Eurographics Symposium on Rendering 2014
Wojciech Jarosz and Pieter Peers
(Guest Editors)

Volume 33 (2014), Number 4

Color Me Noisy: Example-based Rendering of Hand-colored Animations with Temporal Noise Control

J. Fišer,[1†] M. Lukáč,[1] O. Jamriška,[1] M. Čadík,[2] Y. Gingold,[3] P. Asente,[4] and D. Sýkora[1]

[1]CTU in Prague, FEE, [2]Brno University of Technology, [3]George Mason University, [4]Adobe Research

Figure 1: *Examples of hand-colored animations generated using our approach (from left to right): walker (watercolor), teddy (oil pastel), and strongman (watercolor). Note how our method creates variety introducing a desired level of temporal noise while preserving the high-frequency details of the drawing medium and the low-frequency content created by an artist.*

Abstract

We present an example-based approach to rendering hand-colored animations which delivers visual richness comparable to real artwork while enabling control over the amount of perceived temporal noise. This is important both for artistic purposes and viewing comfort, but is tedious or even intractable to achieve manually. We analyse typical features of real hand-colored animations and propose an algorithm that tries to mimic them using only static examples of drawing media. We apply the algorithm to various animations using different drawing media and compare the quality of synthetic results with real artwork. To verify our method perceptually, we conducted experiments confirming that our method delivers distinguishable noise levels and reduces eye strain. Finally, we demonstrate the capabilities of our method to mask imperfections such as shower-door artifacts.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Bitmap and framebuffer operations I.3.4 [Computer Graphics]: Graphics Utilities—Paint systems I.3.m [Computer Graphics]: Miscellaneous—Visual arts

1. Introduction

Hand-colored animation is a type of traditional animation, where each frame is created independently, from scratch. It has a distinct visual style represented by a certain amount of temporal flickering which arises due to misalignment of details in consecutive frames. This characteristic feature lends it a look of liveliness and emotional expressivity, which is being successfully leveraged by critically acclaimed artists such as Bill Plympton and Frédéric Back.

† e-mail:fiserja9@fel.cvut.cz

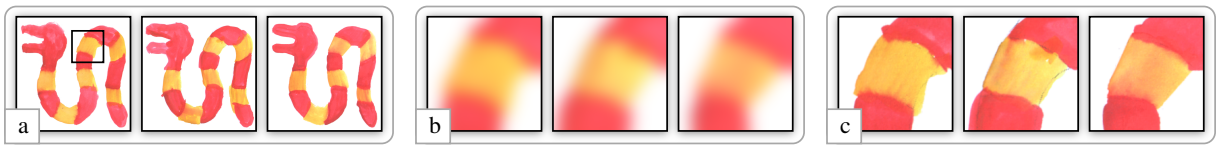

Figure 2: *Motivation—hand-colored animations (a) look temporally coherent when low-pass filtered (b). However, at higher frequencies they contain details that reflect physical properties of the drawing medium and introduce temporal noise (c).*

While temporal noise is usually understood as an undesirable artifact in NPR techniques [BNTS07, BBT09, BBT11, OH12], used judiciously it may serve as an additional medium of artistic expression, either to evoke a hand-crafted look (such as sketchbook scenes in Disney's *Piglet's Big Movie*), or to set a certain mood (e.g., *Shadow World* sequences in *The Lord of the Rings*). In Bill Plympton's more recent work (e.g., *Cheatin'*) noisy, hand-drawn sequences are combined with coherent sequences to convey different moods.

However, the nature of the medium makes it difficult to control the amount of noise, and high noise levels can cause visual fatigue in the viewer. This, in conjunction with the amount of labor involved in production, creates a demand for a more automated process that lets artists control the amount of noise without eliminating it entirely.

Noris et al. [NSC*11] recently presented a system which affords control over the amount of temporal flickering in a sequence of digitally drawn sketches. By registering individual strokes in selected keyframes, they reduce temporal jitter using a weighted combination of original noisy motion and smoothed inbetweening. Although this approach produces impressive reduction of temporal noise level for sketchy vector drawings, it still requires a hand-drawn animation as an input.

Our aim is to reach a more practical workflow that takes a temporally coherent animation created using existing CG pipelines and enriches it with temporal noise synthesized de novo from examples of an arbitrary drawing medium. A similar workflow was recently used by Bénard et al. [BCK*13] in their framework, which extends Image Analogies [HJO*01] to render impressive stylized animations with a specific style or drawing medium given by example. They focus on enforcing temporal coherency using a sophisticated system of correspondence propagation; however, the underlying re-synthesis technique does not permit control over the amount of temporal noise.

In this paper, we propose a novel example-based technique that not only preserve temporal coherency but also introduces a controllable amount of temporal flickering that conveys lively dynamics and visual richness which can be used either to evoke an impression of hand-colored look or provide an additional dimension of expressivity.

2. Related Work

Synthesizing various drawing media is one of the key challenges of non-photorealistic rendering. A wide spectrum of techniques spanning from computational approaches [CAS*97, HLFR07, LXJ12] to realistic example-based methods [ZZXZ09, LBDF13, LFB*13] has been developed. A key issue arises when these techniques are applied to animations, where frame-independent synthesis leads to unpleasant temporal noise that affects viewing comfort.

Many techniques have been proposed to alleviate this issue by enforcing temporal coherency [BNTS07, BBT09, BBT11, OH12]. Although these methods produce visually pleasing results, their visual structure is inconsistent with the natural look of noise typical for hand-colored animation. A similar limitation holds also for procedural noise generation [BLV*10, KP11] which allows for temporally coherent stylization by suppressing temporal components of the generated noise.

Related to synthesis with temporal coherence are methods that try to enforce variety during synthesis [LH05, LH06, RHDG10]. They introduce the ability to vary randomness between scales, but due to being formulated in index domain, they cannot de-couple visual information across scales, which would be necessary for temporal noise control. Related multi-scale texturing approaches [VSLD13] may use a separate source for each scale, but decomposing an example image in this way is problematic.

