NESTED NORMALIZATIONS FOR DECOUPLING GLOBAL FEATURES

Javier Portilla and Eduardo Martínez-Enríquez

Instituto de Óptica, CSIC, Madrid

ABSTRACT

Global features of signals are usually *coupled*, in the sense that not every combination of features' values is mathematically possible. The existence of *forbidden* combinations complicates the topology of the joint feature vector set, and precludes an interpretation of features as independent descriptors. Another consequence is that there is no easy and direct way of imposing a feature to the signal without perturbing others. Therefore, signal analysis, processing and synthesis applications based on global descriptors often resent from feature coupling. Here we first propose a sequential method for modifying arbitrary features independently of each other. Then we derive a decoupling method based on hierarchical normalization that yields new features with gradients mutually orthogonal everywhere in the signal space. Here we apply this method to the first ten marginal moments, emerging new sample statistics that, as we show, significantly improve texture classification.

Index Terms— global features, non-linear orthogonal features; sample statistics; normalization; ODEs; level sets

1. INTRODUCTION

Scalar functions of a vector are usually *coupled*. For instance, among sampling statistics, second raw order moment's range is conditioned by the mean ($\mu_2 \geq \mu_1^2$), the kurtosis (k) by the skewness (s) ($k \geq 1 + s^2$ [1, 2]), and the second autocorrelation factor by the first ($\rho(2) \geq 2\rho(1)^2 - 1$, see, e.g., [3]). Such bounds between sample statistics are purely algebraic, independent of signal statistics. A direct consequence of coupling is that one can not freely choose arbitrary combinations of values when synthesizing or processing signals by adjusting their features' values. Partly because of this, powerful parametric synthesis and feature-transfer approaches are limited in practice to be example-based: they obtain the features' values in the signal to be imitated, and, by successive approximations, impose them to other (synthetic pseudo-random, or real) signal (e.g., [4, 5, 6, 7]). It is desired to have decoupled features not only because this avoids the need of successive approximations during the adjustment process (as required, e.g., in [4, 6]), but more importantly, because it allows to freely explore and monitor the effect of each individual feature on the signal (and, e.g., its neural response [8, 9, 10, 11]). Furthermore, decoupled features will typically be less redundant, which is a desirable property for pattern analysis and classification, and may help to obtain more parsimonious representations (see, e.g. [12]).

Beyond classical standardization of moments, features' coupling has received little attention so far (two exceptions are [13, 14], which proposed a measure of kurtosis less influenced by skewness). Here we propose a general method for decoupling global features that (i) obtains vector fields that are mutually orthogonal everywhere for

successive features; (ii) from these vector fields, obtains the trajectories to be followed for modifying each feature without affecting the others; and (iii) based on sequentially adjusting each feature to reference values (*normalization*), constructs a set of decoupled global features. In this paper we apply it to texture classification.

2. ORTHOGONALIZING GLOBAL FEATURES

2.1. Problem formulation

Definition: A set of differentiable functions with same input vector domain are *decoupled* (*coupled*, otherwise) i.i.f. their gradients are mutually orthogonal everywhere in the domain. **Goal:** Let $\vec{x} \in R^N$ be a discretized signal, and let $\{f_i(\vec{x}), i = 1 \dots M\}$ be an ordered set of functions. We search for a *similar* set $\{\tilde{f}_i(\vec{x}), i = 1 \dots M\}$ fulfilling $\nabla \tilde{f}_i(\vec{x}) \cdot \nabla \tilde{f}_k(\vec{x}) = 0, \forall i \neq k, \forall \vec{x} \in R^N$.

2.2. Orthogonal gradient directions

Instead of aiming at directly computing $\nabla \tilde{f}_k(\vec{x})$, we search for vector fields $\vec{g}_k(\vec{x})$ along the same directions: $\vec{g}_k(\vec{x}) \propto \nabla f_k(\vec{x})$ everywhere. These vector fields do not need to fulfill the requirements of being actual gradients. The calculation of the new gradient directions are obtained by Gram-Schmidt sequential orthogonalization. We first set $\vec{g}_1(\vec{x}) = \nabla f_1(\vec{x})$. At every step for $k > 1$, we take the difference of $\nabla f_k(\vec{x})$ and its projection onto the subspace $\mathcal{L}_{k-1}$ spanned by previous vectors $\{\vec{g}_j, j = 1, \dots, k-1\}$:

$$\vec{g}_k(\vec{x}) = \nabla f_k(\vec{x}) - P_{\mathcal{L}_{k-1}}(\nabla f_k(\vec{x})). \quad (1)$$

Note that vectors $\{\vec{g}_k, k = 1, \dots, M\}$ span, at every point, the same subspace as the original gradients.

