

Perception of Drawing Reference Quality among Professional Hand-drawn Animators

Rachael Schwartz,^{†1,2}  Mark Mullery,²  John Dingliana,¹  Rachel McDonnell¹ 

¹Trinity College Dublin, Graphics & Vision Research Group, Dublin, Ireland

²Cartoon Saloon, Kilkenny, Ireland



Figure 1: *Wolfwalkers* promotional image, provided by and presented with permission from Cartoon Saloon animation studio.

Abstract

We present a preliminary experiment, investigating professional hand-drawn animators' perception of how good one frame is as drawing reference for another. 10 professional hand-drawn animators rated the drawing reference quality of 54 hand-drawn frame pairs, each differing by character pose and region rotation, reflection, and distortion transformations. Our results indicate that animators perceive frames differing by rotation/reflection as better drawing reference than frames differing by distortion.

CCS Concepts

• *Applied computing* → *Fine arts*;

1. Introduction

During hand-drawn animation production, animators keep character design consistent by referencing *model sheets* — character designs with technical drawing instructions — as they work. Since model sheets depict a only small fraction of poses that characters assume in the film, animators also reference existing frames when drawing poses absent from model sheets. However, animators must remember where their desired reference frame appears in the film's 100,000+ frames and tediously sift through dozens of frames to find it. In this paper, we experimentally investigate professional hand-drawn animators' perception of *Reference Frame Similarity (RFS)* — how good one frame is as drawing reference for another — to support automatic reference frame retrieval in future work.

Frames depicting similar body poses can be good drawing reference for one another, and several methods estimate hand-drawn characters' body pose (e.g., [BB22, JZJ*23]). However, two key observations suggest that animators' RFS perception involves both pose similarity and shape/adjacency similarity of *regions*, 2D shapes comprising hand-drawn animated character designs (Figure 2). First, frames with similar pose estimates may be poor drawing reference for each other when region shape/adjacency differs between them (Figure 2, top right). Second, frames with different pose estimates may be good drawing reference for each other when regions differ by rotation/reflection (Figure 2, bottom right), as human visual perception is moderately rotation/reflection-invariant [PESP18] and animators develop enhanced rotation invariance by practicing their trade [Koz01, WWH*11, CKDW19]. This preliminary experiment tests our second observation: professional animators perceive frames depicting physically different poses as good drawing reference when regions differ by planar rotation/reflection.