A different approach to variety synthesis proposed by Assa and Cohen-Or [ACO12] does not rely on texture synthesis but instead decomposes the exemplar into layers, which are then recombined and the result is randomly warped. In our scenario we would like to conform to the user-defined shape, and the remaining small number of discrete varying outputs is insufficient to simulate the variety typical for hand-colored animation.

Our approach is inspired by image morphing techniques [SRAIS10, DSB*12] that extend state-of-the-art image synthesis algorithms [WSI07, SCSI08]. Although these methods have the potential to simulate the look-and-feel of hand-colored animation they do not address the control over the amount of temporal noise.

3. Our Approach

Noise found in hand-colored animations has a specific nature. Artists tend to preserve coherency at a global level—when the sequence is viewed at a distance (see Fig. 2a) or when a low-pass filter is applied (Fig. 2b) the animation is perceived to be temporally coherent. However, at a local level, temporal variance in high-frequency details becomes visible (Fig. 2c). This creates the impression of visual richness, reflecting the real physical properties of the drawing medium used.

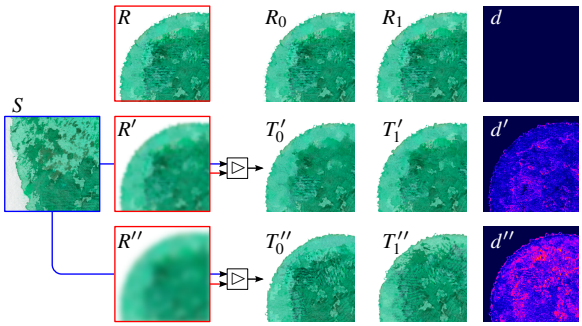

Figure 3: *Synthesis (denoted by ▷ operator) of a noisy target animation T (subscripts denote frame numbers) from a static reference animation R (red input) and a source drawing medium S (blue input). When R is gradually blurred (R', R'') using a low-pass filter h_f with increasing strength f then changes (d', d'') between corresponding synthesized frames (T_0', T_1' and T_0'', T_1'') start to be more apparent and the level of perceived temporal noise increases. Note, however, that individual frames of T' and T'' appear similar when viewed side-by-side with corresponding reference frames in R.*

A characteristic feature of hand-colored animations is that physical properties of the drawing medium are hard to control, so maintaining temporal coherency becomes tedious. The difficulty increases with the scale of details an artist wishes to preserve as coherent. Due to this, hand-colored animations contain specific spatial changes between individual frames that are perceived as a high-frequency *temporal noise* when shown successively. The noise has flat power spectrum (white noise) and gets subjectively stronger when the scale of changing details increases (see Section 4.1 for details).

Vision science offers an explanation of this perception with multi-channel models of human vision [Win05]. When the human visual system processes the temporal signal, two visual mechanisms, the transient and the sustained channels, come into play [KT73, Wat86]. The sustained channel performs a detailed analysis of stationary, or slowly moving, objects (low temporal frequencies) while the transient is involved in signalling the spatial location or change in spatial location (high temporal frequencies). The content of transient channel is therefore perceived as noise, stimulus flickering, or apparent movement [MRW94]. We hypothesize and experimentally measure (see Section 4.1) that the larger the spatial changes in frames, the higher the power spectrum of temporal frequencies, the higher the energy in transient channel, and accordingly the higher the level of perceived noise in animation.

This mechanism motivated us to design a new algorithm that enables control over the amount of perceived temporal noise (see Fig. 3). We render a sequence of images that have similar low-frequency content as the reference animation while high-frequency details are reintroduced by example in a random fashion. The user can then change the frequency threshold to increase/decrease spatial extent of synthesized details and thus control the level of perceived temporal noise.

In the rest of this section we formulate the problem more precisely and propose an algorithm to solve it. We also briefly mention simple extensions that can further improve the quality of the resulting image sequences.

3.1. Problem formulation

The input to our algorithm is a sample of a real drawing medium S and a sequence of N reference images R that represent a coherent, noise-free animation with a similar appearance to S (see Fig. 3). The task is to synthesise a target animation T that satisfies the following three criteria (see Fig. 4):

1. **Fine consistency.** Visual dissimilarity between source S and target T_i should remain small (i is the frame number). This can be accomplished by minimizing established patch-based energy [WSI07]:
$$\sum_{q \in T_i} \min_{p \in S} ||P - Q||_2^2 \qquad (1)$$
where Q denotes a patch of size $w \times w$ centered at the target pixel q, and P is a patch of the same size taken from source pixel p, possibly undergoing additional geometric transformations (we consider rotations and reflections).
2. **Coarse consistency.** Low-frequency content of T_i should be close to the low-frequency content of R_i. Formally we need to minimize the L_2-norm over all pixels of the low-pass filtered signals:
$$||h_f * R_i - h_f * T_i||_2^2 \qquad (2)$$
where h_f is the low-pass filter with tunable strength f and $*$ is the convolution operator.
3. **Temporal noise.** Suppose R is a sequence showing a static image over several frames and q_i is a 1D function yielding the value of a target pixel $q \in T_i$ at the frame i. We would like q_i to contain a signal with white properties, i.e., its power spectrum $\mathcal{Q}(\omega) = |\mathcal{F}(q)|^2$ should have uniformly distributed energies over all frequency bands. Formally we can express this by minimizing the standard

