Now You *Shake* Me: Towards Automatic 4D Cinema
SUPPLEMENTARY MATERIAL

Yuhao Zhou[1] Makarand Tapaswi[1,2] Sanja Fidler[1,2]
[1]University of Toronto [2]Vector Institute
{henryzhou, makarand, fidler}@cs.toronto.edu
http://www.cs.toronto.edu/~henryzhou/movie4d/

We present several details about our dataset, and few additional results. In addition to this document, please note that the supplementary material also includes: (i) several example video clips showing ground-truth effect annotations and predicted results; and (ii) a web page with GIF visualizations (effects are best understood by watching videos instead of static images) for all 9 effects spread across the three intensities. We link to and discuss them in the document below.

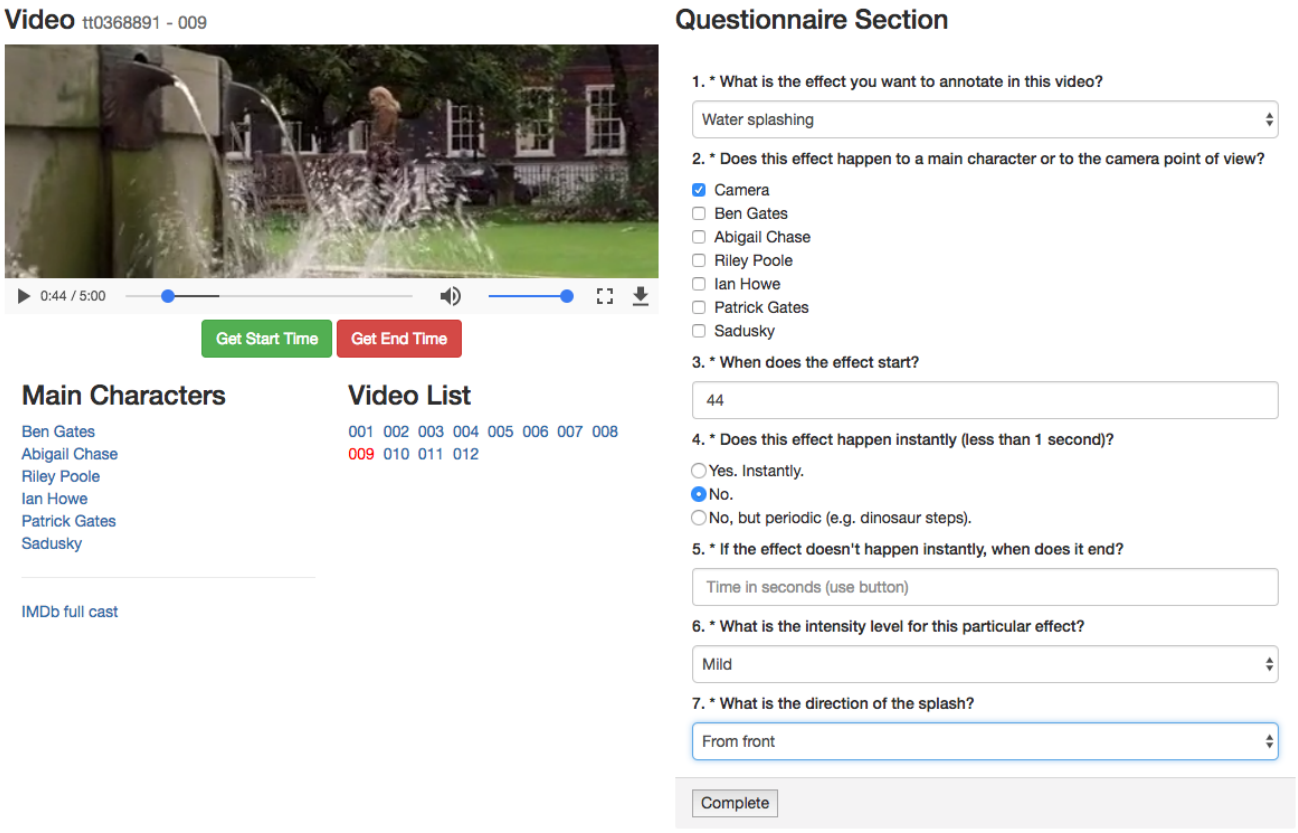

Figure 1: Web annotation interface. The annotators watch the full untrimmed video shown on the left. Upon selecting the effect type, the questionnaire asks to fill in corresponding details related to that effect. The annotators are required to answer all questions and submit the annotation before proceeding.

