Eurographics Symposium on Rendering 2010
Jason Lawrence and Marc Stamminger
(Guest Editors)

Volume 29 (2010), Number 4

A Dynamic Noise Primitive for Coherent Stylization User Study

P. Bénard[1] A. Lagae[2,3] P. Vangorp[3] S. Lefebvre[4] G. Drettakis[3] J. Thollot[1]

[1]Grenoble University [2]Katholieke Universiteit Leuven [3]INRIA Sophia-Antipolis [4]INRIA Nancy Grand-Est / Loria

Figure 1: *A participant performing the user study.*

This document provides more details about the setup, procedure, and analysis of the user study.

1. Procedure and Setup

We opted for a ranking experiment to compare our 6 selected techniques, since ranking tasks are generally considered more stimulating and enjoyable than alternatives such as pairwise comparisons. A ranking task can be regarded as multiple simultaneous pairwise comparisons at the observer's own pace, and is therefore less time consuming than the alternatives while producing comparable results.

We created a setup with two rows of 6 slots as shown in Fig. 1. Initially the top row of slots contains the 6 stimuli (still images or video loops) of the different methods. The users are asked to rank the stimuli from left to right according to a criterion displayed above, by dragging and dropping them to the bottom row (see also the accompanying video to get a better feeling of the experiment).

Synchronized playback of six high-quality videos at 512×512 resolution each, which can be interactively dragged and dropped from one slot to the other, pushes the limits of today's personal computers and excludes a remote web-based setup. We created a local setup to ensure smooth 25fps video playback using a modified version of the *MPlayer* software and the multi-threaded *FFmpeg-mt* codec library. The video streams were encoded into the H.264 format without bitrate limits.

The local setup limits the number of participants, but on the other hand it also gives us more control over the stimulus presentation conditions. It consequently decreases the variance of the acquired data to make up for the smaller number of participants. We used dual 24" Dell UltraSharp 2407WFP LCD monitors at 1920×1200 resolution to provide a large enough work surface for the ranking task. The displays were calibrated to match brightness, contrast, and color reproduction. The experiment was performed in normal office lighting conditions. Participants were unpaid volunteers and had normal or corrected-to-normal vision. They were given written instructions in their native language or in English (Sec. 2), and were otherwise naive as to the aims of the experiment. Participants were asked to report their overall confidence in their rankings and any difficulties they might have had in a post-study questionnaire.

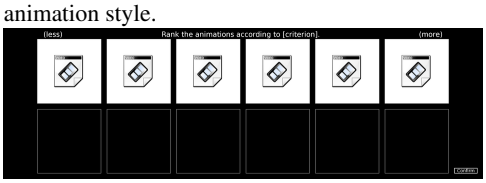
To keep the duration of the sessions below 30 minutes per participant, we decided to split the study into a part involving simple stimuli and a part involving complex stimuli. A total number of 15 volunteers (11 male, 4 female, ages 25–59) participated in both parts of the study. Participants took on average 90 seconds to complete a ranking task, and rarely exceeded 5 minutes even for ranking tasks they reported as difficult.

The black-and-white hatching style was reproduced to the best of our abilities for all the techniques we compared. In particular, a Gabor noise texture with the same parameters was used as input to the previous techniques, and subsequently deformed, advected, or transformed. The color map, in this case a binary threshold, was always applied at the end of the rendering pipeline.

2. Instructions

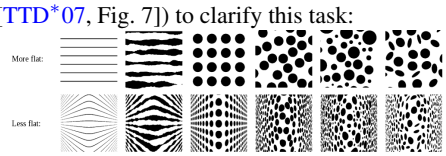
The following instructions were provided to the participants:

Cartoon Animation Perception Survey
The purpose of this survey is to understand how you perceive different variations of a cartoon-like animation style.



We will show 6 images or animation sequences simultaneously. You will be asked to rank them according to a number of very specific criteria, by drag-and-dropping them into the order you decide. A second row of slots is provided for shuffling items around. The items must form a single horizontal axis, i.e. no two items in a single column, before you can continue to the next criterion.
There are no right or wrong answers; the answer reflects only your opinion. If you are unsure about the ranking of some of the items, take your best guess or just put them in a random order.
There are 4 different tasks:

1. "Rank the images according to how flat they appear."
   We provide the following images (taken from [TTD*07, Fig. 7]) to clarify this task:



2. "Rank the animations according to how coherently the pattern moves with the object."
3. "Regardless of the coherence of the motion of the object and the pattern, which you have already evaluated in the previous task, rank the animations according to how much the pattern changes otherwise over time."
4. "Rank the animations according to how pleasant you find them in the context of cartoon animation." (second part only)

You will do 7 rankings (first part) / 4 rankings (second part) in total. The complete survey will typically take less than 30 minutes (first part) / 10 minutes (second part).
Feel free to take a break at any time. You may quit the survey anytime, without having to give a reason and without detriment to you.
Thank you for taking part in this survey!