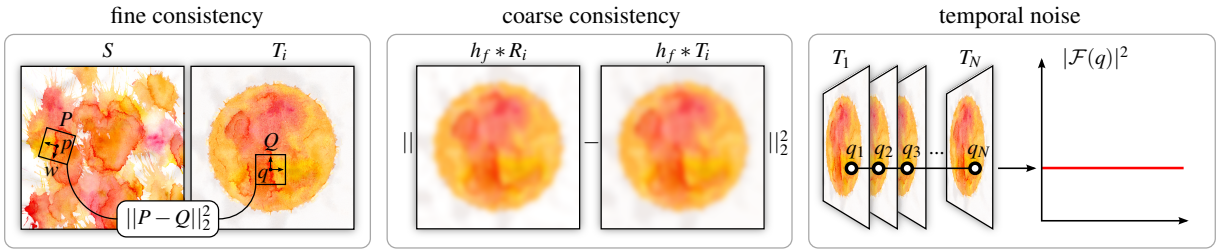

Figure 4: *Problem formulation—using the source drawing medium S and a temporally coherent animation R, the aim is to synthesize a noisy target sequence T that has high-frequency details consistent on a patch level with S (fine consistency) while at lower frequencies being similar to R (coarse consistency). When measuring the power spectrum $|\mathcal{F}(q)|^2$ of motion compensated values stored at a pixel q over time we would like to obtain white noise, i.e., energy distributed uniformly over all frequencies (temporal noise).*

deviation of $\mathcal{Q}$:

$$\frac{1}{N}\sum_{\omega=1}^{N}\left(\mathcal{Q}(\omega)-\frac{1}{N}\sum_{\omega'=1}^{N}\mathcal{Q}(\omega')\right)^2 \tag{3}$$

Such a criterion can also be applied to a more general setting when R contains moving objects. In this case we assume global motion between consecutive frames is compensated before values of q_i are computed.

Note that, surprisingly, in (3) there is no explicit control over the amount of perceived temporal noise. The only aim of (3) is to enforce randomness in the optimization process. Instead, the control is implicitly encoded in the strength f of the low-pass filter h_f used in (2). This follows from our original observation that f can influence the scale of random changes between consecutive frames and thus control the level of perceived temporal noise.

3.2. Algorithm

In this section we propose an algorithm (see Fig. 5) that jointly optimizes the proposed criteria (1–3). It extends the multi-scale EM-like optimization scheme originally proposed by Wexler et al. [WSI07] to find a good local minimum of (1).

Fine consistency. The algorithm of Wexler et al. [WSI07] utilizes image pyramids $\triangle_S$ and $\triangle_T$ to represent the source and target images at multiple scales. It starts with the coarsest level $\ell = 1$ and gradually upsamples the solution until the finest level $\ell = M$ is reached. At each level of the pyramid ℓ the following steps are performed iteratively:

- find nearest neighbor patches $P \subset \triangle_S^\ell$ for all target patches $Q \subset \triangle_T^\ell$ so that $||P-Q||_2^2$ is minimal.
- for each pixel $q \in \triangle_T^\ell$ compute the mode of colors at collocated pixels $p \in \triangle_S^\ell$ that belong to retrieved nearest neighbor patches P.

Coarse consistency. To integrate (2) into the joint optimization process we can exploit the fact that the original Wexler algorithm uses a multi-scale approach to optimize (1). In our setting the synthesis at lower levels of the target pyramid $\triangle_T$ is redundant since from a certain level k a good solution $\triangle_T^k$ is already known: $\triangle_T^k = h_f * R_i \downarrow^f$, where $\downarrow^f$ denotes the downsampling operator that sets an appropriate sampling rate according to the strength f of the low-pass filter h_f. This leads us to propose the following modified version of the original algorithm.

Given the source drawing medium S and the user-specified strength f of the low-pass filter h_f, we initialize source pyramid $\triangle_S$ by low-pass filtering and subsampling S at multiple levels $\ell = 1 \ldots M$:

$$\triangle_S^\ell = h_{\bar{f}(\ell)} * S \downarrow^{\bar{f}(\ell)} \tag{4}$$

where $\bar{f}(\ell)$ is a function which interpolates strength of the low-pass filter h according to the level ℓ. For a box filter where f is the width of the box, $\bar{f}(1) = f$ and $\bar{f}(M) = 1$. Inbetween values are set so that the sampling rate of two consecutive levels decreases with a subtle ratio of 0.85, as in the work of Simakov et al. and Shechtman et al. [SCSI08, SRAIS10], to reach finer granularity during the synthesis and help avoid visual artifacts.

Once the source pyramid is built we create a target pyramid $\triangle_T$ with the same resolution as levels of $\triangle_S$ and enforce (2) by feeding downsampled low-frequency content of the reference animation frame R_i into the coarsest level of $\triangle_T$, i.e., $\triangle_T^1 = h_f * R_i \downarrow^f$. After this initialization the algorithm continues as usual.

Note that successive downsampling of reference animation leads to removal of high-frequency details and introduces fuzziness into the shape of region boundaries. This is a desirable effect which is characteristic for drawing media such as watercolor (see Fig. 2). Nevertheless, there can be situations when these irregularities are unintended. In such case we provide mechanisms that allows to improve the quality of border synthesis using local noise control and source selection. These extensions are further discussed in Section 3.3 and supplementary material.

