RANDOMIZED TEXTURE FLOW ESTIMATION USING VISUAL SIMILARITY

Sunghwan Choi[⋆] *Dongbo Min*[†] *Kwanghoon Sohn*[⋆]

[⋆]School of Electrical and Electronic Engineering, Yonsei University, Korea
[†]Advanced Digital Sciences Center (ADSC), Singapore

ABSTRACT

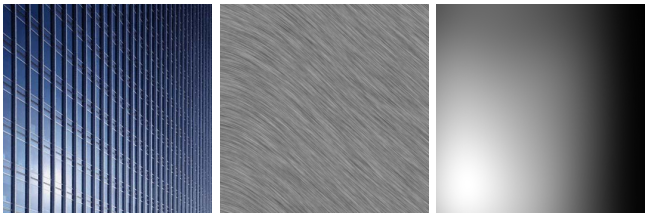
Exploring underlying texture flows defined with orientation and scale is of a great interest on a variety of vision-related tasks. However, existing methods often fail to capture accurate flows due to over-parameterization of texture deformation or employ a costly global optimization which makes the algorithm computationally demanding. In this paper, we address this inverse problem by casting it as a randomized correspondence search along with a locally-adaptive vector field smoothing. When a small example patch is given as a reference, a randomized deformable matching is performed on the very densely quantized label space, enabling an efficient estimation of texture deformation without quality degeneration, *e.g.*, due to quantization artifacts which often appear in the optimization-driven discrete approaches. The visual similarity with respect to the deformation parameters is directly measured with an input texture image on an appearance space. The locally-adaptive smoothing is then applied to the intermediate flow field, resulting in a good continuation of the resultant texture flow. Experimental results on both synthetic and natural images show that the proposed method improves the performance in terms of both runtime efficiency and/or visual quality, compared to the existing methods.

Index Terms— Texture analysis, flow estimation, correspondence search, and joint filtering.

1. INTRODUCTION

Texture flow, typically defined as a locally-varying deformation field in terms of orientation and scale, on a natural photograph is essential to numerous computer vision applications such as scene analysis, segmentation, shape from shading, texture synthesis and editing. It offers a fundamental clue to grasp inherent visual structure embedded in the spatially (and smoothly) varying surface on textures. Texture flow estimation serves as a starting point to interpret the perceptual organization of real world images. Fig. 1 shows the example of texture flow estimated from a natural image by using our method.

Discovering the underlying texture flows through a computational approach, however, is non-trivial, especially due to the deviation in texture appearance. To better model local attributes of the texture, traditional methods usually leverage a parametric model for texture representation. For example, oriented filters were deployed to estimate dominant orientation fields for gradient-like patterns [1]. In [2], a texture feature was represented as a linear array of thresholded pixels (*i.e.*, local binary patterns), followed by the dimension reduction strategy through the principal component analysis (PCA). In [3], Chang and Fisher decomposed a deformed texture into explicit local attributes such as orientation and scale by utilizing a steerable pyramid. In [4], the structure tensor response computed at each pixel is compared to that computed from an example patch in order to discriminate the dominant orientation inherent in the texture.



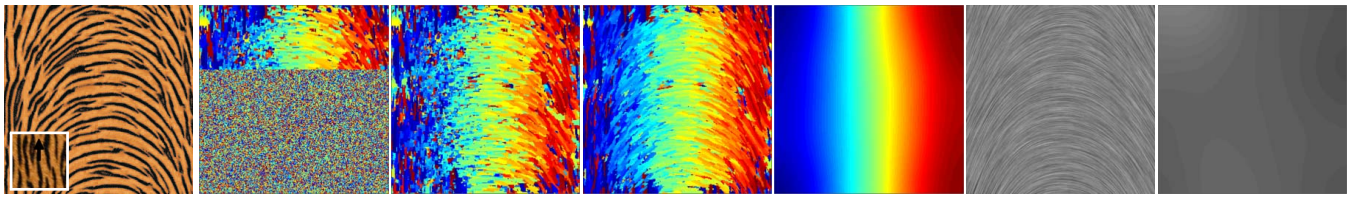
(a) Texture image (b) Orientation (c) Scale

Fig. 1. Examples of texture flow estimation from a natural photograph: (a) the texture image, (b) the estimated orientation field, and (c) the estimated scale field.