2.3. Adjusting the values of the global features

Now we introduce an arbitrary free scalar parameter t in order to obtain a parametric equation of the *trajectory* to be followed for modifying the k -th feature. Let $\vec{y}_k(t)$ describe a certain evolution in t of the points along that trajectory. By differentiating $\vec{y}_k(t)$ we obtain a vector tangent to the trajectory, which we set:

$$\frac{d\vec{y}_k(t)}{dt} = \vec{g}_k(\vec{y}_k(t)), \quad (2)$$

with the initial condition $\vec{y}_k(0) = \vec{x}$. Integrating this system of ordinary differential equations (ODE) until reaching a desired value v_k of a global feature $f_k(\vec{y})$ results in $T_{\vec{g}_k}(\vec{x}, v_k; f_k)$. Because Eq.(2) is autonomous (the differential only depends on its position), the *trajectory* of $\vec{y}_k(t)$, and, thus, its intersection with the manifold $f_k(\vec{x}) = v_k$, only depends on the *direction* of vectors $\vec{g}_k$, not on their magnitude. Thus, $T_{\vec{g}_k}(\vec{x}, v_k; f_k) = T_{\nabla f_k}(\vec{x}, v_k; f_k)$. By sequentially following these trajectories we can jointly adjust arbitrary

sets of differentiable scalar functions to any (possible) combination of values. This method can be used in image processing or synthesis-by-analysis for transferring features. However, it does not produce, by itself, decoupled features.

2.4. Nested normalizations

Intuitively, by normalizing a signal with respect to a certain feature, we remove the direct source of information about that feature. This is a standard procedure in statistics and signal processing, e.g., when subtracting the mean or dividing by the standard deviation. It is also a means of decoupling the involved features.

Given a vector $\vec{x} \in R^N$, a gradient $\nabla \tilde{f}_k$ (or, equivalently, a same-direction vector $\vec{g}_k$) and a global feature f_k , we set a *reference value* $v_{REF}^{(k)}$, and we call $\hat{x}_k$ the *normalized vector* of $\vec{x}$ w.r.t. f_k :

$$\hat{x}_k = T_{\nabla \tilde{f}_k}(\vec{x}, v_{REF}^{(k)}; f_k). \quad (3)$$

As $v^{(k)}(t) = f_k(\vec{y}_k(t))$, $t_{REF} = \arg_t \{f_k(\vec{y}_k(t)) = v_{REF}^{(k)}\}$. We define our decoupled features as those resulting from evaluating the original functions on the successively normalized vectors:

$$\begin{aligned} \hat{x}_0 &= \vec{x} \\ \tilde{f}_1(\vec{x}) &= f_1(\vec{x}) \end{aligned} \quad (4)$$

for $k = 1$ to $M - 1$

$$\hat{x}_k = T_{\nabla \tilde{f}_k}(\hat{x}_{k-1}, v_{REF}^{(k)}; f_k) \quad (5)$$

$$\tilde{f}_{k+1}(\vec{x}) = f_{k+1}(\hat{x}_k). \quad (6)$$

Previous definition is consistent, i.e., $\{\nabla \tilde{f}_i \cdot \nabla \tilde{f}_j = 0, i \neq j\}$ and $\nabla \tilde{f}_k(\vec{x})/||\nabla \tilde{f}_k(\vec{x})|| = \vec{g}_k(\vec{x})/||\vec{g}_k(\vec{x})||$, with $\{\vec{g}_k\}$ fulfilling Eq. (1). We omit the proof here for brevity.