Temporal noise. Suppose we have the same simplified set-

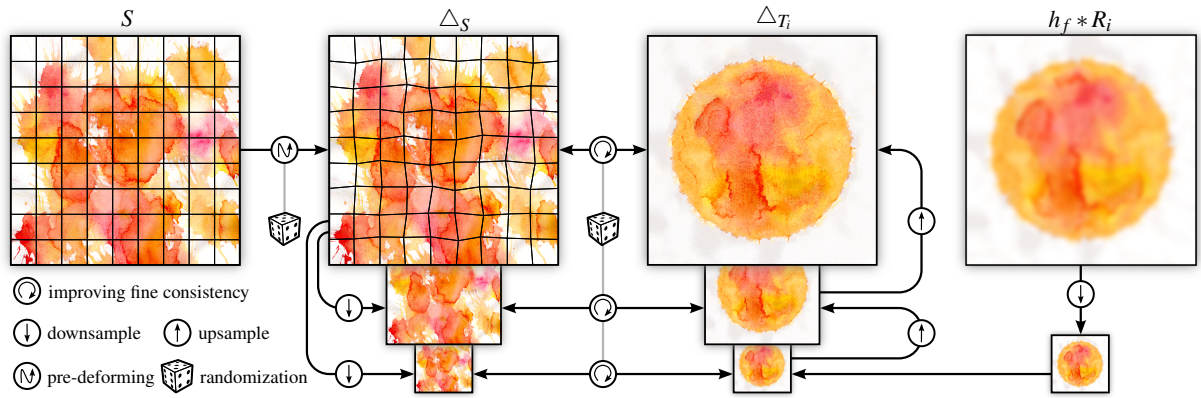

Figure 5: *Algorithm—the source drawing medium S is randomly pre-deformed and image pyramids of source $\triangle_S$ and target frame $\triangle_{T_i}$ are built. The coarsest level of $\triangle_{T_i}$ is initialized by downsampled reference frame R_i. The user-specified strength f of the low-pass filter h_f is used for downsampling. The algorithm starts from coarsest level of $\triangle_{T_i}$ and continues towards finer levels. At each level ℓ fine consistency between $\triangle_S$ and $\triangle_T$ is improved. During this process generalized PatchMatch is utilized to find nearest neighbor patches. The seed for the randomized search is always changed to avoid determinism.*

ting as described in the formulation of *temporal noise* criteria, i.e., a reference animation R that consists of a static image played over several frames. The algorithm proposed so far would lead to a sequence of static images T, where each pixel $q \in T$ would be constant over time. This is the situation we need to avoid as our aim is to produce a noisy sequence.

Direct minimization of (3) would be problematic as it requires computation in the frequency domain, operates over a large amount of data, and for moving objects complex optical flow estimation is necessary to compensate for the global motion. Rather than trying to minimize (3) explicitly we instead synthesize each frame independently and introduce randomness into the original deterministic algorithm by randomly voting over possible patch candidates and pre-deforming the source S. Later (in Section 4.1.1) we demonstrate that such a simplified solution is sufficient to obtain noisy sequences with equally distributed energies over all frequency bands as required by (3).

Recently, *PatchMatch*—a fast approximate nearest neighbor search algorithm [BSFG09, BSGF10] has become popular. Besides significant performance gains, it offers a kind of non-determinism that we can exploit in our scenario. The algorithm uses a random number generator to perform sampling over possible candidates in the space of source patches. Changing the seed of this generator causes the optimization to converge to a different local minimum, changing the appearance of the resulting image.

For low values of f when the synthesis comprises only a few pyramid levels, the likelihood of changes caused by randomized *PatchMatch* reduces significantly. Accordingly, the temporal variance of the resulting sequence T is insufficient to evoke perception of noise in the observer. We attribute this to two known perceptual principles: visual grouping [BL05] and feature fusion [SHK*07]. It was hypothesized that if two visual features have a "common fate" (e.g. they move slowly together in the same direction) and/or are "close enough" in the successive frames, the observer is able to align and fuse them. They are thus perceived as a single object in an apparent motion. An effect of synthetic, unpleasant "floating texture" is then perceived instead the desired noise (see supplemental video for visual inspection).

We address this by randomly pre-deforming the source texture for each synthesized frame, constructing a control lattice with the control points randomly moved in a small radius. The result is deformed using an as-rigid-as-possible moving least squares approach [SMW06]. For our examples we set the grid size to 50 pixels and shift each point 15 to 25 pixels in a random direction. The average offset between synthesized features in two successive frames is above 20 pixels, which corresponds approximately to 20' (visual arcminutes). This value is much higher than the theoretical minimal offset [SHK*07] needed for spatial superposition (2' $\approx$ 2 pixels). This ensures that generated random features are sufficiently far apart to avoid visual fusion.

Note again that the control over the amount of perceived temporal noise is not addressed in this step since it is already encoded in the previous *coarse consistency* phase by setting the strength f of the low-pass filter h_f. The algorithm performs the synthesis starting from the initial coarse solution that corresponds to the low-pass filtered version of R_i and then optimizes for *fine consistency* while using randomization to avoid getting stuck in the same solution. As the scale of randomly synthesized details increases with the increasing strength f the resulting target animation T appears to be more noisy to the observer (see Section 4.1 for evaluation).

3.3. Extensions

The proposed algorithm can be improved further to gain local control over the amount of temporal noise which can help to preserve salient structures (see Fig. 6), make the viewer pay attention to certain parts, or introduce additional channel of artistic expression (see supplementary videos). To enable this control, the isotropic h_f in (2) is replaced with a spatially varying low-pass filter where for each pixel different strength f_p is used. This change is incorporated into our algorithm by setting a different starting level for each pixel, i.e., at pixels with higher f_p the synthesis starts at the coarsest levels of the image pyramid.

Figure 6: *Local noise control—with higher levels of noise the overall shape consistency and presence of small but semantically important features are not guaranteed due to suppression of high-frequency details (left). By specifying the spatially varying strength f_p of the low-pass filter h_f, sensitive parts can be synthesized with a lower noise level and thus preserved (right).*

Orientation of the synthesized strokes (see Fig. 7) can also be controlled locally to emphasize the shape of the animated object or motion orientation. To do that the user can specify two additional orientation fields: O_S for the source drawing medium and O_R for the frames of the reference animation. These can either be obtained automatically, e.g., by computing the per-pixel structure tensor [BCK*13], or painted by the user. When *fine consistency* term (1) is evaluated P is always rotated to compensate for orientation mismatch between P and Q and during the correspondence propagation in PatchMatch [BSFG09], axes-aligned directions are rotated to respect the actual orientation of P.

