  from (5) and (9).
  - 2) Estimate  $\beta(q)$  from (6) by fitting  $M_r(q)$  linearly.
  - 3) Compute  $f(\alpha)$  from (7) and (8).
- 

**Other MFS vectors.** In practice the measurement function  $\mu(B(x, r))$  is not very robust to large illumination changes. But we can define the density function on other quantities. One can imagine many meaningful definitions for  $\mu(B(x, r))$ , such that the corresponding MFS is less effected by illumination changes. One choice is to define  $\mu(B(x, r))$  on the energy of the gradients. Consider  $\{f_k, k = 1, 2, 3, 4\}$  to be four directional differential operators along the vertical, horizontal, diagonal and anti-diagonal directions. Then we define the measurement function  $\mu(B(x, r))$  for the image texture  $I$  as:

$$\mu(B(x, r)) = \left( \int_{B(x, r)} \sum_k (f_k * (G_r * I))^2 dx \right)^{\frac{1}{2}}. \quad (10)$$

Another meaningful choice for  $\mu(B(x, r))$  is the sum of the Laplacians of the image inside  $B(x, r)$ , i.e

$$\begin{aligned} \mu(B(x, r)) &= \int_{B(x, r)} |\nabla^2 (G_r * I)| dx \\ &= \int_{B(x, r)} \left| \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) G_r * I \right| dx. \end{aligned} \quad (11)$$

Different definitions of  $\mu(B(x, r))$  lead to different MFS vectors which capture different aspects of the texture's structure. Multiple MFS vectors could and should be combined to better describe the texture.

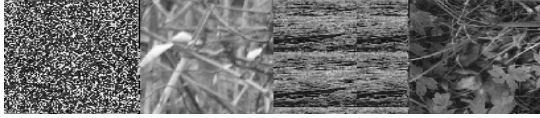


Figure 3. Four textures: cloth, tree, wood, grass from left to right

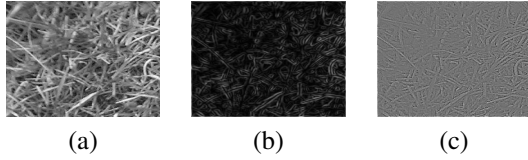


Figure 4. (a) Intensity image of the grass texture. (b) Energy of the first order derivatives as defined in Eqn. (10). (c) Laplacian

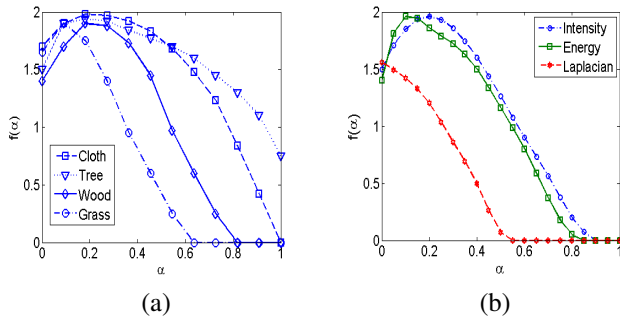


Figure 5. (a) The MFS of the intensity for the four textures shown in Fig. 3. (b) The MFS for all three measurement functions shown in Fig. 4

Fig. 5 (a) shows the MFS (based on (10)) for the four textures in Fig. 3. It demonstrates that the MFS vectors of the different textures are significantly different. Fig. 5 (b) shows the three MFS vectors based on the three measurement functions  $\mu$  described above. From this and other textures we found that the density of the intensity and the density of the edge energy give similar MFS vectors, whereas the density of the Laplacian leads to a significantly different MFS.

### 2.1. Invariance of the MFS to geometric and luminance deformations

**Spatial invariance.** A very attractive property of the MFS is its invariance under the so-called *bi-Lipschitz* transform. A function  $f : R^2 \rightarrow R^2$  is called a bi-Lipschitz transform, if there exist two constants  $c_1 > 0$ ,  $c_2 > 0$  such that for any  $x, y \in R^2$ ,

$$c_1 \|x - y\| < \|f(x) - f(y)\| < c_2 \|x - y\|, \quad (12)$$

where  $\|x - y\|$  is the Euclidean metric between points  $x$  and  $y$ . Basically, any smooth transform is a bi-Lipschitz trans-

form. It is a very general mapping which includes translation, rotation, perspective transformation, and texture warping on regular surfaces. Then we have the following theorem (See Appendix A for the proof).