2.5. Graphical interpretation

Figure 1 illustrates the proposed method with $\vec{x} \in R^3$. The two surfaces are iso-surfaces for f_1 . In gray dotted surface $f_1(\vec{x}) = c$ (arbitrary value), while in the blue solid one $f_1(\vec{x}) = v_{REF}^{(1)}$. Dark curves on the blue surface are the intersection of $f_1(\vec{x}) = v_{REF}^{(1)}$ and successive iso-surfaces of f_2 , $\{f_2(\vec{x}) = c_2^i\}$. The method starts from a given $\vec{x} = \hat{x}_0$. In the first iteration, the method changes $\vec{x}$ until reaching the reference value for the first feature $f_1(\vec{x}) = v_{REF}^{(1)}$, i.e., it normalizes $\vec{x}$ w.r.t. feature $f_1(\vec{x})$ following the curved dashed line with crosses trajectory ($\hat{x}_1 = T_{\vec{g}_1}(\hat{x}_0, v_{REF}^{(1)}; f_1)$). Then a new feature is obtained by evaluating the second function, $\tilde{f}_2(\vec{x}) = f_2(\hat{x}_1)$. Keeping f_1 unchanged, $f_2(\vec{x})$ is adjusted to its reference value $f_2(\vec{x}) = v_{REF}^{(2)}$ (following the dashed gray curve, which is orthogonal to the iso-lines of $f_2(\vec{x})$), obtaining $\hat{x}_2 = T_{\vec{g}_2}(\hat{x}_1, v_{REF}^{(2)}; f_2)$. Finally, the new feature $\tilde{f}_3(\vec{x})$ is obtained as $f_3(\hat{x}_2)$. The black dotted vector is $\nabla f_2(\vec{x})$, the red solid vector is its projection onto the subspace spanned by $\nabla f_1(\vec{x})$, $P_{L_1}(\nabla f_2(\vec{x}))$, and the yellow solid vector is $\vec{g}_2(\vec{x})$ (parallel to $\nabla \tilde{f}_2(\vec{x})$), obtained using (1).¹

3. RAW MOMENTS

We start exploring new features by looking at the raw moments:

$$\mu_j(\vec{x}) = 1/N \sum_{n=1}^N x_n^j,$$

¹Vectors have been magnified for illustration purposes.

Fig. 1. Graphical interpretation of nested normalizations.

which fulfill $\nabla \mu_k(\vec{x}) \propto \vec{x}^{\odot(k-1)}$, where $\vec{x} \odot \vec{y} = (x_1 y_1, \dots, x_N y_N)$ and $\vec{x}^{\odot n} = \vec{x} \odot \vec{x} \odot \dots \odot \vec{x}$ n times.

3.1. Orthogonalization

Using Eq. (1), we get new gradient directions $\vec{g}_k(\vec{x})$. Doing this calculation for higher orders we obtain by induction a recursive expression for any order k (here we omitted constants depending on N , and the dependency on $\vec{x}$):

$$\vec{g}_1 = \vec{1} \quad (7)$$

$$\vec{g}_2 = \vec{x} - \mu_1 \vec{1} \quad (8)$$

$$\vec{g}_3 = \vec{x}^{\odot 2} - \mu_2 \vec{1} - a_{2,3} \vec{g}_2 \quad (9)$$

$\vdots$

$$\vec{g}_k = \vec{x}^{\odot(k-1)} - \sum_{j=1}^{k-1} a_{j,k} \vec{g}_j, \quad (10)$$

with

$$c_{i,j}^{(\ell)} = \begin{cases} \mu_{i+j} - \mu_i \mu_j & \text{if } \ell = 1 \\ c_{i,j}^{(\ell-1)} c_{\ell-1,\ell-1}^{(\ell-1)} - c_{\ell-1,i}^{(\ell-1)} c_{\ell-1,j}^{(\ell-1)} & \text{if } \ell > 1 \end{cases}$$

$$a_{j,k} = \begin{cases} \mu_{k-1} & \text{if } j = 1 \\ c_{j-1,k-1}^{(j-1)} / c_{j-1,j-1}^{(j-1)} & \text{if } j > 1. \end{cases}$$