To further enforce a global consistency, a costly global optimization is often taken into account [2, 3, 5], by minimizing an objective which combines the data constraint with a regularization term enforcing smoothness on the resultant flow fields. Namely, the estimation process is formulated as the MAP solution of a discrete labeling Markov network [2]. However, such optimization-driven methods suffer from the computational burden caused by a high-dimensional label space and/or quantization artifacts inherent in discrete labeling tasks.

What makes our approach novel for texture flow estimation is that we directly utilize intensity values of two patches for computing the correlation metric, unlike existing texture flow estimation approaches [1, 2, 3, 4] that typically rely on parametric models based on texture descriptors. Since the texture has a stochastic property, this kind of intensity-based correlation measure, *i.e.*, a non-parametric model, is traditionally considered unfit in the existing methods. However, it was shown in the texture synthesis literature [6] that the appearance based synthesis approach produces spatially coherent texture images very well by evolving an input example patch with respect to deformation fields consisting of orientation and scale. Following this observation, we demonstrate that the local visual similarity directly measured using a simple intensity correlation metric captures local behaviors of the inherent flow fields very well. Then, such an exhaustive search with respect to orientation and scale becomes computationally feasible by leveraging the randomized deformable matching that can simulate a lot of possible texture deformation fields very efficiently.

In this paper, we mainly focus on the estimation of deformation fields (consisting of orientation and scale) in the per-pixel labeling framework. Different from existing works [2, 3, 5], we introduce a simple yet effective alternative of a global optimization by leveraging both the randomized search concept [7] and the joint filtering technique [8]. It enables achieving substantial gains in terms of both memory requirement and runtime, which is difficult to achieve with existing optimization-driven approaches due to the computational burden of the high-dimensional variable space.



(a) Input texture (b) $\frac{1}{3}$ iteration (c) 1 iteration (d) 3 iterations (e) Field smoothing (f) Orientation field (g) Scale field

Fig. 2. Algorithm overview. Given an input texture with a small example patch (a), our algorithm iteratively updates a deformation field (b)-(d) from random initialization, followed by local field smoothing (e). It produces globally consistent orientation (f) and scale (g) fields.

2. PROPOSED METHOD

Our texture flow estimation begins with the inference of deformation parameters defined in the high-dimensional search space by virtue of the randomized search framework. Once the randomized search is done, the flow field is then adaptively filtered out with the guidance of matching confidence in order to ensure a global coherency. The proposed framework is illustrated in Fig. 2.

2.1. Deformation Model and Overview

Given a texture image I that undergoes smoothly varying deformation in surfaces, our objective is to discover a dense deformation field $f : I \mapsto \mathbb{R}^2$ defined over all pixel coordinates $\mathbf{p} \in I$ through a deformable correspondence search on the appearance space. We assume that a small example patch T is given by a user and represent the input texture image I well. For modeling the underlying visual structure, we define a texture deformation model as a transformation in terms of orientation $\theta_{\mathbf{p}}$ and scale $s_{\mathbf{p}}$:

$$f(\mathbf{p}) = (\theta_{\mathbf{p}}, s_{\mathbf{p}})^T, \quad \forall \mathbf{p} \in I, \tag{1}$$

where $\theta_{\mathbf{p}}$ and $s_{\mathbf{p}}$ represent orientation label and scale label at a pixel $\mathbf{p}$, respectively. Since orientation is periodic, the search range of orientation is constrained to be $0 \leq \theta_{\mathbf{p}} < 2\pi$. To handle scale, the search range of scale is preset to be in the range of $0.25 \leq s_{\mathbf{p}} \leq 2$.

