Modeling 2D+1 textures

G. Menegaz and S. Valaeys
Audiovisual Communications Laboratory, Swiss Federal Institute of Technology

Abstract. We propose a novel wavelet based modeling technique for 2D+1 textures, i.e. static textures shot by a moving camera. The correct *perception* of motion is preserved by keeping unchanged the the temporal correlation between subsequent images, or frames. Global motion estimation is used to determine the movement of the camera and to identify the overlapping region between two successive texture images. Such an information is then exploited for the generation of the texture movies. The proposed method for synthesizing 2D+1 textures is able to emulate any piece-wise linear trajectory and is real-time on PIII processors.

1. Introduction

Dynamic textures are usually meant as multi-dimensional stochastic processes exhibiting some stationarity over time [1]. Some examples are smoke, waves and foliage. This can be regarded as a generalization of the bi-dimensional case, where temporal evolution is a feature of the global stochastic process [1, 2].

The novelty of our contribution is that we address the problem of modeling a different class of dynamic textures, for which the motion is not an intrinsic property of the considered process, but the result of a continuous change of the viewpoint. We aim at modeling the motion features as perceived by a moving observer. To make the distinction with respect of the 3D dynamic processes mentioned above, we call the considered class *2D+1 Texture Movies (TM)*. In this case, the key point is the preservation of the temporal correlation between subsequent images, or frames. We consider here the case of a static texture - the grass - shot by a moving camera, and generalize the DWT based Multiresolution Probabilistic Texture Modeling (MPTM) technique [3] to such a dynamic texture. Probabilistic modeling of static textures aims at generating a new image from a sample texture, such that it is *sufficiently different* from the original yet appears to be generated by the same underlying stochastic process. The goal of the proposed algorithm is to generalize such an idea to the generation of a progressively "growing" texture, where the direction and speed of growth is given *a-priori* by a predefined motion model. More specifically, we focus here on piece-wise linear trajectories. In this case, the main issue is the preservation of the perception of motion, namely the preservation of those visual features. Noteworthy, the trivial juxtaposition of temporally subsequent patches respectively sampled from successive frames is not a solution. The aliasing phenomena due to the sampling as well as the mismatch between the sampling grids associated to two successive frames would result in a discontinuity along the boundary.

This paper is organized as follows. Sec. 2 describes the 2D+1 texture model. Sec. 3 illustrates the movement-simulating algorithm. Results are discussed in Sec. 4 and Sec. 5 derives conclusions.

2. Modeling 2D+1 texture movies

Let Ω be the infinite lattice, and let Ω_t be the domain which is observed at time t, i.e. the spatial support associated with the observation at a given instant in time. Let then $I(\Omega_t)$ be the observation at time t and $\tilde{I}(\Omega_t)$ be the synthetic counterpart. Clearly:

$$\Omega_t \subset \Omega \qquad \forall t \tag{1}$$

Accordingly, Ω_{t_1} and Ω_{t_2} denote the domains covered by the observations at times t_1 and t_2, respectively. The specificity of the proposed approach is that it provides a solution to the following problem: *Given two sub-lattices Ω_{t_1} and Ω_{t_2} such that:*

$$\Omega_{t_1} \cap \Omega_{t_2} = \Omega_{\Delta t} \neq \emptyset, \tag{2}$$

generate a synthetic texture over Ω_{t_2} by growing it from the seed already present on $\Omega_{\Delta t}$ such that the impression of visual continuity is preserved.

If the two sets were disjoint, then the independent generation of the texture over the two domains would have been adequate. Conversely, where there is an overlap between the two domains, the independent generation of the texture would produce an apparent edge at the boundary or, equivalently, a flickering on the representation as a temporal sequence which destroys the continuity of the visual flow.

The key feature of the proposed model is the ability to synthesize a textures $\tilde{I}(\Omega_{t_2})$ over the domain Ω_{t_2} by growing the texture over $\overline{\Omega}_{\Delta t} = \Omega_{t_2} \setminus \Omega_{\Delta t}$ but keeping unchanged the texture already present over $\Omega_{\Delta t}$ and avoiding discontinuities along the boundary. The previous discussion holds unchanged also when the observations are themselves realizations of the stochastic process represented by the considered model for static textures. In this case, the following relation holds:

$$\tilde{I}(\Omega_{t+\Delta t}) = \tilde{I}(\Omega_t) \oplus \tilde{I}(\overline{\Omega}_{\Delta t}) \tag{3}$$