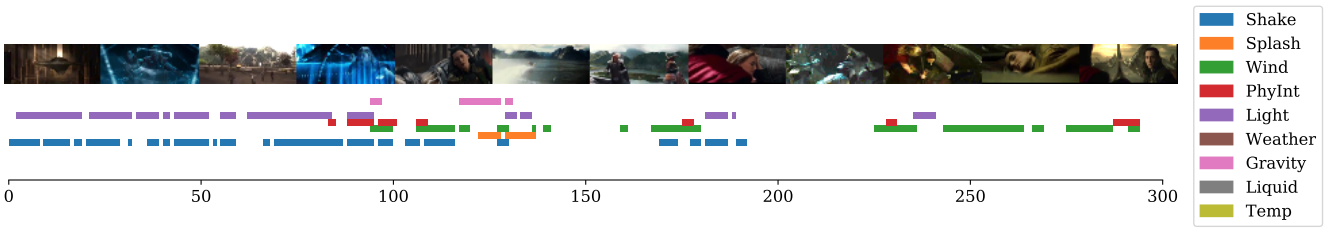

Figure 2: Example annotations made within a clip from the movie: *Thor: The Dark World*. We plot time on the x-axis and create a small rectangle for each effect type (based on color) corresponding to one second in the video. As can be seen, multiple effects do occur concurrently during intense scenes. On the other hand, possibly during dialog, effect annotations are sparse.

1. Annotation interface and effects

Fig. 1 provides a glimpse into our annotation interface. The annotators switch between video clips by clicking on the list of videos (below the video), can see the list main characters in the movie (bottom left), and answer a dynamic questionnaire that chooses questions based on previously chosen answers. Also note that the start and end-timestamps are grabbed from the video by pressing buttons thus minimizing human error in annotating the duration in seconds.

Fig. 2 presents the effect annotations for the clip with highest number of effects (count). Note how multiple effects can co-occur, while other parts of the video have no effects.

Fig. 3 presents all effect types in our movies along with *mild*, *medium*, and *strong* intensities. We explain each sample from the effects in the figure caption in detail. Additionally, as effects are seen best as videos, we create short 2 second GIF visualizations for the same examples. Please open `effect_intensity.html` in your browser to view them.

2. Dataset analysis

We extend Fig. 5 of the main paper (t-SNE [1] plot for 40 characters) in Fig. 4 showing 60 characters grouped by effect duration. A similar semantic grouping is observed, where sci-fi, action, and adventure characters cluster together.

Similarly, in Fig. 5, we show a t-SNE visualization of the movies themselves based on their effect density (fraction of time spent with each effect type).

Finally, we explore correlations between the number of effects in a movie and its "coolness". We evaluate this based on the movie ratings, gross revenue, and budget (all obtained from IMDb). Fig. 6 presents scatter plots for each of the above metrics vs. the amount of time for which a movie has effects (density). While movies with more effects typically require higher budgets, good storytelling does not seem to depend entirely on the density of effects (*e.g.* *Dark Knight*, *Pulp Fiction*). However, having more effects does not hurt in general.

3. Video results

Results for effect detection and classification are best seen in video form. We present several example video clips depicting the ground-truth and predicted effect annotations. For simplicity of video creation and to see the effects clearly, effect and prediction labels change only at each second. Due to the size of the videos, they are included on our website instead of supplementary materials.

4. Effect correlation

We end with confusion matrices for a study of effect correlation. Fig. 7 presents the correlation between a pair of sub-shots with a gap of n in between. We drop sub-shots pairs if either does not have any live effects. We see a strong diagonal indicating effects last for multiple sub-shots and/or are repeated within a period of time. Additionally natural effects (*Shake*, *Wind*, *Temperature*, *Light*) seem to be more correlated as compared to others like *Gravity*.