Figure 7: *Local orientation control—prescribed orientations enable the algorithm to synthesize output that better follows the shape of the target region (right) in contrast to the uncontrolled synthesis (left).*

Besides noise level and orientation, the choice of the source drawing medium can also be controlled locally to improve the quality of the synthesized image. Further details can be found in the supplementary material.

4. Results

We implemented our method using C++ except for PatchMatch [BSFG09], which was implemented in both C++ and CUDA. By default we use simple box filter for the low-pass filter h_f of which the strength f is expressed by the width of the box in pixels. When the source drawing medium contains sharp details a more accurate Lanczos3 filter [DSB*12] can be used to improve visual quality. For the *fine consistency* term we use patches of size $w = 7$ and perform 4 Wexler et al. [WSI07] optimization iterations using 8 PatchMatch iterations at each pyramid level. This number was set empirically to make a balance between effect of randomization and the final visual quality. A lower value causes visual artifacts while a higher value can suppress the effect of randomization as there is higher probability that the algorithm reaches a globally optimal solution.

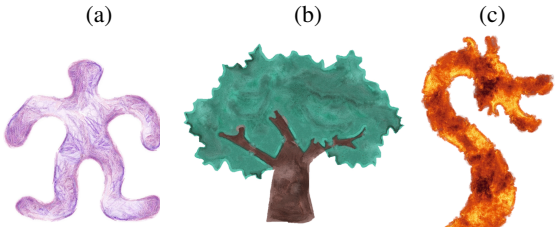

Figure 8: *Results—an additional set of 2D animations: (a) golem [crayon], (b) tree [watercolor], (c) dragon [fire]. See the supplementary video for animations in motion.*

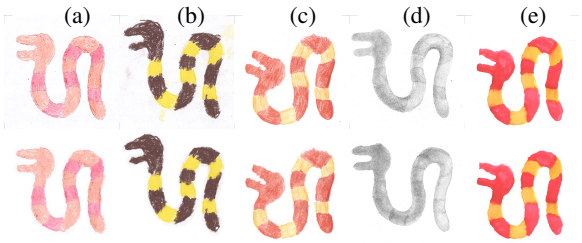

Figure 9: *Examples from the evaluation dataset consisting of hand-colored snakes painted using different drawing media: (a) crayon, (b) chalk, (c) colored pencil, (d) regular pencil, and (e) watercolor. Top row: hand-made, bottom row: synthesized.*

We applied our method to a set of four 2D and two 3D animations (see Fig. 1 and 8). For the 2D cases a shape in a rest pose was created and then a static textured image R_0 was synthesized using [LFB*13] based on a drawing medium S. This image was then deformed using as-rigid-as-possible deformation [SDC09] to produce the temporally coherent animation R. In 3D we mapped textures synthesized from S using [LFB*13] on an animated triangle mesh and rendered the temporally coherent animation R. For each R we synthesized T based on S in various noise levels and played them

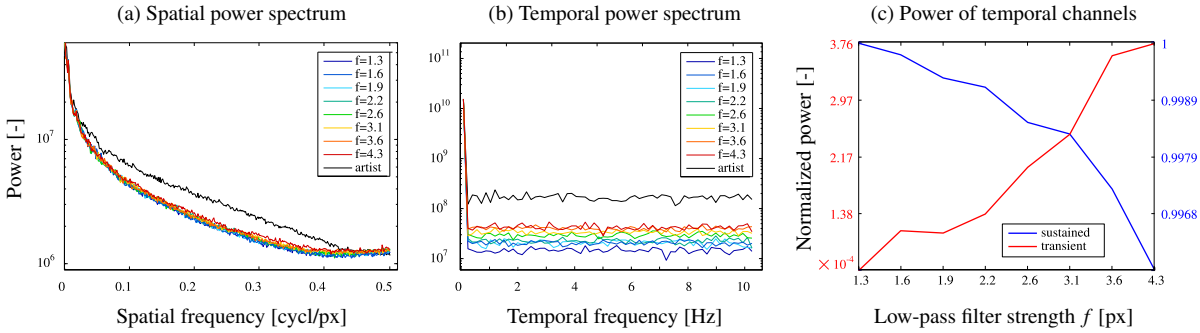

Figure 10: *Spectral analysis—chalk snake in Fig. 9b: (a) Average profiles of the spatial power spectrum of the target frame T_1 synthesized at 8 different strengths f of the low-pass filter h_f and a spectrum profile of the same frame drawn by an artist. (b) Average temporal power spectra of the target sequence T synthesized at 8 different strengths of f and the average spectrum of the animation drawn by an artist (motion in both sequences was compensated). (c) Normalized total power of visual channels for 8 different strengths of f. As f increases, more information is processed by the temporal mechanism and accordingly more temporal noise is perceived.*

sequentially creating the impression of noise slider (see the supplementary video).

The average computation time for one animation frame of size 1Mpix was approximately 30 seconds using one core of a Xeon 3.5GHz and 5 seconds when a CUDA version of PatchMatch was used running on a GeForce GTX 660. A significant speed-up can be reached on multi-core CPUs since the proposed method operates purely in the spatial domain (the temporal coherency of the low-frequency content is implicitly provided by the input sequence) and thus each animation frame can be computed independently.

4.1. Evaluation

To evaluate the proposed method we created a simple experimental animation. It consists of 12 different poses of a striped snake created from a rest pose using as-rigid-as-possible deformation. We printed these poses on a paper using thin outlines and let an artist paint them manually using 5 different drawing media (see Fig. 9, top row and supplementary video). Then we scanned them and performed rectification. As the deformation field is known for each animation frame, we can easily compute its motion compensated version. The resulting hand-colored sequences serve as both ground truth for comparison and examples of drawing media for the synthesis of target sequences (see Fig. 9, bottom row).