[†] Funding provided by the Irish Research Council

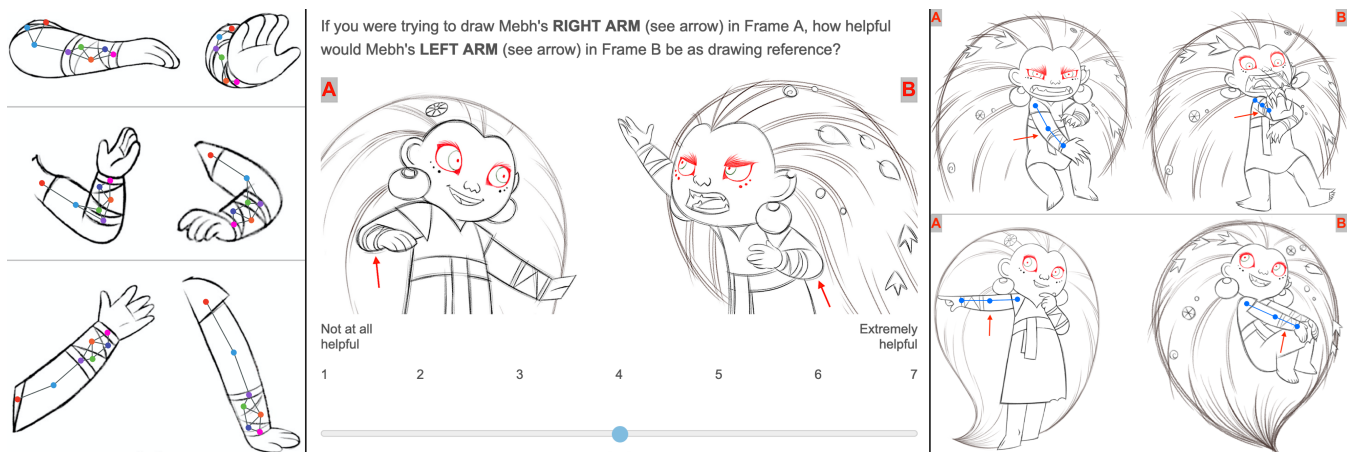


Figure 2: *Left:* Experiment frame pairs demonstrating how each transformation variable affects region shape and adjacency, with region adjacency graphs superimposed for instructional purposes (not shown in experiment). *Top:* [high Distortion, low Reflection, low Rotation]. *Middle:* [low Distortion, high Reflection, low Rotation]. *Bottom:* [low Distortion, low Reflection, high Rotation]. **Center:** Example RFS experiment question. **Right:** Experiment frame pairs with superimposed skeleton poses (in blue), demonstrating that physically similar poses can provide poor drawing reference (*Top*; average RFS rating = 2.09) and physically different poses can provide good drawing reference (*Bottom*; average RFS rating = 5.0). Frames provided by and presented with permission from Cartoon Saloon.

2. Variable & Data Selection

To test our hypothesis, we selected three region transformation variables: Rotation, Reflection, and Distortion. Some combination of these variables usually describes the region shape and adjacency differences between frames depicting the same character's body part in different poses (Figure 2, left). Because Rotation and Reflection preserve region shape and adjacency up to 2D orthogonal transformation, the pose changes they convey do not alter a body part's relative angle to the camera (drawn as *foreshortening*). Only Distortion (non-orthogonal 2D region shape/adjacency change) conveys foreshortening change. We therefore hypothesized that animators' RFS perception would inversely correlate to Distortion more strongly than to Rotation and Reflection.

To optimize variable control, all hand-drawn frames used in our experiment depicted the same character: Mebh, from Cartoon Saloon animation studio's feature film *Wolfwalkers* (<https://www.cartoonsaloon.ie/wolfwalkers/>); frames were provided by Cartoon Saloon. Although Mebh appears in tens of thousands of frames and diverse poses, a frame set offering full variable control over her entire body did not exist. Consequently, participants were asked to consider only one of Mebh's arms in each frame when answering experiment questions (Figure 2, center). Mebh's arm was chosen because it is a complex object comprising multiple regions, appears in myriad articulate poses, and is stylistically consistent throughout the film. All experiment frames show Mebh in the same costume, with the same hairstyle, and without props. Frames were reasonably consistent in scale, line quality, and adherence to model sheet character design. Only frames showing Mebh at eye-level (not from above or below) and facing front (not from behind or side profile) were used. Experiment frames were uncolored to match the majority of frames available at the time in production when animators need drawing references.

Most experiment frames were manually selected from a database of 245 *Wolfwalkers* frames, built by manually scrubbing through the film to find frames with high-quality body pose (e.g., no head shots or blurred motion) and without prop, environment, or other character interactions. These were exported as uncolored line drawings depicting Mebh in isolation. After finalizing experiment design, other frames depicting poses needed for the experiment were found and exported as described above. To model an animation studio setting in our experiment design, it was essential to use original hand-drawn frames rather than editing frames to artificially create specific poses. Edited frames would look noticeably different from original frames and thus would likely affect animators' RFS perception. 64 frames were selected for the experiment and organized into pairs of reference frames.

Rotation, Reflection, and Distortion change between each frame pair's respective arms was evaluated by eye. Finding and implementing an empirical metric accurately simulating animators' visual perception is a complex task, and was beyond the scope of this preliminary experiment. Measurement guidelines were as follows: Rotation: low = 0-10°, medium = 11-89°, high = 90-180°. Reflection: low = no reflection, medium = cross-body reflection (Figure 2, bottom right), high = inside-outside reflection (Figure 2, middle left). Distortion: low = no region shape/adjacency change, medium = mild region shape change without region adjacency change (Figure 2, center), high = strong region shape change with region adjacency change (Figure 2, top left). Professional animators who did not participate in the experiment confirmed that our evaluations were likely to comport with animators' visual perception.