**Theorem 1.** *The MFS is invariant under any spatial bi-Lipschitz transform.*

Thus, the MFS not only is invariant to perspective transformations, but also to non-rigid deformations.

Let us stress that the invariance of the MFS to bi-Lipschitz maps has been proven for images of infinite resolution. In practice we are dealing with finite resolution images. Nevertheless, we found that the MFS is very robust to perspective transforms and surface warpings, even for low resolution images. Usually four to five levels of resolution (defined by radius  $r$ ) are sufficient for a good estimation of the MFS.

Two examples are given: Fig. 7 shows the same plant texture on a plane seen from different view-points. It can be seen that the corresponding MFS vectors in Fig. 8 (a) are nearly the same.

Fig. 6 shows six images of a tree texture on general smooth surfaces under perspective projection. Their MFSs are given in Fig. 8 (b). Although there is some self-occlusions in these images, the MFS are nearly the same.

**Illumination invariance.** Another important factor to be considered is the illumination. In theory it can be proven that the MFS (for all three definitions of measurement function  $\mu$ ) is illumination invariant.

In practice we don't have illumination variance. We can show, however, that for the density defined on first order derivatives and the Laplacian the MFS is locally invariant to affine changes in illumination. For the density defined on the intensity we have invariance to multiplicative changes. Consider, a brightness change due to a constant being added to each image pixel. This will not affect  $\mu(B(x, r))$  (as defined in 10 and 11), because these measures are based on pixel differences. Next, consider a multiplicative change of each pixel value by a constant. This will just multiply  $M(r, q)$  by a constant. Thus, this change will also not affect the estimation of  $\beta(q)$ , because the slope of the line fitting  $\log(M(r, q))$  is invariant to this multiplication. Therefore, the MFS is invariant to affine (multiplicative) changes in illumination. Even non-linear illumination changes do not much effect the MFS, because the fractal dimension is quite robust to small variations of its point set.

Fig. 9 shows nine real texture images taken by a family digital camera. The three images in each row depict the same scene from different view-points under different illumination conditions. The computed MFS is shown in Fig. 10. As can be seen the MFSs of different textures are significantly different, while the MFSs of the same texture under different view-points are almost identical. Fig. 11

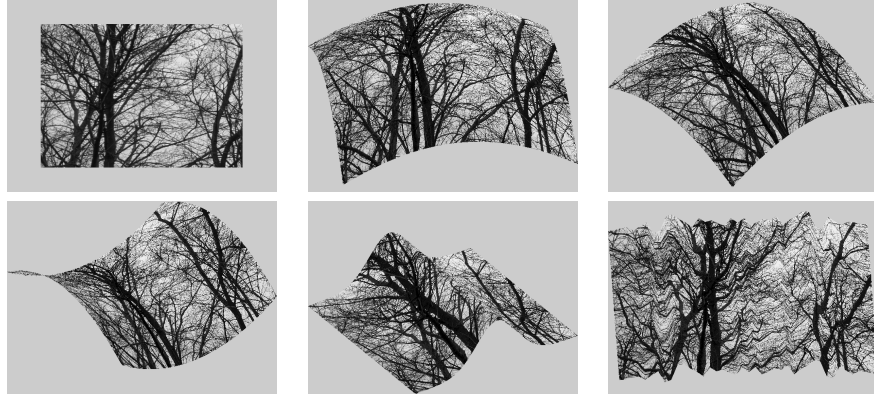


Figure 6. Perspective images of texture *tree* on different general smooth surfaces.



Figure 7. Six perspective texture images of the *foliage* texture.

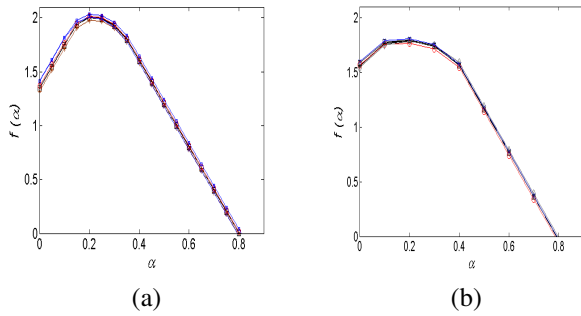


Figure 8. (a) The MFS vector of intensity measurement for six perspective views of the *foliage* textures in Fig. 7. (b) The MFS vectors of intensity measurement for the tree texture warped on six different surfaces as shown in Fig. 6.

shows eight views of a cloth texture on the moving arm of a person. The illumination is not controlled. The corresponding MFSs are shown in Fig. 12, again demonstrating the robustness under geometric and luminance deformations.