The deformation field $f(\mathbf{p})$ is inferred by matching the example patch T and a deformed patch centered at a pixel $\mathbf{p}$ from I. Let E_I and E_G denote respectively a distance function for measuring intensity and gradient similarity of two patches as:

$$E_I(\mathbf{p}, \theta, s) = \sum_{\mathbf{q} \in \mathcal{N}(T)} \left\| T(\mathbf{q}) - I(\phi_{\mathbf{p}}^{(\theta,s)}(\mathbf{q})) \right\|^2, \tag{2}$$

$$E_G(\mathbf{p}, \theta, s) = \sum_{\mathbf{q} \in \mathcal{N}(T)} \left\| \nabla T(\mathbf{q}) - \nabla I(\phi_{\mathbf{p}}^{(\theta,s)}(\mathbf{q})) \right\|^2, \tag{3}$$

where ∇ is a gradient operator, and $\mathcal{N}(T)$ is a set of relative pixel locations, with setting the center of the example patch T to an origin. $\phi_{\mathbf{p}}^{(\theta,s)}(\cdot)$ denotes a warping operator with respect to rotation θ and scale s, which yields

$$\phi_{\mathbf{p}}^{(\theta,s)}(\mathbf{q}) = \mathbf{p} + s \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{q}. \tag{4}$$

Based on these intensity-based distance measures, the visual similarity of two patches is defined as follows:

$$\mathcal{V}(\mathbf{p}, \theta, s) = E_I(\mathbf{p}, \theta, s) + E_G(\mathbf{p}, \theta, s). \tag{5}$$

The deformation field $f(\mathbf{p})$ is then estimated by minimizing the following objective function:

$$f(\mathbf{p}) = \underset{(\theta,s) \in \mathbb{F}}{\arg\min}\, \mathcal{V}(\mathbf{p}, \theta, s), \tag{6}$$

where $\mathbb{F}$ is the set of all possible labels with respect to rotation and scale. Indeed, this minimization problem can be simply solved by exhaustively searching over the discretized label space of $\mathbb{F}$. However, such an exhaustive search is computationally expensive, since the number of possible deformation labels is typically very huge.

To address this problems, we introduce a simple yet effective deformable correspondence search by deploying the randomized search concept recently proposed in [7]. In contrast to search over all possible labels, we smartly traverse *parts* of it using a randomized cooperative hill climbing strategy: propagation and random search. It thus achieves a substantial improvement in runtime efficiency while maintaining its original matching quality.

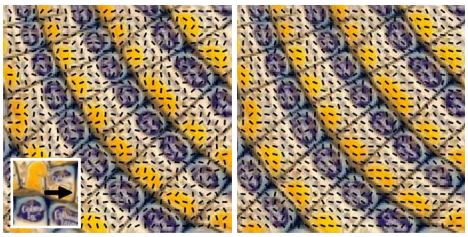
2.2. Randomized Inference

Let us now focus on the problem of finding an optimal deformation field that minimizes the objective function in (6). To handle the computational complexity introduced by a large number of candidate labels, we deploy the randomized search algorithm [7] for inference, which is proven to be highly efficient in the high-dimensional discrete label search. A basic motivation is simple: if deformation fields are initialized by random labels, then correct labels are likely to exist among the set of these random labels. A good guess also guides the rest of pixels to have a good guess by propagating its current labels to the vicinity. This randomized inference process iteratively updates the deformation field f until convergence. For each iteration, good guesses are examined in scan order by alternating between *propagation* and *random search*. Fig. 3(a) shows an example of an initial deformation field f consisting of random labels that are uniformly sampled from $\mathbb{F}$. It should be noted that our algorithm randomly selects arbitrary (floating) values for the orientation and scale within the given search range, and thus it does not suffer from severe quantization artifacts while maintaining its runtime efficiency.

