“Visual textures” are regions of images that exhibit some form of spatial regularity. In applications such as texture synthesis and classification, algorithms require a small texture to be provided as an input, which is assumed to be representative of a larger region to be re-synthesized or categorized. We aim to characterize and infer such representatives automatically. We construct a new representation that compactly summarizes a texture, while using significantly less storage, that can be used for texture compression and synthesis.

To characterize visual textures we use the notions of Markovianity, stationarity and ergodicity. A texture is then defined as a region $\Omega$ of an image $I$ that can be rectified into a sample of a stochastic process that is stationary, ergodic and Markovian. It is parametrized by (a) the Markov neighborhood $\omega$ and its Markov scale $r = |\omega|$, (b) the stationarity region $\bar{\Omega}$ and its stationarity scale $\sigma = |\bar{\Omega}|$, (c) a sufficient statistic $\theta_{\omega}$ defined on $\omega$, and (d) $\Omega$, the texture region. Note that $\omega \subset \Omega$. In describing a texture, we seek the smallest $\omega$, in the sense of minimum area (“scale”) $|\omega| = r$, so the corresponding $\theta_{\omega}$ is a minimal (Markov) sufficient statistic.

In a non-parametric setting, $\theta_{\omega}$ is a collection of intensity values. $\omega = \bigcup_{k=1,...,K} \omega_k$ is the union of $K$ sample regions $\omega_k$. Collectively the neighborhoods capture the variability of the texture. A texture is represented by $\theta_{\omega} = \{\theta_{\omega,1}, \theta_{\omega,2}, \ldots, \theta_{\omega,M}\}$.

Collectively the neighborhoods $\omega$ define the texture space $\bar{\Omega}$. For each neighborhood $\omega_k$, we synthesize a novel instance of the texture, starting from $\omega_k$ and its nearest neighbor in $\omega_k$. The nearest neighbor $\omega_k$ is uniquely specified by the image given $r$ and $\omega_k$ (Fig. 1).

Given a representation $\{\omega, \omega, \theta_{\omega}\}$, we can synthesize novel instances of the texture by sampling from $dP(I(\omega))$ within $\omega_k$. We choose a subset of neighborhoods from $\omega$ that satisfy the compatibility conditions and by construction also respect the Markov structure. We perform this selection and simultaneously also infer $I$ by minimizing [1],

$$E(I, \{\omega_k\}) = \sum_{\omega_k \in \Omega} v_{\omega_k}||I(\omega_k) - I(\omega_k)||^2.$$  

An illustration of the quantities involved is shown in Fig. 1. $v_{\omega_k}$ is used to reduce the effect of outliers. The process is performed in a multi-scale and multi-resolution fashion.

We extend synthesis to video, by performing synthesis using a causal approach. We use the already synthesized frames from previous time steps as a boundary condition and extend the textures to the next frame. Using a causal approach we also synthesize multiple textures simultaneously for video and images without computing a segmentation map. This is useful for applications such as video compression, hole-filling and frame interpolation (see Fig. 2). Boundary conditions are implicitly defined by the computed “structure” regions of the videos.

To evaluate the quality of the texture synthesis algorithm, we need a criterion that measures the similarity of the input, $I$, and synthesized, $\hat{I}$, textures. We introduce the Texture Qualitative Criterion ($TQC$), represented by $E_{TQC}$, which is composed of two terms. The first, $E_1(I, \hat{I})$, penalizes structural dissimilarity, whereas $E_2(I, \hat{I})$ penalizes statistical dissimilarity. Let $\omega = \omega_k$ be patches within $\Omega$, the domains of $\Omega$, and their nearest neighbors be $\omega_k/\omega_k$, which are selected within the domains of $I/\hat{I}$.

$$E_1(I, \hat{I}) = \frac{1}{2N} \sum_{r=1}^{N} ||I(\omega_k) - \hat{I}(\omega_k)||^2,$$

$$E_2(I, \hat{I}) = \frac{1}{2N} \sum_{r=1}^{N} ||\phi(I(\omega_k)) - \phi(\hat{I}(\omega_k))||^2.$$