Directional Image Decomposition Using Retargeting Pyramid

Yuichi Tanaka* and Keiichiro Shirai†
* Tokyo University of Agriculture and Technology, Tokyo, Japan
Email: ytnk@cc.tuat.ac.jp Tel: +81-42-388-7150
† Shinshu University, Nagano, Japan
Email: shirai@cs.shinshu-u.ac.jp Tel: +81-26-269-5486

***Abstract*—Retargeting pyramid (RP) is an alternative method for multiscale image decomposition to the well-known Laplacian pyramid (LP). RP can be obtained by replacing the low-pass filtering process in LP with content-aware image resizing (a.k.a. retargeting), which is a developing technique for computer vision researches. Furthermore, we use RP for contourlet-based directional image decomposition. In experimental results, the proposed decomposition outperforms the LP-based contourlet transform for image denoising.**

I. Introduction

Images often have textures and/or edges laying toward diagonal orientations as well as horizontal/vertical directions. Since it is also well-known that multiscale decomposition, e.g., by wavelet transform [1], is a strong tool for image signal analysis, multiscale-multidirection (MSMD) image decomposition is highly desired and a number of these transforms have been proposed. Without being exhaustive, contourlet [2]–[5], directionlet [6], [7], and curvelet [8], [9] are key methods for MSMD decomposition. As applications, they are used in a wide area of image processing; denoising, enhancement, interpolation, and image coding. They often outperform the conventional separable wavelet-based approach.

In this paper, we propose a new MS image decomposition technique named retargeting pyramid (RP). Its downsampling process is quite different from Laplacian pyramid (LP) [10], a widely-used MS decomposition method. Generally LP is obtained by the separable low-pass filtering along both horizontal and vertical directions followed by the *explicit* downsampling by M. Strictly speaking, LP's downsampling matrix $\boldsymbol{Q} = \mathrm{diag}(M, M)$. Instead of the explicit approach, RP is constructed by the *implicit* donwsampling and filtering by utilizing a developing computer vision technique, content-aware image resizing, a.k.a. *image retargeting* [11]–[21].

The technique whose concept is the same as RP was proposed in [22]. It is based on quad mesh-based retargeting and yields a MS pyramid. However, its deformation is separable, i.e., all deformed meshes still keep rectangular shapes. Furthermore, effective applications have not been shown. In contrast, RP is based on nonseparable mesh deformation: a deformed mesh is allowed to have a (convex or concave) quadrilateral shape. Additionally, to obtain MD decomposition as well as MS one, we combine RP with directional filter bank (DFB) [23], [24] to construct a MSMD decomposition similar to that of the contourlet transform. The proposed MSMD decomposition is applied to image denoising and it outperforms the conventional contourlet transforms.

II. Contourlet Transform

The contourlet transform [2] can be divided into two phases; MS and MD decomposition phases. First, an input image is decomposed by LP which produces a (redundant) MS pyramid of the image. Then, directional filter bank (DFB) [23], [24] is applied to each level of the pyramid to obtain the full MSMD decomposed image.

Let $\boldsymbol{x}_0$ be the input image signal. The i-th level (i is a nonnegative integer) output $\boldsymbol{x}_i$ and $\hat{\boldsymbol{x}}_i$ by LP is formalized as follows:

$$\boldsymbol{x}_{i+1} = \mathbf{L}\boldsymbol{x}_i \tag{1}$$

$$\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i - \mathbf{L}^{-1}\boldsymbol{x}_{i+1} \tag{2}$$

where $\mathbf{L}$ is the LP operator. The LP operator is further divided into

$$\mathbf{L} = (\downarrow 2)\mathbf{F} \tag{3}$$

where $\mathbf{F}$ is 2-D low-pass filtering and $(\downarrow 2)$ is a downsampling by 2 for both horizontal and vertical directions. In the contourlet transform, the residual signal $\hat{\boldsymbol{x}}_i$ is further transformed by DFB:

$$\hat{\boldsymbol{y}}_i = \mathbf{D}_{i,n}\hat{\boldsymbol{x}}_i \tag{4}$$

where $\mathbf{D}_{i,n}$ is i-th level DFB decomposition with 2^n directional subbands. Finally, the J-level full MSMD image can be obtained by $\boldsymbol{x}_J$ and a set of $\hat{\boldsymbol{y}}_i$ $(0 \le i \le J-1)$. The contourlet decomposition is illustrated in Fig. 1.