Propagation. In the first step, a propagation proceeds in order to improve an intermediate deformation label $f(\mathbf{p})$ by considering current best label pairs $\Omega_{\mathbf{p}}$ of its neighboring pixels including itself, for instance, $\Omega_{\mathbf{p}} = \{f(\mathbf{p}), f(\mathbf{p}-(1,0)), f(\mathbf{p}-(0,1))\}$ on odd-numbered iteration. The hypothesis test is then performed as follows:

$$\begin{aligned} f(\mathbf{p}) &\leftarrow \underset{(\theta,s) \in \Omega_{\mathbf{p}}}{\arg\min}\, \mathcal{V}(\mathbf{p}, \theta, s), \\ c(\mathbf{p}) &\leftarrow \mathcal{V}(\mathbf{p}, f(\mathbf{p})) \end{aligned} \tag{7}$$

where $\mathcal{V}$ is the visual similarity distance defined in (5) and $\leftarrow$ is an assignment operator. Intuitively, the current deformation label



(a) Random initialization (b) 1 iteration

(c) 3 iterations (d) After smoothing

Fig. 3. Intermediate texture flows during the randomized inference process: (a) random initialization, (b) 1 iteration, (c) 3 iterations, (d) after smoothing.

$f(\mathbf{p})$ is replaced with the label that provides the smallest matching cost among candidate labels $\Omega_{\mathbf{p}}$. Also, the smallest matching cost is stored in the distance map $c(\mathbf{p})$. This process helps improve the convergence, since neighboring pixels tend to have similar orientation and scale in natural images. On even-numbered iteration, the propagation is performed in reverse scan order: $\Omega_{\mathbf{p}} = \{f(\mathbf{p}), f(\mathbf{p}+(1,0)), f(\mathbf{p}+(0,1))\}$.

Random Search. In the second step, a random search proceeds to prevent the flow result being trapped in a local minimum. We update the current optimal label $f(\mathbf{p})$ by a sequence of random trials which are constructed by sampling around $f(\mathbf{p})$ at an exponentially decreasing distance as

$$(\theta_{\mathbf{p}}^i, s_{\mathbf{p}}^i)^T = (\theta_{\mathbf{p}}, s_{\mathbf{p}})^T + \alpha^i \mathbf{R}_i \mathbf{Z}, \;\; i = 0, 1, 2, \ldots, \tag{8}$$

where $\mathbf{R}_i$ is a 2×2 diagonal matrix whose diagonal entries are uniform random numbers in $[-1, 1] \times [-1, 1]$, α^i is the i^{th} exponential of a ratio $\alpha = 0.5$, and $\mathbf{Z} = (\pi, 2.0)^T$ is the maximum search range. The index i increases until the search radius of the first entry in $\alpha^i \mathbf{Z}$ is below 1. Using this sequence, the current label $f(\mathbf{p})$ is refined if the target random pair has a smaller cost.

Fig. 3 shows intermediate results of the flow field during iterations. Starting from random orientations of Fig. 3(a), the orientation fields are progressively evolved during iterations. As a result, the resultant orientation fields are locally aligned with the visual structure of an input texture image, as shown in Fig. 3(c). In addition, despite the variation in appearance between the example patch and the sample patch from the input image, the local behavior of underlying texture flows are well captured using the visual similarity directly measured on the appearance space, thanks to our randomized deformable matching in the high-dimensional search space. However, the inference mechanism is inherently local, thus often missing a spatial coherency in the estimation. Instead of using the costly global optimization, we resolve this problem by explicitly imposing the smoothness prior on the deformation field via a local filtering approach, which will be detailed in the following section.

(a) Distance map $c(\mathbf{p})$ (b) Initial $s_{\mathbf{p}}$ (c) Smoothed $s_{\mathbf{p}}$

Fig. 4. Results on scale smoothing: (a) distance map, scale fields (b) without smoothing and (c) with smoothing.