3. Method

Rotation, Reflection, and Distortion were each modulated across low, medium, and high levels (Figure 2, left), creating 27 cases

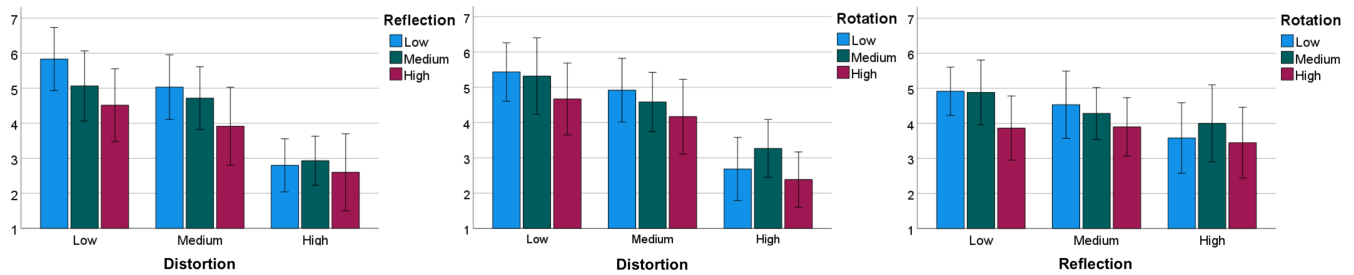


Figure 3: Significant Interactions between (left) Distortion and Reflection, (center) Distortion and Rotation, and (right) Reflection and Rotation. 1 on the scale indicated that the drawing reference was not at all helpful, while 7 meant that it was extremely helpful.

overall. Two frame pairs were selected for each case, yielding 54 experiment questions measuring animators' RFS perception. For each question, both frames were resized and cropped to maximize their visual similarity. A red arrow showed participants which arm to consider in each frame when answering the question (Figure 2, center). The experiment questions were presented on Qualtrics software. Each RFS question clearly described the participants' task: *Given two frames A and B, rate the helpfulness of Frame B as drawing reference for Frame A on a 1-7 scale slider, with 1 being "Not at all helpful" and 7 being "Extremely helpful"* (Figure 2, center). This rating was our measurement of animators' RFS perception between Frames A and B — specifically, of Frame B with respect to Frame A. Qualtrics was used to randomize the order in which questions appeared during the experiment.

Out of 30 professional hand-drawn animators employed by Cartoon Saloon at the time of the experiment, 10 volunteered to participate. Animators were recruited by internal email and were not paid to participate. Before starting the experiment, participants provided information on their professional history. Seven participants had 3-6 years of experience working for hand-drawn animation studios and the remaining three participants had 14-30+ years of experience; on average, participants had 9.6 years of industry experience. Seven participants had an additional 2-4 years of hand-drawn animation experience from attending animation college. Eight participants worked on *Wolfwalkers* and had experience drawing Mebh. Since requiring participants to draw all 54 Frame As during the experiment was impractical, participants were asked to imagine that they were drawing Frame A using Frame B as drawing reference.

To model real-life animation studio conditions, participants were given Mebh's model sheets and asked to reference them during the experiment, as they would while animating. Participants were asked to imagine that they could manipulate Frame B on their computers as they would while animating (e.g., via rotation, reflection, etc.) to make it better drawing reference for Frame A. Because this experiment aimed to measure animators' RFS perception with respect to their professional workflows, it was important for their in-practice software usage to impact experiment results. Participants were asked to complete the experiment on the computer monitor or drawing tablet they used while animating to mimic their working conditions (participants' reported screen widths = 15-32"). Because Mebh's hand pose varied too widely across frames to be experimentally controlled, participants were asked to ignore differences in hand pose when answering experiment questions. Our commit-

ment to using original hand-drawn frames precluded us from editing out hand poses. Participants completed two RFS training questions to familiarize themselves with the question format, and were required to confirm that they understood all experiment directions before proceeding to the RFS perception experiment questions. After completing all RFS questions, participants answered debriefing questions. They were asked about what factors determined their answers to RFS perception questions and how much they referenced Mebh's model sheets during the experiment.