In the original contourlet transform, the separable wavelet transform is used for $\mathbf{F}$. To improve MS decomposition performance, a new contourlet transform with sharp frequency localization was proposed [3]. It changes $\mathbf{L}$ in the original contourlet transform as follows:

$$\boldsymbol{x}_{i+1} = \tilde{\mathbf{L}}\boldsymbol{x}_i \tag{5}$$

$$\hat{\boldsymbol{x}}_i = \tilde{\mathbf{G}}\boldsymbol{x}_i \tag{6}$$

where

$$\tilde{\mathbf{L}} = (\downarrow d)\tilde{\mathbf{G}} \tag{7}$$

and $\tilde{\mathbf{G}}$ is a high-pass filter making the perfect reconstruction pair with $\tilde{\mathbf{L}}$. Furthermore, the downsampling factor d may

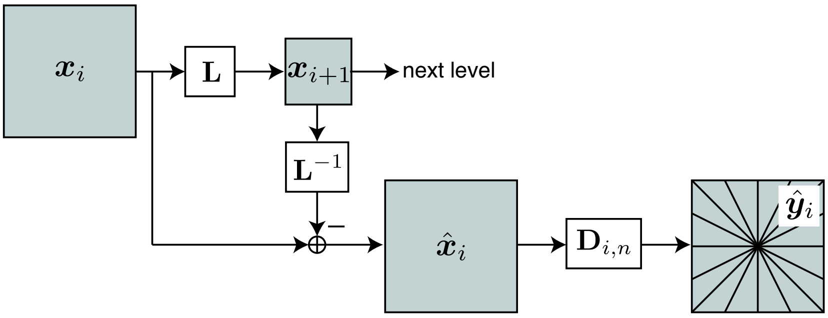

Fig. 1. The contourlet transform with Laplacian pyramid.

be different from 2. The modified contourlet transform shows better performance than the original one for denoising.

III. MSMD Image Decomposition with RP

In this section, we present RP as an alternative to LP. As mentioned in Section II, the donwsampling factors of the contourlet transforms are fixed to 2 or d. In other words, every part of the low-pass filtered signal is uniformly downsampled by 2 or d at the downsampling process. It leads to that every portion of the image is considered to have the same *significance*. Generally speaking, some regions in an image are more prominent than others, since we usually focus on the foreground quality rather than the background clarity. To reflect this assumption correctly, we replace the explicit low-pass filter and downsampling in the contourlet transform with a more sophisticated signal-oriented downsizing, so-called retargeting.

A. Retargeting Pyramid

RP is represented as follows:

$$\boldsymbol{x}_{i+1} = \mathbf{R}\boldsymbol{x}_i \tag{8}$$

$$\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i - \mathbf{R}^{-1}\mathbf{R}\boldsymbol{x}_i. \tag{9}$$

The operation $\mathbf{L}$ in LP is replaced by a retargeting $\mathbf{R}$. In (9), $\mathbf{R}^{-1}$ referred to as the inverse signal mapping corresponding to $\mathbf{R}$. Any size of the retargeted image can be permitted unless the width/height does not exceed the original size. As a result, RP gives more flexibility about the redundancy ratio of the pyramid-based MS image decomposition.

The operation $\mathbf{R}$ can be an arbitrary retargeting method, such as seam carving [11], [12]. In this paper, we customize one of retargeting methods based on mesh deformation, scale-and-stretch [14].