2.3. Field Smoothing

After the deformation labels are inferred through the randomized search, the nonlinear vector field smoothing is performed with the guidance of the distance map $c(\mathbf{p})$ as a matching confidence in order to enforce global consistency. We extend the work of [8] by introducing the scale field as well as the orientation field into the smoothing process as a two-tuple flow vector. Intuitively, dominant flow vectors having a smaller matching distance are preserved, while weak flow vectors are directed to follow neighboring dominant ones. The field smoothing is defined as follows:

$$\hat{f}(\mathbf{p}) = \mathbf{K}^{-1}(\mathbf{p}) \sum_{\mathbf{q} \in \mathcal{N}_f(\mathbf{p})} \mathbf{W}(\mathbf{p}, \mathbf{q}) f(\mathbf{q}), \tag{9}$$

where $\mathbf{K}^{-1}(\mathbf{p})$ is a diagonal matrix for a normalization and $\mathcal{N}_f(\mathbf{p})$ denotes the neighborhood of $\mathbf{p}$. $\mathbf{W}$ is a 2×2 diagonal weighting matrix, which is defined as:

$$\mathbf{W}(\mathbf{p}, \mathbf{q}) = \begin{bmatrix} w_r(\mathbf{p}, \mathbf{q}) w_d(\mathbf{p}, \mathbf{q}) & 0 \\ 0 & w_r(\mathbf{p}, \mathbf{q}) \end{bmatrix}, \tag{10}$$

where w_r and w_d represent the range kernel and the direction kernel, respectively. The range kernel w_r encourages dominant orientations and scales to be preserved during smoothing, which is defined as follows:

$$w_r(\mathbf{p}, \mathbf{q}) = \frac{1}{2}(1 + \tanh[\bar{c}(\mathbf{p}) - \bar{c}(\mathbf{q})]), \tag{11}$$

where $\bar{c}(\mathbf{p})$ represents a normalized matching distance across an entire image. $\tanh(\cdot)$ is a monotonically increasing function with respect to the distance difference $\bar{c}(\mathbf{p}) - \bar{c}(\mathbf{q})$, and thus bigger weights are assigned to the neighboring pixels $\mathbf{q}$ whose matching distances are lower than that of the center $\mathbf{p}$. Accordingly, the pixels having lower matching distances contribute more in the filtering of the deformation field. The direction kernel w_d helps tighter alignment of neighboring orientations, which is defined as:

$$w_d(\mathbf{p}, \mathbf{q}) = |\cos(\theta_{\mathbf{p}} - \theta_{\mathbf{q}})|. \tag{12}$$

The direction weight increases as the difference of two orientations approaches to 0 or π. Note that w_d is only applicable to the orientation field, since the scale field is not directional. For further enhancing global coherency of the intermediate flow fields, the field smoothing is iteratively applied η times. In Fig. 3(d), our smoothing improves global coherency of the orientation field, resulting in a good continuation of texture flows. In addition, as shown in Fig. 4, our smoothing helps remove outliers in the scale estimation, which are caused by severe variations on texture appearance.

3. EXPERIMENTAL RESULTS

We validate the performance of the proposed method on various texture images including both synthetic and natural photographs. Test