## 2.2. Comparison to other texture descriptors

The most popular global texture descriptor is the histogram. It could be interpreted as a descriptor defined on

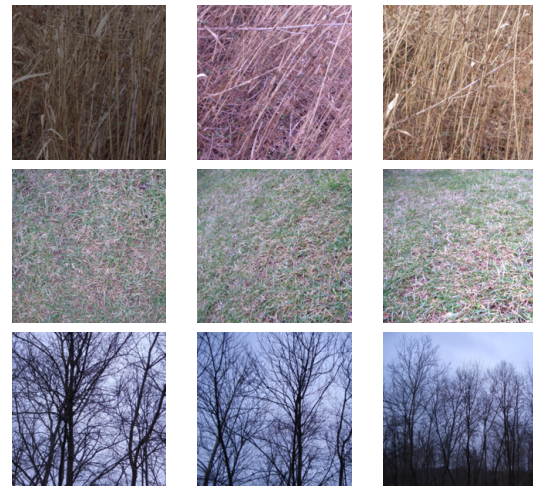


Figure 9. Three real 3D texture images under different lighting condition and view points.

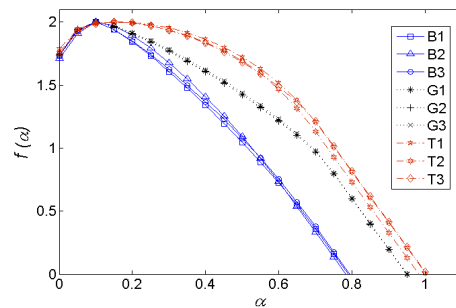


Figure 10. The MFS of intensity measurement for the nine texture images in Fig. 9. B1,B2,B3 are bulrushes, G1, G2, G3 are grasses and F1, F2, F3 are trees.

some categorization of the image, such as for example a categorization of pixels based on intensity thresholding. The descriptor for each category is simply the number of pixels



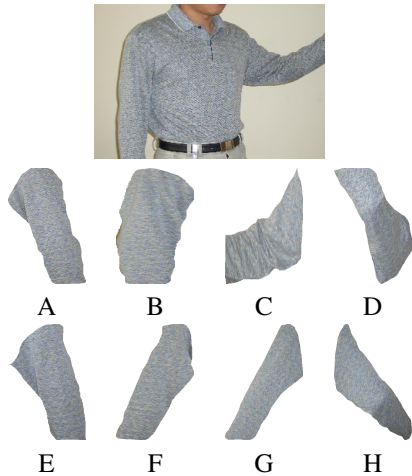


Figure 11. The arm part of the human gesture video under different luminance conditions and views.

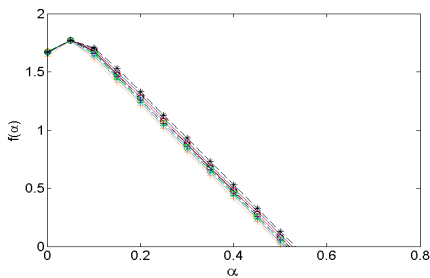


Figure 12. The MFS of intensity measurement corresponding to Fig.11

in the category. The drawback of such a description is that the spatial information of how pixels are distributed is lost. In contrast, the fractal dimension encodes the spatial information (as irregularity). The intensity-based histogram is quite robust to small changes, but not invariant to perspective transformations, and it is very sensitive to changes in illumination. The modified texton-based histograms could reduce the effects of illumination but in trade-off to larger sensitivity to global geometric transforms. In contrast, the MFS is capable of incorporating various low-level local feature measurements, such as the Laplacian, the energy, or SIFT features without sacrificing global spatial invariance.

Another type of texture descriptor utilizes clusters of locally spatial invariant descriptors of textons (elements of textures). In the following section we compare our approach against the sophisticated texture representation of Lazebnik et al [6], which we call the LA-method. The basic idea of the LA-method is to characterize the texture by clusters of elliptic regions. The elliptic regions are localized with the Harrison (H) and Laplacian (L) operators. The ellipses are transformed to circles. Thus this descriptor is locally in-

variant to affine transforms. Each region is represented by two descriptors: one encoding the histogram of the intensity distribution in a neighborhood (S); the other encoding the histogram of edges in a neighborhood (R) (a variation of SIFT features).