While it would be possible to compare the visual plausibility of generated animations against these sequences using a two-alternative forced choice subjective experiment, it should be noted that such a comparison would not in itself be rigorous. This is because the natural animation contains multiple unknown hidden parameters, such as locally varying noise levels or orientation field flickering for anisotropic media, that would have to be matched first. Furthermore, such a comparison would be aesthetic at best, because it is impossible to judge the plausibility of the temporal noise separately from the plausibility of the still image, which is significantly affected by the selected synthesis method.

4.1.1. Spectral Analysis

We analysed the spectral properties of the synthesized sequences for increasing strengths f of the low-pass filter h_f, both in spatial and temporal domains after motion compensation. Results for the chalk sequence (see Fig. 9b) are presented in Fig. 10 (for other media see supplementary material).

In the spatial domain the power spectra of the frames synthesized using different strengths are similar (Fig. 10a), i.e., the overall visual characteristic does not change significantly. The method does not introduce notable over-smoothing with increasing f; only a subtle sharpening effect is visible. When compared to the power spectrum of the real frame a more notable difference indicating subtle over-smoothing is apparent, i.e., the synthesized images do not look as sharp as the original painted by the artist. The amount of this smoothing effect varies across drawing media and is typically small enough so that the synthesized images look convincing (see supplementary material).

In the temporal domain the average power spectrum of T has the energy distributed equally over all frequency bands (Fig. 10b), which corresponds to our aim to obtain characteristics of white noise. It is also visible that the higher the strength f, the higher the overall energy in the temporal spectrum. This indicates increased perception of temporal noise, which can be further verified by measuring power of sustained and transient channels [Win05, ACMS10]. Results

are illustrated in the supplementary material, and overall energies for the chalk sequence are plotted in Fig. 10c. These measurements confirm that the energy in the transient channel grows with the increasing strength f of the low-pass filter h_f and thus the perception of noise level increases.

4.1.2. Subjective Experiments

The spectral analysis above shows evidence that the increasing strength f of the low-pass filter h_f results in corresponding growth of temporal noise. However, the relation between the strength f set by the user and the real quantity of *perceived* temporal noise remains to be investigated. Furthermore, it is also not clear how specific properties of the drawing medium (e.g., crayon, watercolor) affect the visibility of increasing noise level in animations and how this influences eye strain of the observer.

To that end we designed two subjective experiments with 50 and 64 participants, respectively. In the first experiment participants were asked to compare pairs of random sequences generated using our method and for each pair select a sequence that appears more noisy to them. In the second experiment we show just one sequence per question and ask the participants to rate the degree of eye strain they experienced while watching it. There were 4 simulated media: crayon, chalk, colored pencil, and watercolor (we excluded the failure case, regular pencil), and 8 generated levels of noise for each animation, i.e., 32 video stimuli in total.

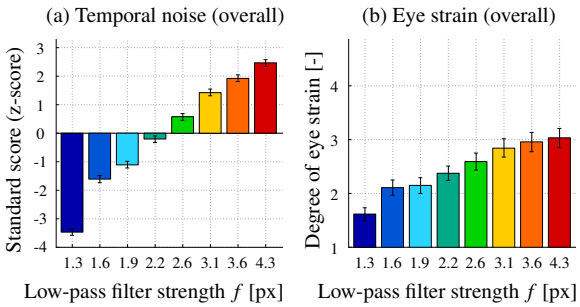

Figure 11: *Overall results of subjective experiments—(a) the two-alternatives-forced-choice study on the perception of temporal noise and (b) the study of eye strain in hand-colored animations synthesized using our method. Error bars show the standard errors.*

The overall results of both studies are shown in Fig. 11. According the ANOVA tests [MR99] the null hypothesis "there is perceptually no difference between levels of temporal noise in the presented sequences" can clearly be rejected ($p \ll 0.001$), meaning that change of low-pass filtering strength f produces sequences with perceptually different noise level. The same holds also for eye strain. The multiple comparison test (Tukey's honestly significant differences [HT87]) returns an overall ranking of the individual noise levels and the eye strain with an indication of the significance of the differences. In the first experiment, there is a statistically significant difference between each level of temporal noise produced by each value of f. The second experiment exhibits two statistically significant groupings: chalk, crayon, watercolor (greater visual discomfort) and watercolor, pencil (lesser visual discomfort).

Furthermore, the first experiment did not show any statistically significant effect of the simulated medium on the level of perceived temporal noise. Nevertheless, the second experiment indicated that there may be a small effect of medium type on the eye strain. Results also indicate slight non-linear relationship between the strength of the low-pass filter f and perceived amount of temporal noise. This motivated possible perceptual linearisation of our method, as shown in the supplementary material.

In summary, both studies confirmed there is a relationship between setting the strength f of the low-pass filter h_f and the levels of perceived temporal noise and eye strain. With increasing f the level of noise and eye strain increases. More details about experiment setup and obtained results can be found in the supplementary material.

4.2. Comparison

For comparison purposes, we have attempted to adapt the methods of [LH05] and [BCK*13] to synthesize results close to our hand-colored animation scenario. Clips comparing these algorithms with our approach are included in the supplementary video.

We have extended [LH05] in order to synthesize an animation with a configurable amount of temporal noise by manipulating the noise/scale settings and using an appropriate source of randomness. While the result was a reasonably consistent noisy sequence, as compared to our approach the algorithm is unable to preserve either high-frequency details of the original drawing medium or the prescribed low-frequency content. It also cannot easily provide local control over the amount of noise.

When attempting to use [BCK*13] we encountered the problem that even when setting different weights to the temporal coherence term the synthesis tends to converge to near-identical results on consecutive frames when the shape or color of the object of interest do not change significantly. The only way to obtain noisy sequence was to deactivate advection vectors and let the algorithm synthesize each frame independently. However, this solution offers only one noise level which cannot be further controlled and there is no guarantee that the resulting sequence will be temporally coherent (see supplementary material for further details).