3.2. Nested normalizations

We have chosen $\{v_{REF}^{(k)} = \mu_k^G\}$, the moments of a zero-mean, unity variance Gaussian: $\mu_k^G = 0$, for k odd and $\mu_k^G = \prod_{j=1}^{k/2} (2j - 1)$, for k even. According to Eq. (4), for the first order $\tilde{\mu}_1(\vec{x}) = \mu_1(\vec{x})$. For the second order, from Eq. (7), Eq. (2) is $y'_{j,1}(t) = 1$, where j indicates the j th element of $\vec{y}$. This yields $y_{j,1}(t) = x_j + t$, and the parameter making $\mu_1 = \mu_1^G = 0$ is $t_{REF} = -\mu_1(\vec{x})$. Then $\hat{x}_1 = T_{\nabla \tilde{f}_1}(\vec{x}, \mu_1^G; \mu_1) = \vec{x} - \mu_1(\vec{x}) \vec{1}$, and, applying Eq. (6) it yields $\tilde{\mu}_2(\vec{x}) = \mu_2(\vec{x} - \mu_1(\vec{x}) \vec{1})$, which is a (biased) estimate of the variance. For the third order, from Eq. (8) and Eq. (2) we have $y'_{j,2}(t) = y_{j,2}(t) - \mu_1(\vec{x})$. Because this gradient is orthogonal to $\nabla \tilde{\mu}_1 = \vec{1}$, μ_1 holds constant, and it is immediate to solve the ODE: $y_{j,2}(t) = (x_j - \mu_1) \exp(t) + \mu_1$, and the parameter making $\mu_2 = \mu_2^G = 1$ is $t_{REF} = -1/2 \log(\mu_2(\vec{x}) - \mu_1^2)$. Then $\hat{x}_2 = T_{\nabla \tilde{f}_2}(\hat{x}_1, \mu_2^G; \mu_2) = (\vec{x} - \mu_1(\vec{x}) \vec{1}) / \sigma(\vec{x})$, being $\sigma^2(\vec{x}) = \mu_2(\vec{x}) - \mu_1(\vec{x})^2$ the sample variance. Finally, by applying Eq. (6) it yields $\tilde{\mu}_3(\vec{x}) = \mu_3((\vec{x} - \mu_1(\vec{x}) \vec{1}) / \sigma(\vec{x}))$, which is the classical skewness. So far, by applying

the proposed method we obtained classical expressions for sample mean, variance and skewness.

For the fourth order feature, because $\vec{g}_3 \perp \nabla \tilde{\mu}_1$ and $\vec{g}_3 \perp \nabla \tilde{\mu}_2$ both $\tilde{\mu}_1 = \mu_1$ and $\tilde{\mu}_2$ (therefore, also μ_2) hold constant in $\vec{y}_3(t)$. However, Eq. (9) also involves $\mu_3(\vec{x})$ (within the $a_{2,3}$ factor), which is *not* constant here. Therefore, we can not solve each marginal ODE independently like in the first three cases, and we face a system of ODEs with no (as far as we know) analytical solution. Something similar happens with higher orders. In these cases, we have performed the normalization by solving Eq. (2) numerically, using the Runge-Kutta integration method. In the fourth order case, the numerical integration is used to obtain a zero-skewness sample, and a new sample statistic $\tilde{\mu}_4(\vec{x})$ emerges, corresponding to measuring the fourth order moment of $\hat{x}_3$, a (somehow) symmetrized signal, normalized in mean, variance and also in skew. Similarly, we obtain the higher order new sample statistics, by measuring the k -order moment on a vector normalized in the previous $k - 1$ sample statistics.

3.3. Visualizing skewness, kurtosis, and orthokurtosis

Figure 2 illustrates how skewness and kurtosis are coupled features, and how to decouple them. In order to visually represent the involved features we choose $N = 4$ (e.g., a 2×2 pixel image). By representing the signal samples' values in a new orthonormal basis including the $[1, 1, 1, 1]/2$ vector along the mean direction we decouple the sample mean, the other three components representing zero-mean vectors. By using a $\sqrt{N} = 2$ radius 3D sphere we show the unit sample variance manifold. The skewness and kurtosis (upper and middle panel of the figure, respectively) are obtained by computing the sample third and fourth order moments on the sphere. We have marked the iso-skewness and iso-kurtosis lines, to show how far they are from being orthogonal: when the skewness is high in absolute value, so it is the kurtosis, and only for small absolute skewness both features are close to orthogonal. In contrast, bottom panel shows the trajectories on the sphere obtained by integrating Eq. (2) with the vector of Eq. (9), which are orthogonal to the iso-skewness lines. Each trajectory also represents a set of vectors having the same *orthokurtosis* $\tilde{\mu}_4$ (μ_4 measured on the normalized zero-skew vector), which is decoupled from mean, variance and skewness. Note how in the new representation skewness and orthokurtosis have decoupled output domains: any skewness value is compatible with any admissible value of orthokurtosis and vice versa.