Note that $\mathbf{R}$ cannot be decomposed into the low-pass filtering and downsampling operation, since they are integrated with retargeting. Instead, $\mathbf{R}$ can be divided into the following two phases:

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Phi} \tag{10}$$

where $\mathbf{\Phi}$ is a mapping of the original pixel position $\boldsymbol{p}$ represented as

$$\mathbf{\Phi} : \boldsymbol{p} \to \boldsymbol{p}' \tag{11}$$

where $\boldsymbol{p}'$ is the deformed pixel position. Then $\mathbf{\Lambda}$ picks up the (interpolated) data points at the uniform grid positions $\boldsymbol{\eta} = [\eta_0, \eta_1]^T$ $(0 \le \eta_0 \le W_t - 1,\ 0 \le \eta_1 \le H_t - 1$, in which W_t and H_t are the target width and height, respectively) of the deformed image.

B. Mesh-based Retargeting

Our implementation of image retargeting is summarized below.

1) Calculate a significance map.
2) Construct a weighted Laplacian matrix using the significance map.
3) Optimize the coordinates of the mesh by solving a sparse linear system.
4) Deform the image by using the optimized mesh.
5) Uniformly resize the deformed image into the target size.

In the rest of this subsection, we describe a few key techniques for the image retargeting.

1) Significance Map: The significance map $\boldsymbol{S}$ is defined as follows:

$$\boldsymbol{S} = \hat{\boldsymbol{S}} + \frac{1}{\max(\Delta_{\boldsymbol{x}})}\Delta_{\boldsymbol{x}}, \tag{12}$$

where $\hat{\boldsymbol{S}}$ is a saliency map proposed by Itti et al. [25], and $\Delta_x = ((\frac{\partial}{\partial x}\boldsymbol{x})^2 + (\frac{\partial}{\partial y}\boldsymbol{x})^2)^{1/2}$ is the $L2$-norm of the gradient.

2) Weighted Laplacian Matrix: The energy among edges represented as a pair of vertices $\{k, l\}$ is defined as follows:

$$U = \sum_{\{k,l\}} w_{kl} \|\boldsymbol{u}_k - \boldsymbol{u}_l\|^2, \tag{13}$$

where $\boldsymbol{u} = (u, v)$ is a coordinate of a vertex, and w_{kl} is the weight to control the mesh stretch. We set the weight as:

$$w_{kl} = \exp\left(-\frac{|\nu_{kl}|^{0.5}}{\sigma_w}\right), \quad \nu_{kl} = \frac{S(\boldsymbol{u}_k) + S(\boldsymbol{u}_l)}{2}, \tag{14}$$

where ν_{kl} is the essential weight given as the mean of significance which is then adjusted in w_{kl} by taking a sparse prior considering the sparseness of the significance map. σ_w is an arbitrary standard deviation.

The optimal vertex positions are obtained by differentiating U:

$$\frac{\partial U}{\partial \boldsymbol{u}} = \sum_{l \in \mathcal{N}(k)} w_{lk}(\boldsymbol{u}_k - \boldsymbol{u}_l) = \mathbf{0} \tag{15}$$

TABLE I
MEAN ENERGY OF $\hat{\boldsymbol{x}}_0$

Image	CT-LP	CT-MD	CT-RP
Lena	22.13	49.21	17.97
Pepper	36.49	68.84	25.96

where $\mathcal{N}(k)$ is the set of k's adjacent pixels. The optimal vertex positions can be obtained by solving (15), subject to the constraint of the positions of boundary vertices. Finally, the optimal vertex positions $\boldsymbol{u}_{opt} = \boldsymbol{p}'$ are used to deform the image pixels.