The post-study questionnaire consisted of the following questions:

Questionnaire

- How confident are you in your rankings for each criterion?
- Were there any criteria you didn't understand? If so, describe your interpretation.
- Were there any ambiguously formulated criteria? If so, describe the ambiguity.
- Were there any criteria for which it was very difficult to rank the video sequences? If so, describe the difficulty.
- Are there other relevant criteria for cartoon animation that weren't covered? If so, describe.
- Was the survey's duration too long, about right, or too short? Were the tasks fun or tiresome?
- Did the written instructions cover everything you needed to know about the survey? What else should have been included?
- Were the written instructions clear? How could they be improved?

3. Analysis

The results of the ranking tasks were analyzed using Thurstonian scaling [Thu27,Ges97] to derive interval scales. The statistical significance of observed trends is confirmed by the Wilcoxon rank-sum hypothesis test [Wil45]. Fig. 2 shows the interval scales and the similarity groups based on the Wilcoxon test at 95% significance level. Fig. 3 and 4 shows the same interval scales with confidence intervals based on 10,000 bootstrap resamplings [Efr79].

References

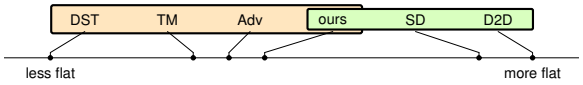
[Efr79] EFRON B.: Bootstrap methods: another look at the jackknife. *Annals of Statistics 7*, 1 (1979), 1–26. 3

[Ges97] GESCHEIDER G. A.: *Psychophysics: The Fundamentals*, 3rd ed. Lawrence Erlbaum Associates, 1997. 3

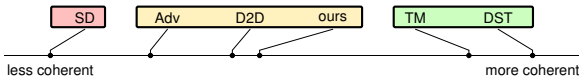
[Thu27] THURSTONE L. L.: A law of comparative judgment. *Psychology Review 34*, 4 (1927), 273–286. 3

[TTD*07] TODD J. T., THALER L., DIJKSTRA T. M. H., KOENDERINK J. J., KAPPERS A. M. L.: The effects of viewing angle, camera angle, and sign of surface curvature on the perception of three-dimensional shape from texture. *J. Vision 7*, 12 (2007), 1–16. 2

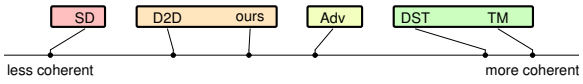
[Wil45] WILCOXON F.: Individual comparisons by ranking methods. *Biometrics Bulletin 1*, 6 (1945), 80–83. 3



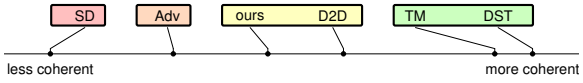
(a) Simple stimuli – Flatness



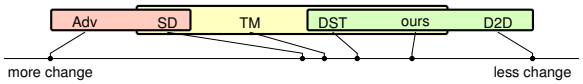
(b) Simple stimuli – Coherent motion – Translation



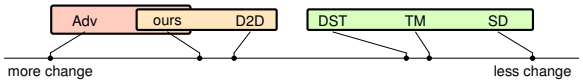
(c) Simple stimuli – Coherent motion – Rotation



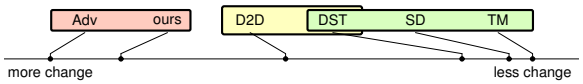
(d) Simple stimuli – Coherent motion – Zoom



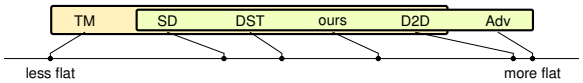
(e) Simple stimuli – Temporal continuity – Translation



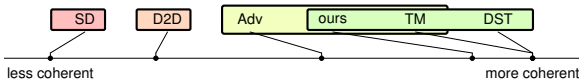
(f) Simple stimuli – Temporal continuity – Rotation



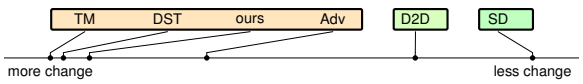
(g) Simple stimuli – Temporal continuity – Zoom



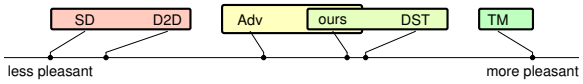
(h) Complex stimuli – Flatness



(i) Complex stimuli – Coherent motion

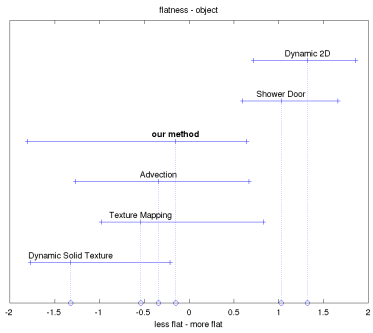


(j) Complex stimuli – Temporal continuity

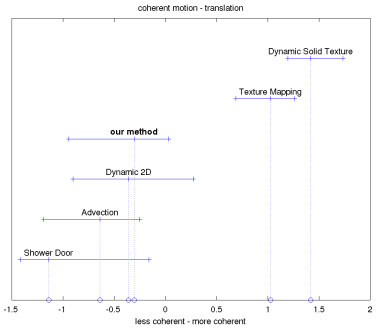


(k) Complex stimuli – Pleasantness

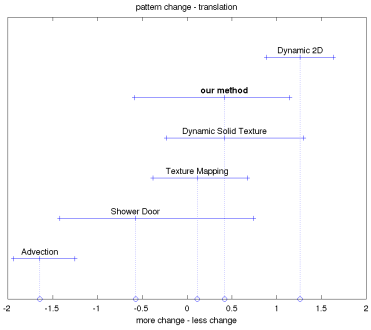
Figure 2: *Interval scales indicating the relative merits of each method. The methods are classified into similarity groups based on the Wilcoxon rank-sum hypothesis test at 95% significance level.*