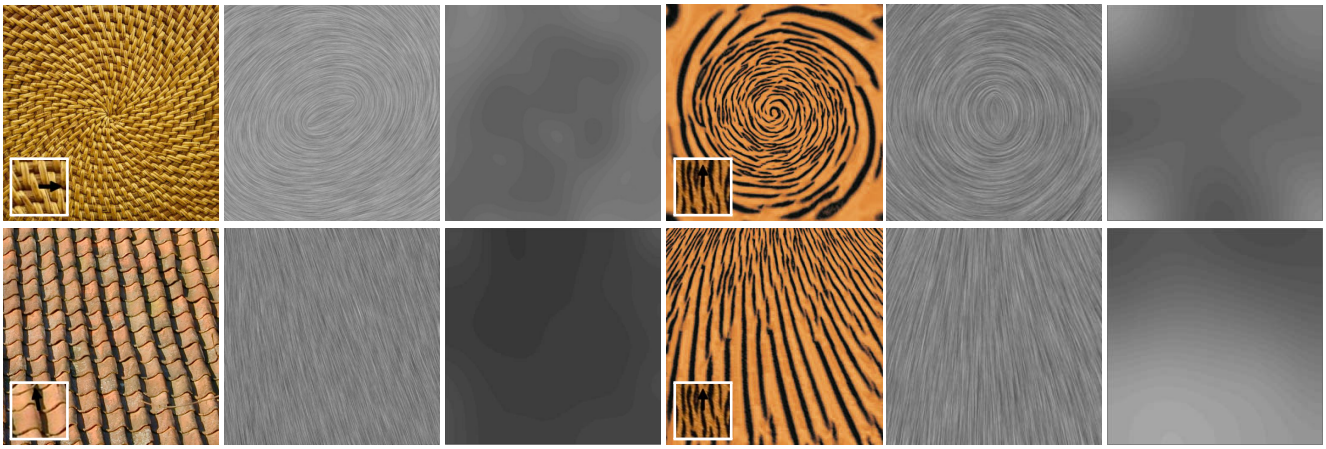
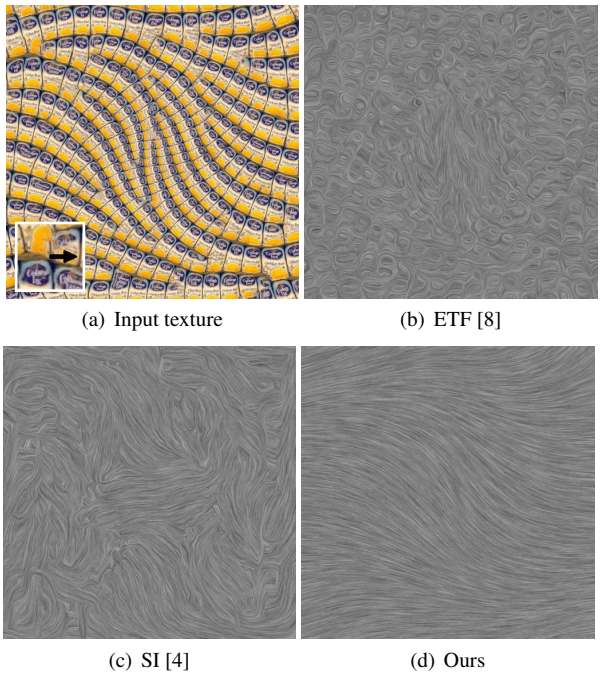


Fig. 5. Experimental results on natural and synthetic images: (from left to right) input textures of size 256×256 with example patches of size 64×64, and estimated orientation and scale fields.

images were selected which undergo sufficient deformation such as rotation and scale transformation, and also show a clear distinction enough to be interpreted by the human visual system (HVS). The user specified example patches with their reference orientation are used that are not distorted and contain sufficient texture elements to represent the input texture image. The proposed method was implemented in Matlab and was simulated on a PC with Quad-core CPU 2.93GHz. In all experiments, the maximum iteration of the randomized inference process is set to 3. For each test image, the window size $\mathcal{N}_f$ and the maximum iteration η of field smoothing is empirically set in the range of $[15, 25]$ and $[5, 15]$ according to an image resolution, respectively. For flow visualization, the line integral convolution (LIC) [9] is used.

Fig. 5 shows the estimated deformation fields on both natural and synthetic textures, which are visually consistent with the human perception in terms of scale and orientation. As shown in the top row of Fig. 5, our method can capture the inherent flows of the circular pattern very well. Moreover, it can be seen that our method produces consistent flows when an input texture has a regular property.