5. Applications

By combining TexToons [SBCC*11] with our approach, one can produce hand-colored animation from a sequence of

outline-only hand-drawn sketches. Moreover, our technique can mask shower door artifacts that sometimes appear because of the approximative nature of the original TexToons framework (see Fig. 12 and the supplementary video).

Figure 12: *TexToons—the output from the TexToons algorithm is used as a reference for resynthesis. The shifted texture in the original sequence is denoted by the green curve (upper row). With our approach (bottom row) consecutive frames do not suffer from the "shower door" effect.*

Other possible applications of our framework such as stylization and imperfection masking in particle simulations, or painterly rendering of photos and videos can be viewed in supplementary materials.

6. Limitations

An implicit assumption of our method is that areas in the reference animation R have counterparts in the source S that are similar in the RGB domain. As our method draws the samples exclusively from S, absence of a suitable source will change the color of the output to match the most similar one in S (see Fig. 13a–c). If this is not acceptable, color matching could be applied or different exemplar images provided.

A similar situation occurs when S contains multiple areas that have similar average intensity and chroma values and are only distinguished by their fine-scale structure. As the filter eliminates this information above a certain width, the distinction between these areas is lost (see Fig. 13e–f). Synthesis in such situations could be improved if some sort of structural descriptor was taken into account.

When R contains large areas of solid color the algorithm starts to produce artifacts (see Fig. 13g–i). It will also cause the noise level settings to be ineffectual and hinder synthesized frames from carrying the coherency information between frames. To rectify this, one may add some temporally-coherent texture to the solid areas of R as an overlay, using the workflow described in Section 4.

7. Conclusion and Future Work

We presented a new framework that allows the transfer of a hand-colored look to 2D and 3D CG animations. Its ability to control the amount of temporal noise provides a new channel of artistic expression, and enables the creation of longer sequences that are less distracting to the observer yet still preserve a lively hand-colored look. We showed that simply varying the strength of the spatial low-pass filter is sufficient to control the amount of perceived temporal noise, and demonstrated that the algorithm can mask visual artifacts in temporally coherent animations. As a future work we plan to extend it to handle more challenging situations (such as automatically distinguishing areas with different fine-scale structure) and extend local noise control to automatically suppress temporal noise in areas with high edge or saliency detector response.

Acknowledgements

We would like to thank Bill Plympton and Kristýna Mlynaříková for providing excerpts from their animations and all anonymous reviewers for suggestions and constructive comments. This research was funded by Adobe and partly supported by the Technology Agency of the Czech Republic under the research program TE01020415 (V3C – Visual Computing Competence Center), by the Czech Science Foundation under research program P202/12/2413 (OPALIS), by the Grant Agency of the Czech Technical University in Prague, grant No. SGS13/214/OHK3/3T/13 (Research of Progressive Computer Graphics Methods), and by SoMoPro II grant (financial contributions from the EU 7 FP People Programme (Marie Curie Actions), REA 291782, and from the South Moravian Region).

References

[ACMS10] AYDIN T. O., ČADÍK M., MYSZKOWSKI K., SEIDEL H.-P.: Video quality assessment for computer graphics applications. *ACM Transactions on Graphics 29*, 6 (2010), 161. 7

[ACO12] ASSA J., COHEN-OR D.: More of the same: Synthesizing a variety by structural layering. *Computers & Graphics 36*, 4 (2012), 250–256. 2

[BBT09] BÉNARD P., BOUSSEAU A., THOLLOT J.: Dynamic solid textures for real-time coherent stylization. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D)* (2009), pp. 121–127. 2

[BBT11] BÉNARD P., BOUSSEAU A., THOLLOT J.: State-of-the-art report on temporal coherence for stylized animations. *Computer Graphics Forum 30*, 8 (2011), 2367–2386. 2

[BCK*13] BÉNARD P., COLE F., KASS M., MORDATCH I., HEGARTY J., SENN M. S., FLEISCHER K., PESARE D., BREEDEN K.: Stylizing animation by example. *ACM Transactions on Graphics 32*, 4 (2013), 119. 2, 6, 8

[BL05] BLAKE R., LEE S.-H.: The role of temporal structure in human vision. *Behavioral and Cognitive Neuroscience Reviews 4*, 1 (2005), 21–42. 5

[BLV*10] BÉNARD P., LAGAE A., VANGORP P., LEFEBVRE S., DRETTAKIS G., THOLLOT J.: A dynamic noise primitive for coherent stylization. *Computer Graphics Forum 29*, 4 (2010), 1497–1506. 2

[BNTS07] BOUSSEAU A., NEYRET F., THOLLOT J., SALESIN D.: Video watercolorization using bidirectional texture advection. *ACM Transaction on Graphics 26*, 3 (2007), 104. 2

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D. B.: PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics 28*, 3 (2009), 24. 5, 6

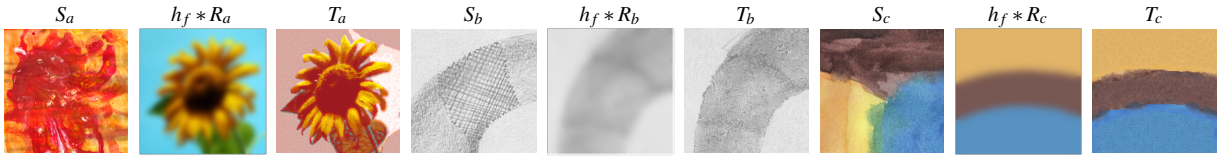

Figure 13: *Limitations—synthesized frames ($T_{a,b,c}$) may not properly convey the look of the drawing medium ($S_{a,b,c}$) when it contains different colors (S_a) from the reference (R_a) or subtle high-frequency details (S_b) that cannot be distinguished by intensity level or color, or when the reference frame contains solid colors (R_c).*

[BSGF10] BARNES C., SHECHTMAN E., GOLDMAN D. B., FINKELSTEIN A.: The generalized PatchMatch correspondence algorithm. In *Proceedings of European Conference on Computer Vision* (2010), pp. 29–43. 5