References

[1] L. J. P. van der Maaten and G. E. Hinton. Visualizing High-Dimensional Data using t-SNE. *JMLR*, 9(11):2579–2605, 2008. 2

Figure 3: Example frames for each effect at the three different intensity levels. Below we describe in one sentence the scenario for each example. **SHAKE** 1a. Characters are driving a car on a bumpy road. 1b. A character is sitting in a spacecraft traveling at extremely high speed. 1c. A missile hit the target, resulting in a strong explosion. **SPLASH** 2a. Dogs are playing in the water at the beach. 2b. Characters are fighting on a motorboat sailing in the sea. 2c. A pirate ship comes up from under the sea. **GRAVITY** 3a. A character jumps off from a high ground (zero-gravity for a short period of time). 3b. A character accelerates on his motorcycle at high speed. 3c. Multiple characters are floating at zero-G inside the space capsule. **LIGHT** 4a. A flashlight flaring towards the camera. 4b. A projector shines at the back of the room. 4c. A character travels through a portal to another world. **WIND** 5a. A gentle breeze touches character's face on a cliff. 5b. The sea wind blows away sailor's hat. 5c. A character rides a flying create hunting from the sky (wind in his face due to motion). **WEATHER** 6a. Light snow at a party. 6b. Rain in London. 6c. Soldiers are crossing a jungle in the rain. **TEMPERATURE** 7a. Characters gather around at a campfire. 7b. A character visits a village on a plateau in Tibet. 7c. A character was surrounded by burning woods. **LIQUID** 8a. Two characters are trapped on a ship which is about to sink. 8b. A character finds herself awoken in the sea. 8c. A character sinks into a lake. **PHYSICAL INTERACTION** 9a. Two characters are shaking hands. 9b. The middle character is being forced out of a carriage as a hostage. 9c. A character (on the right) is about to counter a gorilla attack.