The LA-method essentially represents an image texture by its pattern of texture elements. It doesn't have global spatial invariance, because the pattern itself is the 2D projection of a 3D pattern. Different view-points lead to different patterns. The dimension of the LA texture descriptor is very high (thousands for a typical image texture). Furthermore, the LA-method requires sophisticated preprocessing to compute the texture elements and k-mean clustering to construct a good texture descriptor. In comparison, the MFS is simple and straightforward to compute as it neither relies on feature detection nor on clustering techniques.

### 3. Experimental Evaluation

We evaluated the performance of the MSF descriptor on classic texture *retrieval* and *classification* and compared it to two top methods: the LA-method (described before) and VZ-Joint method ([16]). Varma and Zisserman's VZ-Joint method is a non-invariant algorithm which uses a dense set of 2D textons. The descriptor is one-dimensional texton histogram encoding the joint distribution of all pixel values in the neighborhood. For the comparison we use the dataset provided in ([6]), which consists of 1000 uncalibrated, unregistered images of size  $640 \times 480$ : 25 different textures each with 40 samples (available at <http://www-cvr.ai.uiuc.edu/ponce>).

Within each class, significant viewpoint changes and scale differences are present, and the illumination conditions are uncontrolled. Additional sources of variability include non-planarity of the textured surface, significant non-rigid deformations between different samples of the same class, inhomogeneities of the texture pattern, and viewpoint-dependent appearance variations. (For details see [6].)

We used three MFS vectors (as defined by the measurement functions 9, 10 and 11), and for each vector we computed 26 values. There is no clustering in our method. As distance function between the MFS vectors we used the weighted  $L_1$  norm. The weight coefficients are from the residuals in the computation of  $\beta(q)$ .

Retrieval and classification are implemented as described in [6]. Briefly, for retrieval, given a query image we selected other images from our database in an increasing order of the distance, i.e., from the most similar to the least similar. Each image in the database is used as a query image once, and the performance is summarized as a plot of average recall vs. the number of retrievals. For classification we used the nearest-neighbor technique. The training set is selected as a fixed size random subset of the class, and

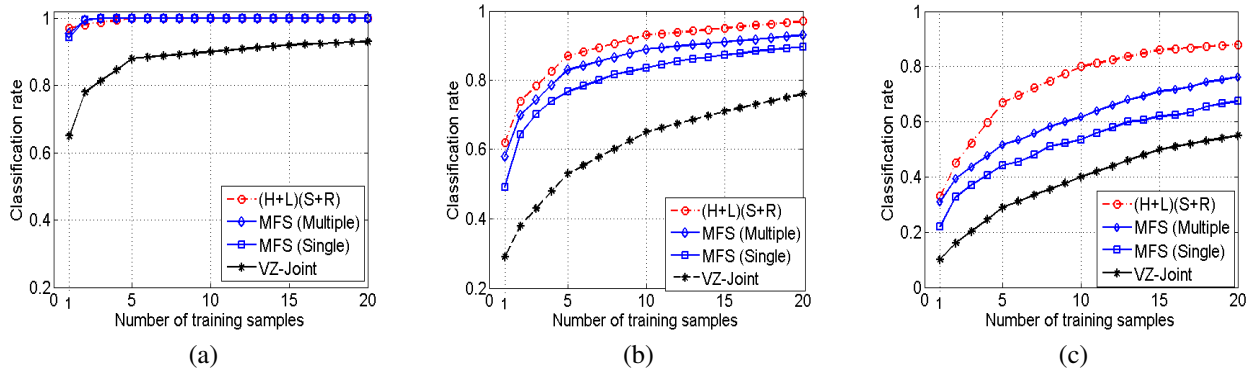


Figure 14. Classification rate vs. number of training samples. Three methods are compared: the MFS method, the (H+L)(S+R) method in [6] and the VZ-Joint method in [16]. (a) Classification rate of the best class. (b) Mean classification rate for all 25 classes. (c) Classification rate of the worst class.