[CAS*97] CURTIS C. J., ANDERSON S. E., SEIMS J. E., FLEISCHER K. W., SALESIN D. H.: Computer-generated watercolor. In *Proceedings of SIGGRAPH 97* (1997), pp. 421–430. 2

[DSB*12] DARABI S., SHECHTMAN E., BARNES C., GOLDMAN D. B., SEN P.: Image melding: Combining inconsistent images using patch-based synthesis. *ACM Transactions on Graphics 31*, 4 (2012), 82. 2, 6

[HJO*01] HERTZMANN A., JACOBS C. E., OLIVER N., CURLESS B., SALESIN D. H.: Image analogies. In *Proceedings of SIGGRAPH 2001* (2001), pp. 327–340. 2

[HLFR07] HAEVRE W. V., LAERHOVEN T. V., FIORE F. D., REETH F. V.: From dust till drawn. *The Visual Computer 23*, 9-11 (2007), 925–934. 2

[HT87] HOCHBERG Y., TAMHANE A. C.: *Multiple Comparison Procedures*, 1st ed. Wiley, 1987. 8

[KP11] KASS M., PESARE D.: Coherent noise for non-photorealistic rendering. *ACM Transaction on Graphics 30*, 4 (2011), 30. 2

[KT73] KULIKOWSKI J. J., TOLHURST D. J.: Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology 232*, 1 (1973), 149–162. 3

[LBDF13] LU J., BARNES C., DIVERDI S., FINKELSTEIN A.: RealBrush: painting with examples of physical media. *ACM Transactions on Graphics 32*, 4 (2013), 117. 2

[LFB*13] LUKÁČ M., FIŠER J., BAZIN J.-C., JAMRIŠKA O., SORKINE-HORNUNG A., SÝKORA D.: Painting by feature: Texture boundaries for example-based image creation. *ACM Transaction on Graphics 32*, 4 (2013), 116. 2, 6

[LH05] LEFEBVRE S., HOPPE H.: Parallel controllable texture synthesis. *ACM Transactions on Graphics 24*, 3 (2005), 777–786. 2, 8

[LH06] LEFEBVRE S., HOPPE H.: Appearance-space texture synthesis. *ACM Transactions on Graphics 25*, 3 (2006), 541–548. 2

[LXJ12] LU C., XU L., JIA J.: Combining sketch and tone for pencil drawing production. In *Proceedings of International Symposium on Non-Photorealistic Animation and Rendering* (2012), pp. 65–73. 2

[MR99] MONGOMERY D. C., RUNGER G. C.: *Applied Statistics and Probability for Engineers*, 2nd ed. John Wiley & Sons, 1999. 8

[MRW94] MÄKELÄ P., ROVAMO J., WHITAKER D.: Effects of luminance and external temporal noise on flicker sensitivity as a function of stimulus size at various eccentricities. *Vision Research 34*, 15 (1994), 1981–1991. 3

[NSC*11] NORIS G., SÝKORA D., COROS S., WHITED B., SIMMONS M., HORNUNG A., GROSS M., SUMNER R.: Temporal noise control for sketchy animation. In *Proceedings of International Symposium on Non-photorealistic Animation and Rendering* (2011), pp. 93–98. 2

[OH12] O'DONOVAN P., HERTZMANN A.: AniPaint: Interactive painterly animation from video. *IEEE Transactions on Visualization and Computer Graphics 18*, 3 (2012), 475–487. 2

[RHDG10] RISSER E., HAN C., DAHYOT R., GRINSPUN E.: Synthesizing structured image hybrids. *ACM Transactions on Graphics 29*, 4 (2010), 85. 2

[SBCC*11] SÝKORA D., BEN-CHEN M., ČADÍK M., WHITED B., SIMMONS M.: TexToons: Practical texture mapping for hand-drawn cartoon animations. In *Proceedings of International Symposium on Non-photorealistic Animation and Rendering* (2011), pp. 75–83. 8

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2008). 2, 4

[SDC09] SÝKORA D., DINGLIANA J., COLLINS S.: As-rigid-as-possible image registration for hand-drawn cartoon animations. In *Proceedings of International Symposium on Non-photorealistic Animation and Rendering* (2009), pp. 25–33. 6

[SHK*07] SCHARNOWSKI F., HERMENS F., KAMMER T., ÖĞMEN H., HERZOG M. H.: Feature fusion reveals slow and fast visual memories. *Journal of Cognitive Neuroscience 19*, 4 (2007), 632–641. 5

[SMW06] SCHAEFER S., MCPHAIL T., WARREN J.: Image deformation using moving least squares. *ACM Transactions on Graphics 25*, 3 (2006), 533–540. 5

[SRAIS10] SHECHTMAN E., RAV-ACHA A., IRANI M., SEITZ S. M.: Regenerative morphing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2010), pp. 615–622. 2, 4

[VSLD13] VANHOEY K., SAUVAGE B., LARUE F., DISCHLER J.-M.: On-the-fly multi-scale infinite texturing from example. *ACM Transactions on Graphics 32*, 6 (2013). 2

[Wat86] WATSON A. B.: Temporal sensitivity. In *Handbook of Perception and Human Performance*, Boff K. R., Kaufman L., Thomas J. P., (Eds.). John Wiley and Sons, New York, 1986. 3

[Win05] WINKLER S.: *Digital Video Quality: Vision Models and Metrics*. Wiley, 2005. 3, 7

[WSI07] WEXLER Y., SHECHTMAN E., IRANI M.: Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 3 (2007), 463–476. 2, 3, 4, 6

[ZZXZ09] ZENG K., ZHAO M., XIONG C., ZHU S.-C.: From image parsing to painterly rendering. *ACM Transactions on Graphics 29*, 1 (2009), 2. 2