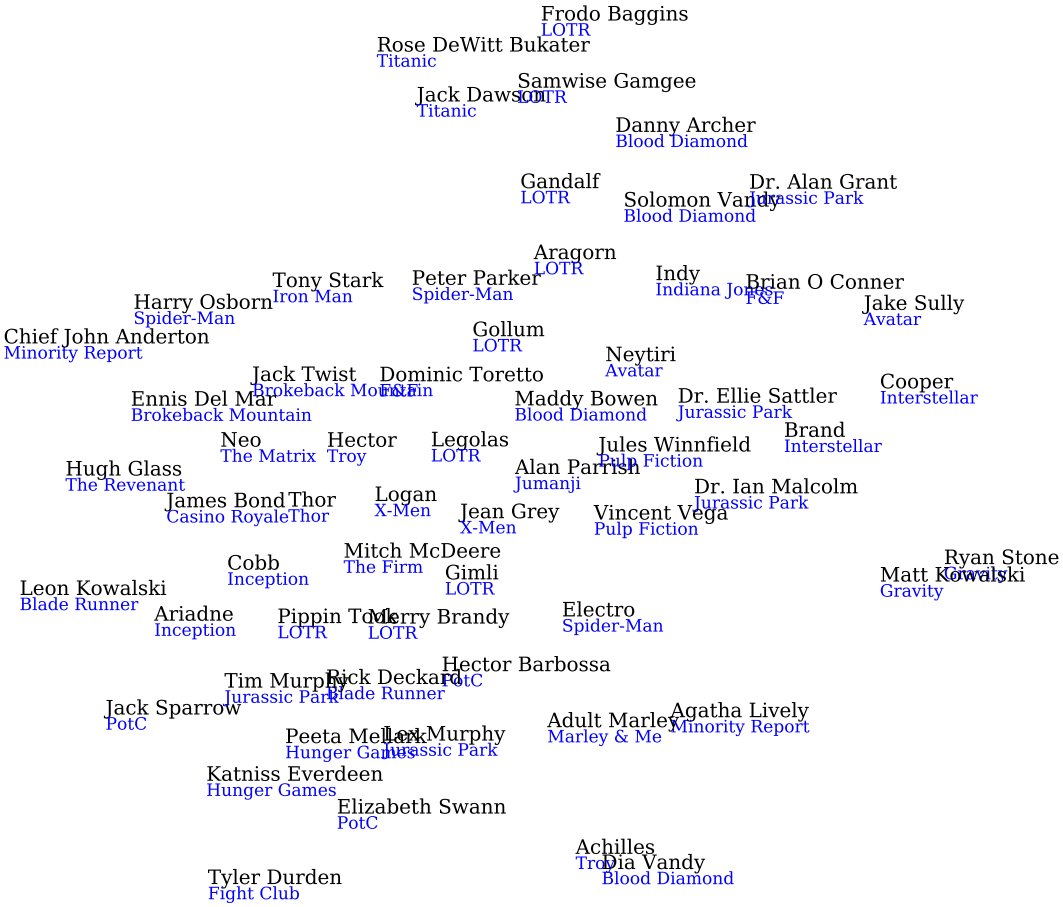

Figure 4: Character t-SNE based on effect duration. Using the top 60 characters and plot them. The same grouping behavior still exists: Sci-Fi movies cluster on the right, action movies cluster on the left, adventure movies cluster in the bottom.

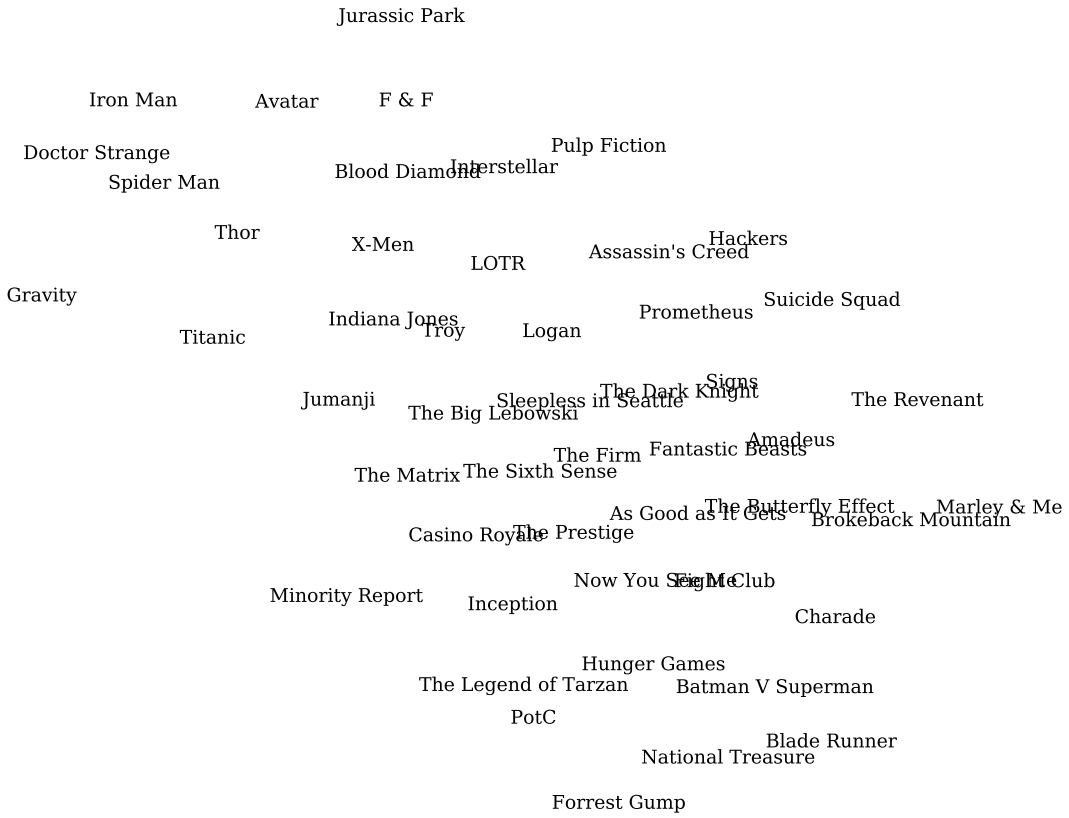

Figure 5: Movies are embedded using t-SNE. We computed the density of each effect for each movie and select the top 50 movies containing the most effects. The cluster shown in the t-SNE visualization reveals genre and even movie content. Sci-Fi movies form a cluster in the top-left of the figure. Drama and comedy are mostly seen on the right. And adventure/action movies are mainly clustered in the bottom.