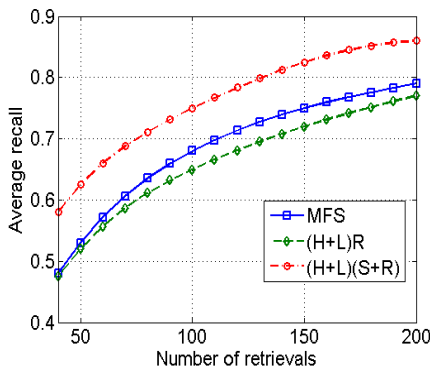


Figure 13. Retrieval curves for the Ponce database by our method and the methods in [6].

all remaining images are used as test set. The reported classification rate is the average over 200 random subsets. The results for the methods [6] and [16] are from the paper [6].

Fig. 13 presents a comparison of our method to the techniques in [6] for retrieval. Our method performs better than the "(H+L)R" channel of the LA-method, but not better than the "(H+L)(S+R)" channel. The reason is that the MFS is not as robust as [6]'s methods to illumination changes. [6]'s robustness is due to the use of only edge and corner information. In addition there is parameter-dependent clustering. It is worth stressing that the dimension of the MFS is significantly smaller than the ones of the methods in [6] (seventies vs thousands).

Fig. 14 evaluates classification and shows the influences of the number of samples on the classification rate. The mean classification rate is plotted for our method using 3 MFS vectors as well as for a the single MFS vector defined on the density of the intensity (26 numbers only). The figure demonstrates that our method has a classification rate which is slightly worse than the LA-method, but generally

much better than the VZ-Joint method. It is worth mentioning that the MFS performs worse than the LA-method only in four of the 25 classes; it is those with significant illumination changes. For all other classes it actually has a very impressive classification rate (see Figure 14a and c).

Finally, we would like to address that a major advantage of our method is its low dimension. Also, our algorithm will behave much better for larger images. The images used in the experiment are not large enough to obtain very good estimates of the fractal dimension. As a result the estimated MFS is not perfectly invariant anymore. The performance of our method increases a lot with an increase image size. In comparison, the feature-based method ([6]) will not benefit much from larger image size, as it strongly depends on clustering.

#### 4. Summary and Conclusions

We showed both in theory and experiment that the MFS offers a new and promising efficient texture description. Not only it is comparable to other top methods in classifying textures, but also it has a mathematically justified global bi-Lipschitz invariance. To our knowledge, no other approach proposed in the past has such global properties. In practice, the MFS is very robust to spatial transformations and somewhat less robust to changes in illumination, but comparable to feature-based methods. Experiments on standard datasets demonstrated its effectiveness and efficiency in texture retrieval and classification tasks, achieving a performance comparable to the top feature based methods but with far lower dimension (seventies vs thousands). Furthermore, the MFS is very efficient and simple to compute without requiring feature detection and clustering. In future work, we will investigate how to better combine MFS vectors from different measurement functions in order to achieve greater robustness to illumination changes.

## References

- [1] R. Azencott, J. Ping, and L. Younes, "Texture Classification Using Windowed Fourier Filters," *PAMI.*, Vol. 19(2), pp. 148-153, 1997.
- [2] F. Espinal, B. D. Jawerth and T. Kubota, "Wavelet-based Fractal Signature Analysis for Automatic Target Recognition," *Opt. Eng.*, Vol. 37(1), pp. 166-174, 1998.
- [3] K. J. Falconer, *Techniques in Fractal Geometry*, John Wiley, 1997.
- [4] C. Kervrann and F. Heitz, "A Markov Random Field Model-based Approach to Unsupervised Texture Segmentation Using Local and Global Spatial Statistics," *IEEE Trans. Image Process*, Vol. 4(6), pp. 856-862, 1995.
- [5] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *PAMI.*, Vol. 15(11), pp. 1186-1191, 1993.
- [6] S. Lazebnik, C. Schmid and J. Ponce, "A Sparse Texture Representation Using Affine-Invariant Regions," *PAMI*, 2004
- [7] B. B. Mandelbrot, *The Fractal Geometry of Nature*, San Francisco, CA: Freeman, 1982.
- [8] L. Kam and J. Blanc-Talon, "Are Multifractal Multipermuted Multinomial Measures Good Enough for Unsupervised Image Segmentation", *CVPR*, pp. 58-63, 2000.
- [9] L. M. Kaplan, "Extended Fractal Analysis for Texture Classification and Segmentation", *PAMI* Vol. 8(11), pp. 1572-1585, 1999.
- [10] A. Conci and L. H. Monteiro, "Multifractal Characterization of Texture-Based Segmentation", *ICIP*, pp. 792-795, 2000.
- [11] T. Mihran, "Moment Based Texture Segmentation," *Pattern Recognition Letters*, Vol. 15, pp. 659-668, 1994.
- [12] F. Mindru, T. Tuytelaars, L. Van Gool and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination", *Computer Vision and Image Understanding*, Vol. 94(1-3), pp. 3-27, 2004.
- [13] J. L. Mundy and A. Zisserman, *Geometric invariance in Computer Vision*, MIT Press, 1992.
- [14] S. Peleg, J. Naor, R. Hartley and D. Avnir, "Multiple resolution texture analysis and classification," *PAMI*, Vol. 6, pp. 518-523, 1984.
- [15] A. Teuner, O. Pichler, and B. J. Hosticka, "Unsupervised Texture Segmentation of Images Using Tuned Matched Gabor Filters," *IEEE Trans. Image Process*, Vol. 4(6), pp. 863-870, 1995.
- [16] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence", *CVPR*, Vol. 3, pp. 255-271, 2003.
- [17] I. Weiss, "Geometric Invariants and Object Recognition," *IJCV*, Vol. 10(3), pp. 207-231, 1993.