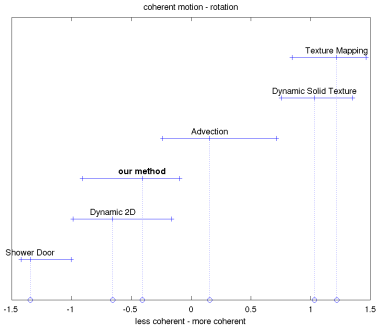
(a) Flatness



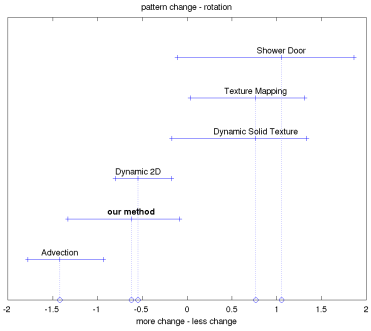
(b) Coherent motion – Translation



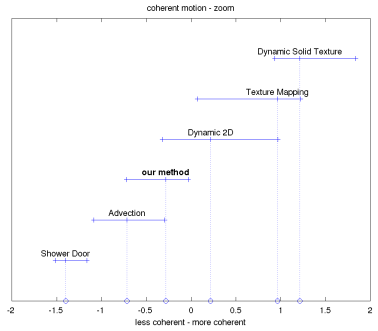
(e) Temporal continuity – Translation



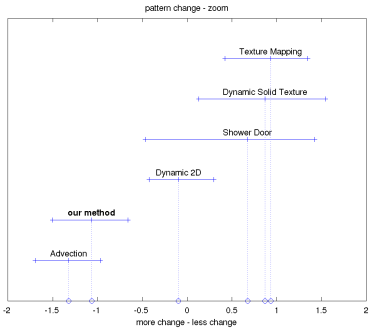
(c) Coherent motion – Rotation



(f) Temporal continuity – Rotation

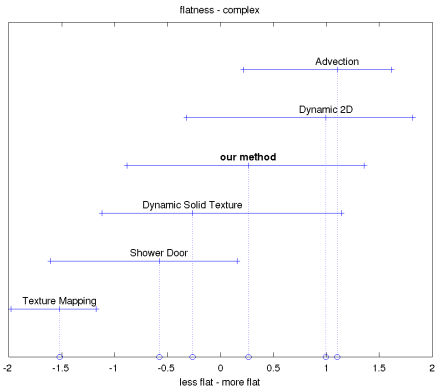


(d) Coherent motion – Zoom

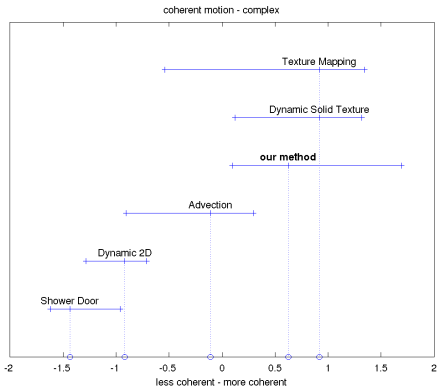


(g) Temporal continuity – Zoom

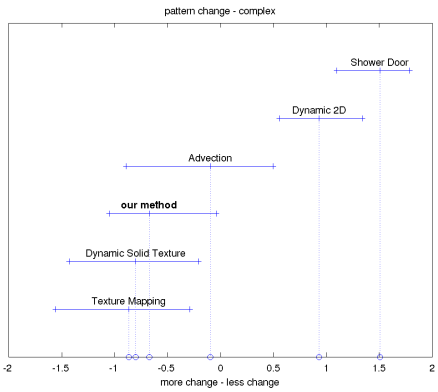
Figure 3: *Simple stimuli: interval scales and 95% confidence intervals based on 10,000 bootstrap resamplings.*



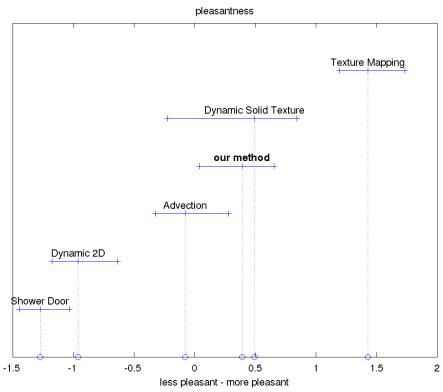
(a) Flatness



(b) Coherent motion



(c) Temporal continuity



(d) Pleasantness

Figure 4: *Complex stimuli: interval scales and 95% confidence intervals based on 10,000 bootstrap resamplings.*