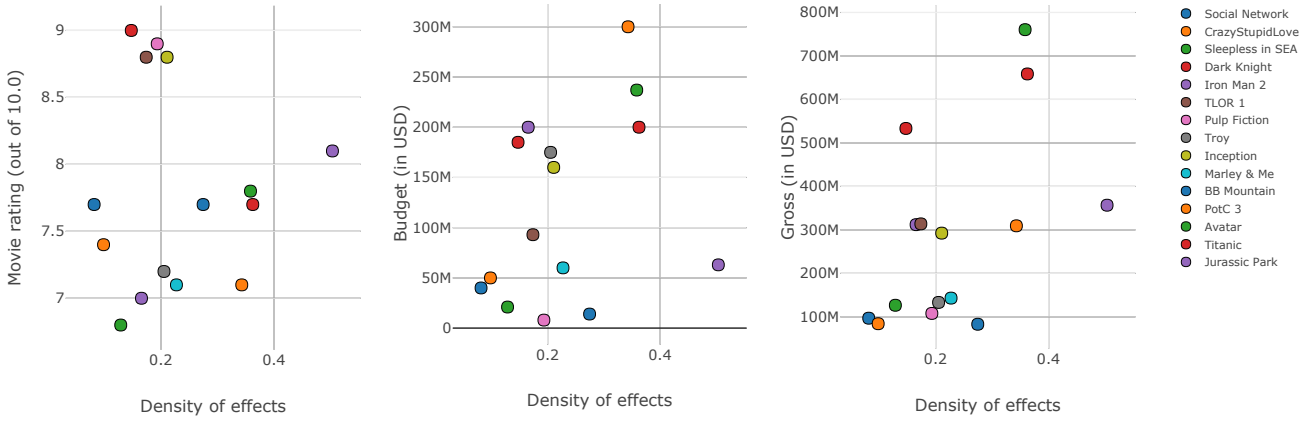

Figure 6: Movie performance as measured by rating, budget, and gross revenue vs. density of effects. More effects seem to require higher budges, but also have more revenue. However, the movie rating depends more on the storytelling rather than density of effects.

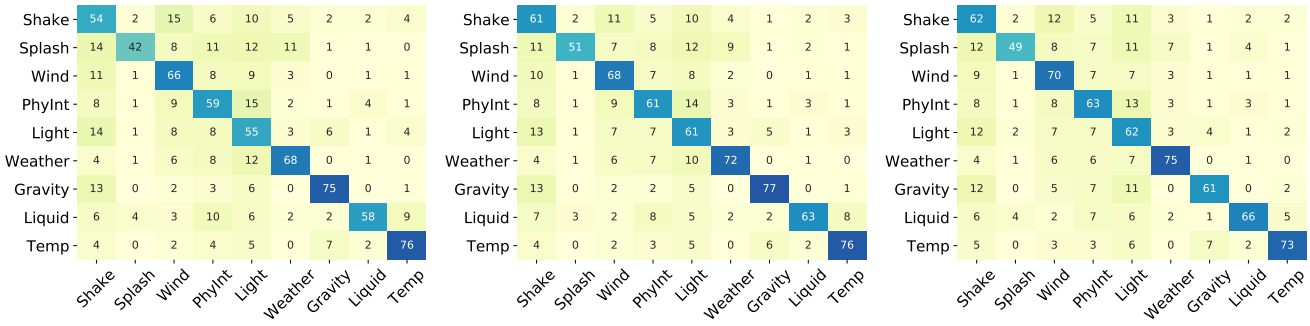

Figure 7: Effect correlation graph: we examine the causal effect of effects. For each sub-shot, we look ahead the next n subshots and see if the current effect has an impact on the next sub-shots. From left to right of the figure, n is 1, 5, 10 sub-shots ahead. We see that a strong diagonal for all plots indicating that effects of the same type tend to last for a long period of time, or occur again soon.