4. EXPERIMENTS

Here we apply the proposed method for texture classification. We compare the performance of two classifiers using features derived (i) from standardized moments, from 2nd to 10th, (ii) from a generalized form of Ref. [13], and (iii) from the decoupled features. 54 textures of 640×640 pixels were selected from the Brodatz [15] database under the criterion of looking homogeneous in 64×64 pixel patches². Every texture was divided into 10×10 disjoint patches. The problem consists of classifying patches in their corresponding textures. To extract the features we first applied the TILs representation [16], a tight frame acting as bar and edge detector. Discarding the low-pass band, the representation provides 8 subbands. Then, for every patch and every subband, three set of features were obtained: (i) classical features: variance (v), skewness (s), kurtosis (k)

²The list of selected texture and the ®Matlab code of the experiments available upon publication in <https://www.researchgate.net/project/Nested-Normalizations-for-Decoupling-Global-Features>.

Fig. 2. Representing sample skewness (up), sample kurtosis (middle) and new fourth order measurement (bottom, see text for details.)

Table 1. Averaged mutual information

Features	(3,4)	(3,5)	(4,5)	(3,6)	(4,6)	(5,6)
Classical	0.26	0.61	1.39	0.55	3.10	3.11
Mod. [13]	0.23	0.61	1.39	0.55	3.07	3.03
Proposed	0.22	0.04	0.08	0.09	0.54	0.09

and rest of higher order standardized moments; (ii) modified moment set: same as classical, except that now even order moments are shift-minimized ($\tilde{\mu}_n = \min_{\alpha_n} \mathbb{E}\{(\hat{x} - \alpha_n)^n\}$, $\hat{x}$ being standardized observation) [13]; and (iii) proposed features set: v (equal to $\tilde{\mu}_2$); s (equal to $\tilde{\mu}_3$); and the rest of orthogonal higher order features.

4.1. Decoupling effect and mutual information

In order to quantify the redundancy between features, we have estimated the mutual information (MI) between pairs of features [17]. Some mean values across subbands are shown in Table 1. We see that, except for the (3,4)-th case, the proposed features present a drastic reduction of redundancy compared to the classical and modified [13] ones. Figure 3 shows a 3D subset of the classical (left) and proposed (right) features (i.e., before and after our decoupling method), to illustrate the decoupling effect on the data distribution in the feature space. Shown data correspond to a single subband (number 5 in the representation), of textures D103 and D111, selecting just 3 features for the classical (s , k and 5th) and the proposed ($\tilde{\mu}_3$, $\tilde{\mu}_4$ and $\tilde{\mu}_5$) sets. We include the projections onto the three orthogonal planes. The decoupling between every pair of features is apparent, specially for the (3,5)-th and (4-5)-th order cases.

Fig. 3. Illustrating the decoupling effect in the feature space. Left: Using classical features; Right: Using proposed features.

4.2. Classification: Results and discussion

We used two classifiers: a naïve (univariate) Gaussian and a parameter-optimized Support Vector Machine (SVM) using Radial Basis Functions.³ We applied cross validation with 4 folds, and averaged 8 runs for each result. Figure 4 shows the classification results for the three compared sets and the two classifiers (see legend). We observe a totally different behavior between using classical and modified [13] moments, vs. the decoupled ones: whereas the former achieve their optima when using only variance, skewness and kurtosis ($n = 4$), roughly achieving 3% and 2% error ratios for naïve and SVM, respectively, the latter keep on decreasing the error when adding higher order features, reaching 1.34% and 0.86%, respectively, for $n = 10$. We see how [13] produces just a marginal improvement. It is also very significant how, for the classical features, the naïve method behaves very differently from SVM, whereas, for the decoupled features, results of SVM and naïve classifiers run in close parallel. We believe this is due to the strong assumption made by the naïve method (namely, that features are mutually independent), which holds approximately true *after* the feature decoupling, but not before.