C. RP-based Contourlet Transform

Finally, our proposed RP is applied to replace LP. In our preliminary experiments, RP's retargeting ability is fully utilized when RP is performed at the first level, i.e., retargeting for the original image $\boldsymbol{x}_0$. Consequently, the proposed MSMD decomposition is represented as follows:

$$\boldsymbol{x}_{i+1} = \begin{cases} \mathbf{R}\boldsymbol{x}_i & i = 0 \\ \tilde{\mathbf{L}}\boldsymbol{x}_i & i > 0 \end{cases} \tag{16}$$

$$\hat{\boldsymbol{x}}_i = \begin{cases} \boldsymbol{x}_i - \mathbf{R}^{-1}\mathbf{R}\boldsymbol{x}_i & i = 0 \\ \tilde{\mathbf{G}}\boldsymbol{x}_i & i > 0 \end{cases} \tag{17}$$

$$\hat{\boldsymbol{y}}_i = \mathbf{D}_{i,n}\hat{\boldsymbol{x}}_i \quad \forall i. \tag{18}$$

Note that for $i > 0$, the input signal is a retargeted (deformed) signal whereas $\hat{\boldsymbol{x}}_0$ has an original structure. Therefore, it can be considered as an effective MSMD decomposition which focuses on prominent regions in the image. Note that the average downsampling factor can be determined by a user-defined size of the retargeted image. The proposed structure of the MSMD decomposition is illustrated in Fig. 2.

IV. EXPERIMENTAL RESULTS

In this section, performances of the proposed method are compared with the conventional contourlet transforms. Three MSMD transforms, the original contourlet transform [2] (denoted as CT-LP), the modified contourlet transform [3] (denoted as CT-MD), and our proposed transform with RP (denoted as CT-RP), are used for the experiment. For CT-MD, the downsampling factor d is fixed to 2. For CT-RP, the full-size image is resized (retargeted) to half width/height and σ_w in (14) is set to 0.25. Consequently, all the transforms have the same redundancy ratio 1.33.

The decomposition level J is set to 5. For the DFB decomposition level n, different settings have been used: CT-LP is $[0, 0, 4, 4, 5]$, CT-MD is $[2, 2, 3, 4, 5]$, and CT-RP is $[2, 2, 3, 4, \underline{3}]$[1], where the leftmost number is the coarsest (low-frequency) scale and the rightmost one is the finest (high-frequency) scale. Note that the DFB decomposition levels for the conventional CTs are the same as the original papers [2], [3].

[1]The underlined number represents the DFB decomposition level for $i = 0$.

TABLE II
PSNR OF DENOISED IMAGES (dB)

Lena					
σ	5	10	15	20	25
CT-LP	34.61	31.68	29.90	28.65	27.61
CT-MD	34.70	32.35	31.14	30.24	29.44
CT-RP	**35.61**	**33.07**	**31.62**	**30.48**	**29.53**
Pepper					
σ	5	10	15	20	25
CT-LP	33.53	31.03	29.45	28.25	27.39
CT-MD	33.81	31.93	30.82	29.97	**29.30**
CT-RP	**34.70**	**32.63**	**31.23**	**30.05**	29.09

A. Retargeting Performance

First, we examine the essential performance of RP. Fig. 3 presents retargeting results of our mesh-based retargeting. Clearly edge regions are stretched and smooth regions are shrank by our retargeting method. Furthermore, Fig. 4 shows $\hat{\boldsymbol{x}}_0$ of *Lena* produced by three transforms. Additionally, the mean energies of $\hat{\boldsymbol{x}}_0$ are summarized in Table I. Clearly the CT-RP yields smaller residual values than those of the other contourlet transforms. Since CT-MD has a narrower passband shape than that of the CT-LP [3], it results larger energy in $\hat{\boldsymbol{x}}_0$. In comparison with the CT-LP, the CT-RP gives significantly lower energies of the residual signal.

B. Denoising Performance

We applied the simple hard-thresholding technique for denoising with the threshold 3σ where σ is the standard deviation of noise and it can be estimated by the robust median estimator [26]. PSNRs of the denoised images are summarized in Table II. It is clear that CT-RP presents uniformly better performance than the other transforms except $\sigma = 25$ of *Pepper*. Especially, its performance gain is significant when the noise level is low since noise-free images are assumed for the retargeting process. The denoised *Pepper* images and their enlarged portions are shown in Figs. 5 and 6, respectively. Similar to the objective quality comparison, the denoised image by CT-RP is much better than those of the conventional contourlet transforms.