Fig. 6 compares our results with those obtained by two recent deformation estimation approaches: the edge tangent flow filter (ETF) [8] and the statistical invariance approach (SI) [4]. Note that similar to our method, they infer the deformation field locally, not leveraging a global reasoning. Given the input texture of size 256×256 and the example patch of size 64×64, the running times are respectively around 4 seconds (ETF) and 10 seconds (SI) on average, while our method takes 79 seconds (75 seconds for randomized inference and 4 seconds for vector field smoothing) under the typical setting to estimate the deformation field. For a complex input texture as in Fig. 6(a), these two local approaches fail to produce a coherent flow field. These local methods directly calculate the deformation field with a over-simplified model (*e.g.*, intensity gradient), and thus they run faster than ours but cannot discriminate dominant orientation in the presence of complex structure. Moreover, the ETF [8] does not consider the scale field. In contrast, though inherently local, our method formulates the deformation field estimation as the per-pixel labeling framework based on a non-parametric deformation model. This labeling algorithm shares a similar spirit with several global optimization-driven approaches, but our randomized search strategy along with the vector field smoothing enables a much faster inference, with a comparable estimation quality to global approaches. For instance, the global approaches typically take about



(a) Input texture (b) ETF [8]

(c) SI [4] (d) Ours

Fig. 6. Comparison of the estimated flow field with competing methods: (a) input texture, (b) ETF [8], (c) SI [4], (d) proposed method.

$10 \sim 20$ minutes [2]. In addition, our inference algorithm requires only a little extra memory for storing a distance map, unlike existing optimization-driven approaches [2, 5] that typically need huge memory usage to handle the high-dimensional label space.

4. CONCLUSIONS

In this paper, we have addressed the inverse estimation of underlying texture deformation fields. The efficient randomized search enables the direct application of the non-parametric texture model to a high-dimensional search space, and the locally-adaptive vector field smoothing provides an excellent alternative for costly optimization based approaches. Our method produces very promising results compared with existing local methods, and also achieves substantial efficiency gains over global methods.

5. REFERENCES

[1] S. Paris, H.M. Briceño, and F.X. Sillion, "Capture of Hair Geometry From Multiple Images," *in Proc. ACM SIGGRAPH '04*, pp. 712-719, 2004.

[2] Y.-W. Tai, M.S. Brown, and C.-K. Tang, "Robust Estimation of Texture Flow via Dense Feature Sampling," *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.

[3] J. Chang and J.-W. Fisher III, "Analysis of Orientation and Scale in Smoothly Varying Textures," *in Proc. IEEE Int'l Conf. Computer Vision*, pp. 881-888, 2009.

[4] X. Liu, L. Jiang, T.-T. Wong, and C.-W. Fu, "Statistical Invariance for Texture Synthesis," *IEEE Trans. Visualization and Computer Graphics*, vol. 18, no. 11, pp. 1836-1848, Nov. 2012.

[5] M. Park, K. Brocklehurst, R.-T. Collins, and Y. Liu, "Deformed Lattice Detection in Real-World Images Using Mean-Shift Belief Propagation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1804-1816, Oct. 2009.

[6] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, "Texture Optimization for Example-based Synthesis," *in Proc. ACM SIGGRAPH '05*, 2005.

[7] C. Barnes, E. Shechtman, A. Finkelstein, and D.-B. Goldman, "PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing," *in Proc. ACM SIGGRAPH '09*, 2009.

[8] H. Kang, S. Lee, and C.K. Chui, "Flow-Based Image Abstraction," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 1, pp. 62-76, Jan. 2009.

[9] B. Cabral and L.C Leedom, "Imaging Vector Fields Using Line Integral Convolution," *in Proc. ACM SIGGRAPH '93*, pp. 263-270, 1993.