## A. Proof of Theorem 1.

Let  $f$  denote the bi-Lipschitz transform, then the new image after applying such a transform on the original image  $I(x)$

could be written as  $I'(x) = I(f(x))$ . In order to prove the invariance of the MFS to bi-Lipschitz transforms, we first prove that for each  $\alpha$ ,

$$E'_\alpha = f(E_\alpha),$$

where  $E'_\alpha$  denotes the corresponding level set of  $I'(x)$ . Following, we prove that  $f(E_\alpha)$  has the same fractal dimension as  $E_\alpha$ .

For a constant  $\alpha$ , let  $x$  be any point in the set of  $E_\alpha$ , then from (3) we have

$$\lim_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r} = \alpha.$$

Thus, for the point  $f(x)$  in the new image  $I'(x) = I(f(x))$ , we have

$$\begin{aligned} \frac{\log \mu(B(f(x), r))}{\log r} &\leq \frac{\log c_2^2 \mu(B(x, r))}{\log r} \\ &= \frac{2 \log c_2 + \log \mu(B(x, r))}{\log r}. \end{aligned}$$

Similarly, we have

$$\frac{\log \mu(B(f(x), r))}{\log r} \geq \frac{2 \log c_1 + \log \mu(B(x, r))}{\log r}.$$

Taking the limit on both sides of the two inequalities above, we obtain

$$\lim_{r \rightarrow 0} \frac{\log \mu(B(f(x), r))}{\log r} = \lim_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r} = \alpha.$$

Thus, we conclude that  $E'_\alpha = f(E_\alpha)$ .

To complete the proof, we need to show that the fractal dimension of  $f(E_\alpha)$  and the fractal dimension of  $E_\alpha$  are the same. Suppose  $E_\alpha$  is covered by  $N(\delta, E_\alpha)$  sets, which is the smallest number of sets with diameter less than  $\delta$ . Then the  $N(\delta, E_\alpha)$  images of these sets under  $f$  form a cover of  $f(E_\alpha)$  by sets of diameter less than  $c_2 \delta$ . By the definition the fractal dimension, we have

$$\begin{aligned} \dim(f(E_\alpha)) &= \lim_{c_2 \delta \rightarrow 0} \frac{\log N(c_2 \delta, f(E_\alpha))}{-\log c_2 \delta} \leq \lim_{\delta \rightarrow 0} \frac{\log N(\delta, E_\alpha)}{-\log c_2 \delta} \\ &= \lim_{\delta \rightarrow 0} \frac{\log N(\delta, E_\alpha)}{-\log c_2 - \log \delta} = \lim_{\delta \rightarrow 0} \frac{\log N(\delta, E_\alpha)}{-\log \delta} \\ &= \dim(E_\alpha). \end{aligned}$$

On the other hand, suppose  $f(E_\alpha)$  is covered by  $N(\delta, f(E_\alpha))$  sets, which is the smallest number of sets of diameter at most  $\delta$ . The same argument yields

$$\dim(E_\alpha) \leq \lim_{\delta \rightarrow 0} \frac{\log N(\delta, f(E_\alpha))}{-\log c_1^{-1} - \log \delta} = \dim(f(E_\alpha)).$$

Thus,  $\dim(E_\alpha) = \dim(f(E_\alpha))$ . The MFS of the new image  $I(f(x))$  equals the MFS of the original image  $I(x)$ .  $\square$