V. CONCLUSIONS

In this paper, we presented a new MS image decomposition with the mesh-based retargeting. It realizes an efficient MSMD representation of images via the contourlet transform framework. Its denoising performance is superior to the conventional contourlet transforms. Our future work includes to investigate more efficient structures for other image processing applications.

ACKNOWLEDGMENTS

This work was supported in part by Grant-in-Aid for Young Scientists (B) 24760288 and SCAT Research Grant.

REFERENCES

[1] G. Strang and T. Q. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge, MA, 1996.

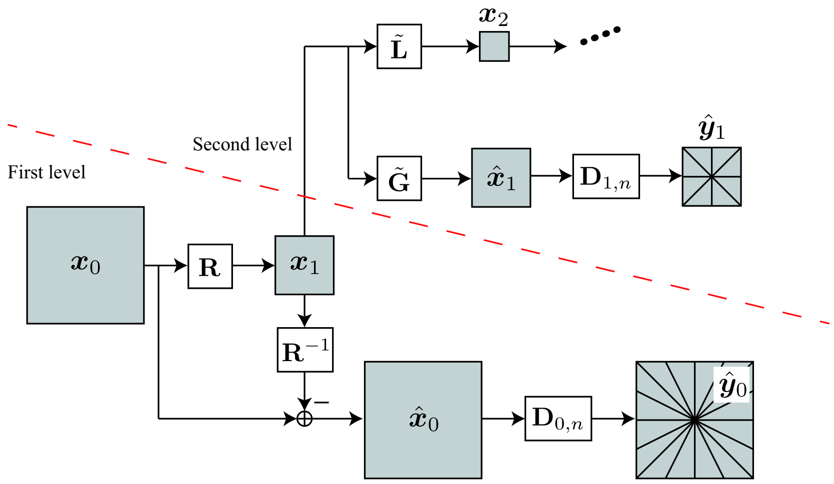

Fig. 2. The contourlet transform with retargeting pyramid.