4. Results

We analyzed animators' RFS scores via Three-Way ANOVA with repeated measures: Distortion, Reflection and Rotation. We conducted Mauchly's test for sphericity, and where the assumption was violated, we applied the Greenhouse-Geisser correction (indicated with an asterisk). We found a main effect of Distortion ($F^*(1.18, 10.64) = 74.40, p^* < 0.001, \eta_p^2 = 0.892$), where posthoc analysis showed that all 3 levels of Distortion were rated significantly differently from each other, low (5.139), medium (4.556) and high (2.778) ($p < 0.004$ in all cases). We also found a main effect of Reflection ($F^*(1.27, 11.49) = 11.773, p < 0.001, \eta_p^2 = 0.567$), where posthoc showed that low (5.556) was rated the same as medium (4.239), while high (3.678) was significantly different to both ($p < 0.03$ in both cases). We also found a main effect of Rotation ($F^*(1.79, 16.08) = 14.652, p \approx 0, \eta_p^2 = 0.619$), where posthoc showed that low (4.344) was rated the same as medium (4.389), while high (3.739) was significantly different to both ($p < 0.02$ in both cases). The higher effect size of Distortion compared to those of Reflection and Rotation supports our hypothesis that animators' RFS perception inversely correlates to Distortion more strongly than to Rotation and Reflection.

We also found a significant interaction between Distortion and Reflection ($F(4, 36) = 4.127, p < 0.007, \eta_p^2 = 0.314$), where pairwise comparisons showed us that, at low Distortion, there were 3 significantly different ratings for Reflection levels low, medium and high ($p < 0.04$ in all cases). For medium Distortion, low and medium Reflection were rated the same, while high Reflection was rated significantly worse than the other two ($p < 0.04$ in both cases). For high Distortion, there was no effect of Reflection level, indicating that the more Distortion present, the less important the differences in Reflection levels (Figure 3, left).

A significant interaction between Distortion and Rotation was

also found ($F^*(2.04, 18.38) = 3.999, p < 0.035, \eta_p^2 = 0.308$), where at low Distortion, low and medium Rotation were rated the same, while high Rotation was significantly worse than both ($p < 0.01$ in both cases). For medium Distortion, low Rotation was rated significantly higher than either medium or high Rotation ($p < 0.02$ in all cases), while at high Distortion, medium Rotation was rated significantly higher than low or high Rotation. This last result was most likely an artifact of particular selected frames (Figure 3, center).

Finally, a significant interaction occurred between Rotation and Reflection ($F(4, 36) = 4.088, p < 0.008, \eta_p^2 = 0.312$), where the pattern matched that of the Distortion and Rotation interaction (Figure 3, right). No three-way interaction occurred.

Qualitative Results Participants reported that their RFS ratings depended on foreshortening difference (6 participants), body pose difference (4), overall visual similarity (3), and how well the reference frame matched model sheet character design (2). One participant remarked that "even though an arm could look similar, a tilt in volume or rotation in space that is different might induce mistakes of interpretation of the design style." As region Distortion corresponds to foreshortening changes, this further indicates Distortion's predominant effect on RFS perception. 4 participants reported that referencing model sheets during the experiment helped them preserve character design — which naturally varies across animators and hand-drawn frames — and 3 reported that model sheets helped them draw basic poses. According to one participant, "if the foreshortening/exaggeration was too much for the B pose, I found it not very helpful [as drawing reference] and referred to the model sheets for clarity on the design on the arm." Animators with ≤ 6 years of industry experience gave more detailed answers about model sheet helpfulness, possibly indicating that they are still actively learning skills for extrapolating novel poses from drawing references.