where $\tilde{I}(\Omega_{t+\Delta t})$ is the synthetic texture simulating the observation at time $t+\Delta t$, $\tilde{I}(\Omega_{\Delta t})$ is the texture seed, and the operator $\oplus$ indicates the juxtaposition of the textures stated. The spatial position of $\Omega_{t+\Delta t}$ can be easily recovered from the underlying motion model. Let $x, y \in \mathbb{R}$ be the spatial coordinates of the upper left corner of Ω_t and let h and w, with $h, w \in \mathbb{R}_*^+$, be the height, respectively the width, of the spatial domain Ω_t, assumed to be of rectangular shape. Given the estimated speed $\overrightarrow{v} = (v_x, v_y)$ at which the viewpoint moves, it is straightforward to derive the position of the domain $\Omega_{t+\Delta t}$ concerned by the observation at time $t+\Delta t$ as the one whose upper left corner has coordinates:

$$\begin{aligned} x + \Delta x &= x + v_x \cdot \Delta t \\ y + \Delta y &= y + v_y \cdot \Delta t \end{aligned} \tag{4}$$

Therefore, one can easily identify $\Omega_{\Delta t}$ and $\overline{\Omega}_{\Delta t}$.

3. Generalizing the DTW-MPTM to 2D+1 textures

The proposed method is s generalization of the DWT-MPTM to 2D+1 textures. It consists in synthesizing a texture area larger than the video frame size, preserving the texture over $\Omega_{\Delta t}$ while generating a limited amount of new texture, only when necessary, to cover $\overline{\Omega}_{\Delta t}$ avoiding discontinuities. It is worth pointing out that the straightforward solution of synthesizing each frame independently with the DWT-MPTM is not suitable because it creates a disjointed succession of rapid texture changes that fails to generate an impression of movement. One also quickly comes to the conclusion that a cut-and-paste approach at image level, in which the common section is correctly displaced and remaining empty parts of the frame are filled with newly synthesized patches of texture, creates unacceptable discontinuities. Another trivial solution would be to synthesize a much larger texture area than the frame size and to select the covered domain to be part of the frame according to the camera movement. This method is however suffers of some shortcomings. First, the required size of the synthetic texture should be known *a-priori*. Moreover, large amounts of texture could be produced without ever being needed.

A way to answer those concerns is to work in feature space. Although the DWT used for compression purposes is in general not shift-covariant, covariance properties hold for translations in transform space which correspond to translations at image level that can be broken down in horizontal and vertical shifts of $k \cdot 2^N$ and $h \cdot 2^N$ pixels, respectively, where

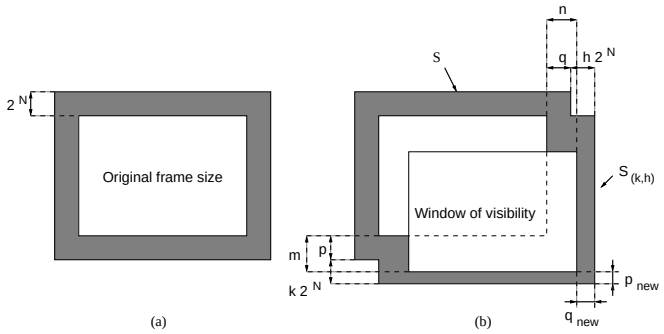

Figure 1. Simulation of movement. The size of S is bigger than that of the original. When the added border is not enough to simulate the required movement, a shifted version of S is created. The window of visibility is then moved inside $S_{(k,h)}$ to reproduce the correct movement.

$k,h \in \mathbb{Z}$ and N is the number of decomposition levels of the DWT. Working in feature space, we are consequently able to generate from a synthetic image S the following set:

$$\Gamma = \{S_{(k,h)}, k,h \in \mathbb{Z} | S_{(k,h)} = T_{(k,h)} S\} \tag{5}$$

where $T_{(k,h)}$ is the translation operator that applied to S produces a shift of $k \cdot 2^N$ and $h \cdot 2^N$ units in the horizontal and vertical directions, respectively. As the translation from S to $S_{(k,l)}$ takes in fact place in feature space, the remaining empty parts of $S_{(k,l)}$ can be filled by applying the DWT-MPTM algorithm locally without creating discontinuities. Any random translation can be obtained by extending the size of S so as to add to it a border of 2^N pixels on all sides. Accordingly, simulating a random translation is a two-step process: obtaining the correct $S_{(k,h)}$; selecting the correct area of $S_{(k,h)}$ which corresponds to the video frame. An example is shown in fig. 1. Let p and q, with $p,q \in \{0, 2 \cdot 2^N\}$ be the width in pixels of the border zone respectively in the horizontal and vertical direction of movement. To generate a horizontal, respectively vertical, movement of m, respectively n, pixels at image level, with $m \geq p$ corresponding to Δx and $n \geq q$ corresponding to Δy in Sec. 2, the correct $S_{(k,h)}$ is chosen so that:

$$k = \min_{\tilde{k}} \{(p + \tilde{k} \cdot 2^N) \geq m\} \tag{6}$$

$$h = \min_{\tilde{h}} \{(q + \tilde{h} \cdot 2^N) \geq n\} \tag{7}$$

The window of visibility is then correctly positioned inside $S_{(k,h)}$, namely:

$$new\_p = p + k \cdot 2^N - m \tag{8}$$

$$new\_q = q + h \cdot 2^N - n \tag{9}$$

4. Results and Discussion

The performance of the proposed system has been evaluated in terms of *preservation of the perceptual features*. Before tackling this subject, it is important to mention that despite the great amount of research devoted to the identification of the *perceptual features* which determine texture perception, the problem is still unsolved. Two main guidelines can be identified. The first is based on the assumption that there exists a set of statistics which is *necessary and sufficient* to identify a *texture class*. Under such an hypothesis, a pair of textures sharing those statistics are *perceptually equivalent* [4, 8]. The problem is faced in

an information theoretic manner, and leads to the definition of models based on statistical parameters. The way such parameters map to the hypothesized necessary and sufficient set is still unknown. The second consists in looking at the problem in a different perspective and aims at identifying and characterizing the *visual mechanisms* which are responsible for texture perception (referring to particular tasks like analysis and discrimination) on a psychophysical or, more in general, neuro-physiological basis [5, 6, 7]. The focus is on the *local* parameters that are relevant for the analysis of the visual stimulus with regard to the considered task, as opposed to the analysis by synthesis approach followed in the other case. The problem is very complex and further investigation is needed to understand and model the involved visual processes. In this contribution, we do not put forward a general theory for texture perception neither a golden rule for the evaluation of a modeling technique. In our opinion, the visual mechanisms which subserve these phenomena need to be investigated further before being able to formulate a general theory. Instead, we have focused on a particular case - driven by an application - and we have faced it in an empirical way leading to what can be considered a "first order" solution. The identification of the features which determine the classification of the texture as belonging to a given class as well as the impression of motion will be an essential part of our future research.

The evaluation of the ability of any texture model, either static or dynamic, to reproduce the perceptual features of the original texture implies *ad-hoc* subjective tests respecting the paradigm set by psychophysics. However, as mentioned above, such a task was beyond the scope of this contribution. Nevertheless, some informal subjective tests involving non trained people of the laboratory revealed that the majority of the subjects were not able to distinguish between the original and synthetic samples.

5. Conclusion

We propose a novel generative model for 2D+1 textures, suitable for model-based coding of video. The integration of the motion information within the DWT-MPTM algorithm for static textures results in an dynamic generative model able to synthesize any 2D+1 texture movie with any piece-wise linear trajectory. A texture seed is extracted from a frame of the original sequence and is used as model for synthesizing the other frames. The motion vector field is estimated at any frame and is used to constrain the generating process such that the correct temporal correlation between the images is preserved. Among the issues deserving further investigation are the emulation of other camera functions like zooming and rotation as well as the rendering of perspective.

References

[1] Soatto S, Doretto G and Wu Y N Dynamic Textures, Proc. of the International Conference on Computer Vision (ICCV), 2001

[2] Bar-Joseph Z, El-Yaniv R, Lischinski D and Werman M, Proc. IEEE Visualization, 403-410, Oct. 2001

[3] Menegaz G, DWT-Based Non-Parametric Texture Modeling, Proc. of the International Conference on Image Processing (ICIP), Thessaloniki, Greece, Oct. 2001

[4] Simoncelli E P and Olshausen B A, Natural Image Statistics and Neural Representation, Annual Review of Neuroscience, 24, 1193-1215, 2001

[5] Landy M S, Texture perception, Encyclopedia of Neuroscience, Elsevier, Amsterdam, The Neatherlans 1996

[6] Wolfson S S and Landy M S, Examining Edge- and Region-based Texture Analysis Mechanisms, 38(3), 439-446, Vision Research, 1998

[7] Li Z, Modeling pre-attentive stereo grouping by intracortical interactions in early visual cortex, Journal of Vision, 1(3), 2001

[8] Portilla J and Simoncelli E P, A parametric texture model based on joint statistics of wavelet coefficients, International Journal of Computer Vision, 40(1), 49-71, Dec. 2000