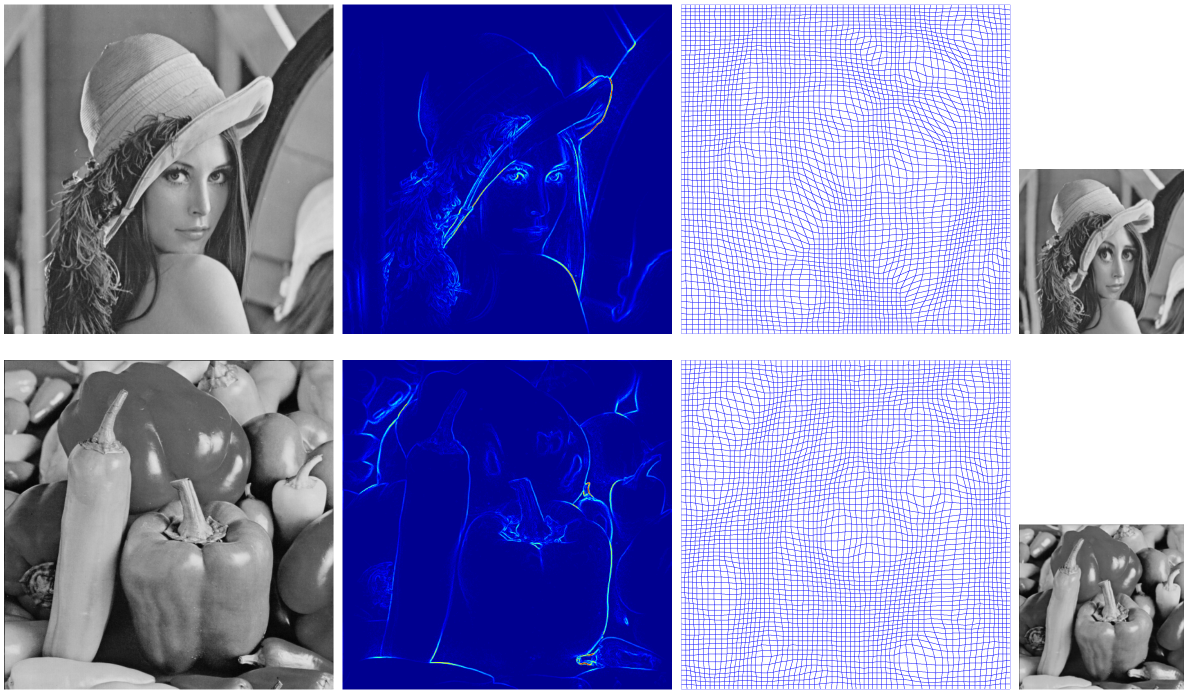

Fig. 3. Retargeting results. From left to right: original image, significance map $\boldsymbol{S}$, deformed mesh, and retargeted image. Top row: *Lena*. Bottom row: *Pepper*.

[2] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, 2005.

[3] Y. Lu and M. N. Do, "A new contourlet transform with sharp frequency localization," in *Proc. ICIP'06*, 2006, pp. 1629–1632.

[4] Y. Lu and M. N. Do, "CRISP-contourlet: a critically sampled directional multiresolution image representation," in *Proc. SPIE Conf. on Wavelet Applications in Signal and Image Processing*, 2003, pp. 655–665.

[5] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, 2006.

[6] V. Velisavljević, B. Beferull-Lozano, M. Vetterli, and P. L. Dragotti, "Directionlets: Anisotropic multi-directional representation with separable filtering," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1916–1933, 2006.

[7] V. Velisavljević, B. Beferull-Lozano, and M. Vetterli, "Space-frequency quantization for image compression with directionlets," *IEEE Trans. Signal Process.*, vol. 16, no. 7, pp. 1761–1773, 2007.

[8] E. Candès and D. L. Donoho, "Curvelets — a surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surfaces Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Vanderbilt University Press, Saint-Malo, 1999.

[9] J. L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, 2002.

[10] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, 1983.

[11] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.

[12] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.

[13] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. ICCV'07*, 2007.

[14] Y. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-

Fig. 4. The finest residual signals $\hat{\boldsymbol{x}}_0$ of the noise-free *Lena* image. From left to right: CT-LP, CT-MD, and CT-RP. The values are emphasized for clearer visualization.

Fig. 5. Denoised *Pepper* image with $\sigma = 10$. From left to right: CT-LP, CT-MD, and CT-RP.

Fig. 6. Enlarged portions of Fig. 5. From left to right: CT-LP, CT-MD, and CT-RP.

stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, 2008.

[15] B. Chen and P. Sen, "Video carving," in *Proc. Eurographics*, 2008.

[16] Z. Li, P. Ishwar, and J. Konrad, "Video condensation by ribbon carving," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2572–2583, 2009.

[17] Y. Guo, F. Liu, J. Shi, Z. H. Zhou, and M. Gleicher, "Image retargeting using mesh parametrization," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 856–867, 2009.

[18] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. ICCV'09*. IEEE, 2009, pp. 151–158.

[19] D. Domingues, A. Alahi, and P. Vandergheynst, "Stream carving: An adaptive seam carving algorithm," in *Proc. ICIP'10*, 2010.

[20] X. Fan, X. Xie, H. Q. Zhou, and W. Y. Ma, "Looking into video frames on small displays," in *Proc. 11th ACM International Conference on Multimedia*. ACM, 2003, pp. 247–250.

[21] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Trans. Graph.*, vol. 28, no. 5, 2009.

[22] C. S. Won and S. Shirani, "Size-controllable region-of-interest in scalable image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1273–1280, 2011.

[23] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, 1992.

[24] S.-I. Park, M. J. T. Smith, and R. M. Mersereau, "Improved structures of maximally decimated directional filter Banks for spatial image analysis," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1424–1431, 2004.

[25] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[26] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, 2000.