5. Discussion

Our preliminary experiment results support our hypothesis that animators' RFS perception is sensitive to region shape/adjacency change (Distortion) and somewhat stable under 2D orthogonal transformation (Reflection and Rotation). Therefore region shape/adjacency similarity is a key factor in animators' RFS perception, superseding body pose similarity in several cases (e.g., Figure 2, right). Pairwise interactions indicate that influence of orthogonal transformation on RFS may diminish as Distortion intensifies, but further testing is needed to confirm. We expect this to be true, since Distortion alters visual information available in frames, and most participants reported that visual information preservation between frames influenced their RFS ratings.

Prior perceptual experiments on professional artists study their mental rotation abilities alongside general perceptual tasks like mentally combining objects [WWH*11] and Gestalt completion [Koz01], finding that such abilities improve over time as artists practice their drawing skills [CKDW19]. Rather than evaluating general visual perception, our experiment explores a perception task specific to its participants' vocation, supporting future development of a reference frame retrieval tool valuable in industry settings. Since past work shows that artists outperform laypeople on

general visual perception tasks, we reasoned that hand-drawn animators are likely uniquely expert in the perception tasks specific to their craft. While our preliminary results support those of previous studies, detailed inferences about professional hand-drawn animators' RFS perception could not be made from existing literature. We conducted our experiment to address this knowledge gap.

Our experiment has several limitations. The small number of examples per variable case likely affected pairwise interactions. Classifying variable cases by eye was appropriate for our preliminary study, but may have impacted results. Future work can use empirical shape comparison metrics for classification, but such metrics' differences to human visual perception may create new limitations. The small number of participants working at the same animation studio and use of a single character and body part prevent us from generalizing our findings to all hand-drawn animators, art styles, characters, and body parts. Expanding our existing experiment design to include more frames, participants, art styles, characters, and (combinations of) body parts will address most of these limitations. While it was not possible for our participants to draw target frames during the experiment, future work might explore this option. Our experiment was designed for body parts with left-right symmetry, and could be reformulated for non-symmetric character designs in future. Participants reported that reference frames' visual clarity, correctness of character designs, and information they provided about how to draw the target pose strongly influenced their RFS ratings, so these factors should be studied in future work. Future work can also compare animators' and laypeople's RFS perception, provide more reference frames per question to supplement available visual information, and evaluate RFS interactions between characters' depicted body pose and region shape/adjacency change. The last of these will be especially critical in developing an automatic reference frame retrieval method balancing pose estimation with region shape and adjacency awareness.

References

- [BB22] BRODT K., BESSMELTSEV M.: Sketch2pose: estimating a 3d character pose from a bitmap sketch. *ACM Trans. Graph.* 41, 4 (July 2022). doi:10.1145/3528223.3530106. 1
- [CKDW19] CHAMBERLAIN R., KOZBELT A., DRAKE J., WAGEMANS J.: Learning to see by learning to draw: A longitudinal analysis of the relationship between representational drawing training and visuospatial skill. *Psychology of Aesthetics Creativity and the Arts* (04 2019). doi:10.1037/aca0000243. 1, 4
- [JZJ*23] JU X., ZENG A., JIANAN W., QIANG X., LEI Z.: Human-art: A versatile human-centric dataset bridging natural and artificial scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023). doi:10.1109/CVPR52729.2023.00067. 1
- [Koz01] KOZBELT A.: Artists as experts in visual cognition. *Visual Cognition - VIS COGN* 8 (12 2001), 705–723. doi:10.1080/13506280042000090. 1, 4
- [PESP18] PASUPATHY A., EL-SHAMAYLEH Y., POPOVKINA D. V.: Visual shape and object perception, 01 2018. doi:10.1093/acrefore/9780190264086.013.75. 1
- [WWH*11] WALKER C., WINNER E., HETLAND L., SIMMONS III S., GOLDSMITH L.: Visual thinking: Art students have an advantage in geometric reasoning. *Creative Education* 2 (01 2011), 22–26. doi:10.4236/ce.2011.21004. 1, 4