5. CONCLUSIONS AND FUTURE WORK

We have proposed a hierarchical gradient orthogonalization method for, given an ordered set of global features, obtaining a certain decoupled set that minimizes a constrained L_2 metric w.r.t. the original gradients. The developed methods, by means of a differential Gram-Schmidt orthogonalization procedure, allow for a sequential individual adjustment of the values of any set of differentiable features. Furthermore, by applying successive (*nested*) normalizations, new decoupled features with orthogonal gradients emerge. These belong to a compact joint feature domain where each feature can take any value in its domain independently of the others. Here we have shown how to apply this local orthogonalization to classical sampling moments of arbitrary order. Interestingly, the three first solutions, corre-

Fig. 4. Classification error, in % (see text for details).

spond to the sample mean, variance and skewness. However, higher order moments produce completely new features. Actually, they have provided a drastic decrease of percentage error in a texture classification experiment with 52 classes, decreasing to less than half the percentage error, and also outperforming previous decoupling recipes [13]. Furthermore, thanks to the reduced redundancy amongst new features, excellent classification results can be obtained using simple methods. The proposed decoupled features may be relatively costly to compute, lack analytical expressions, and may require new, perhaps unfamiliar, interpretations. Nevertheless, based on what we have shown here, the method seems highly promising for pattern analysis. And, because it is based on non-linearly transforming the signal, it also provides a new general-purpose frame for both signal processing (e.g., features transfer) and sampling/simulation (e.g., texture and other signal synthesis, for neural stimuli, etc.). We plan to explore and characterize new decoupled feature sets, like spatial functions, and other pdf's descriptors.

³We thank the authors of the Pattern Recognition Toolbox (<http://covartech.github.io>), which we used in preliminary experiments.

6. REFERENCES

- [1] K. Pearson, "IX. Mathematical contributions to the theory of evolution.—XIX. Second supplement to a memoir on skew variation," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 216, no. 538-548, pp. 429–457, 1916.
- [2] R. Sharma and R. Bhandari, "Skewness, kurtosis and Newton's inequality," *ArXiv e-prints*, Sept. 2013.
- [3] K. Swamy, "On Sylvester's criterion for positive-semidefinite matrices," *IEEE Transactions on Automatic Control*, vol. 18, no. 3, pp. 306–306, Jun 1973.
- [4] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings., International Conference on Image Processing*, Oct 1995, vol. 3, pp. 648–651 vol.3.
- [5] J. Portilla, R. Navarro, O. Nestares, and A. Taberero, "Texture synthesis-by-analysis based on a multiscale early-vision model," *Optical Engineering*, vol. 35, no. 8, pp. 2403–2417, 1996.
- [6] J. Portilla and E.P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40(1), pp. 49–71, 2000, $\text{\textcircled{R}}$ Matlab code publicly available at <http://www.cns.nyu.edu/~lcv/texture/>.
- [7] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 2414–2423.
- [8] Jeremy Freeman, Corey M Ziemba, David J Heeger, Eero P Simoncelli, and J Anthony Movshon, "A functional and perceptual signature of the second visual area in primates," *Nature Neuroscience*, vol. 16, 05 2013.
- [9] Gouki Okazawa, Satoshi Tajima, and Hidehiko Komatsu, "Image statistics underlying natural texture selectivity of neurons in macaque v4," *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. E351–E360, 2015.
- [10] Margaret S. Livingstone, Justin L. Vincent, Michael J. Arcaro, Krishna Srihasam, Peter F. Schade, and Tristram Savage, "Development of the macaque face-patch system," *Nature Communications*, vol. 8, 03 2017.
- [11] Srivatsun Sadagopan, Wilbert Zarco, and Winrich A Freiwald, "A causal relationship between face-patch activity and face-detection behavior," *eLife*, vol. 6, pp. e18558, apr 2017.
- [12] J. Balas, "Texture synthesis and perception: Using computational models to study texture representations in the human visual system," *Vision Research*, vol. 46, no. 3, pp. 299 – 309, 2006.
- [13] D. C. Blest, "A new measure of kurtosis adjusted for skewness," *Australian and New Zealand Journal of Statistics*, vol. 45, no. 2, pp. 175–179, 2003.
- [14] J.F. Rosco, A.Pewsey, and M. C. Jones, "On Blest's measure of kurtosis adjusted for skewness," *Communications in Statistics: Theory and Methods*, vol. 44, no. 17, pp. 3628–3638, 2015.
- [15] P. Brodatz, *Textures: a photographic album for artists and designers*, Dover pictorial archives. Dover Publications, 1966, Available at https://multibandtexture.recherche.usherbrooke.ca/original_brodatz.html.
- [16] J. Portilla, A. Tristan-Vega, and I.W. Selesnick, "Efficient and robust image restoration using multiple-feature L2-relaxed sparse analysis priors," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5046–5059, Dec 2015, $\text{\textcircled{R}}$ Matlab code publicly available at DOI: 10.13140/RG.2.1.3531.5280.
- [17] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.