Continuous Real-time Annotation Fusion via Rank-based Signal Warping

Brandon M. Booth†, Karel Mundnich‡, Shrikanth S. Narayanan‡

†Department of Computer Science, ‡Department of Electrical Engineering

Motivation

Continuous human annotations are noisy and prone to unintended influence from personal bias, task ambiguity, environmental distractions, and more. **Can we remove these artifacts?**

Try this annotation challenge:

How goofy are these faces on a [0,1] scale?

Why is this hard? Goofiness does not have an intuitive scale. Now try to find which two faces are more similar.

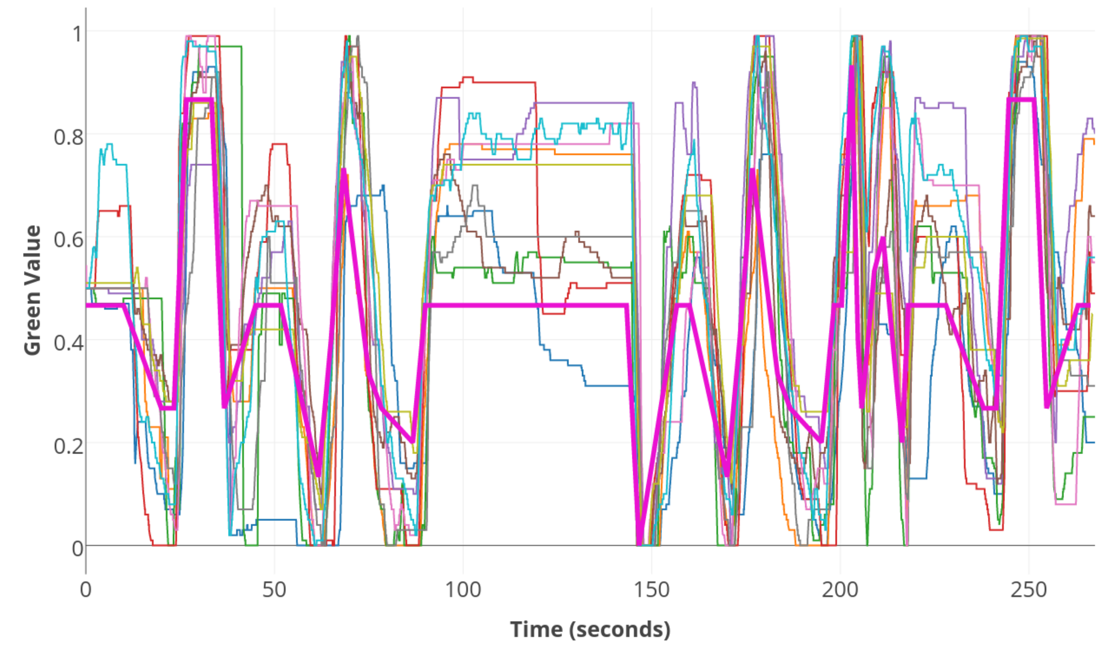
People are better at ranking than rating. [3]

Our Approach

- Propose **rank-based signal warping** to complement existing annotation fusion methods
- **Validate on experiment** with an objective truth

Experiment

Ten annotators rate the intensity of the color green in a video in real-time.



Annotations alongside the true value (**bold**)

Annotators **cannot capture trends while preserving self-consistency**.

Rank-based Warping

1. Apply any state-of-the-art annotation fusion method
2. Extract nearly constant intervals using total variation denoising [1]
3. Collect additional annotations comparing triplets of constant intervals
4. Construct ordinal embedding from constant intervals (using t-STE) [2]
5. Warp signal to align with embedding (Figure 1)

AGREEMENT MEASURES FOR BASELINE AND WARPED FUSED ANNOTATION APPROACHES

Signal Type	Pearson	Spearman	Kendal's Tau	NMI
Simple Average	0.775	0.795	0.636	0.302
Warped Average	**0.811**†	0.738	0.584	0.307
EvalDep Average	0.906	0.946	0.830	0.484
Warped EvalDep	**0.967**†	0.939	0.835	0.562

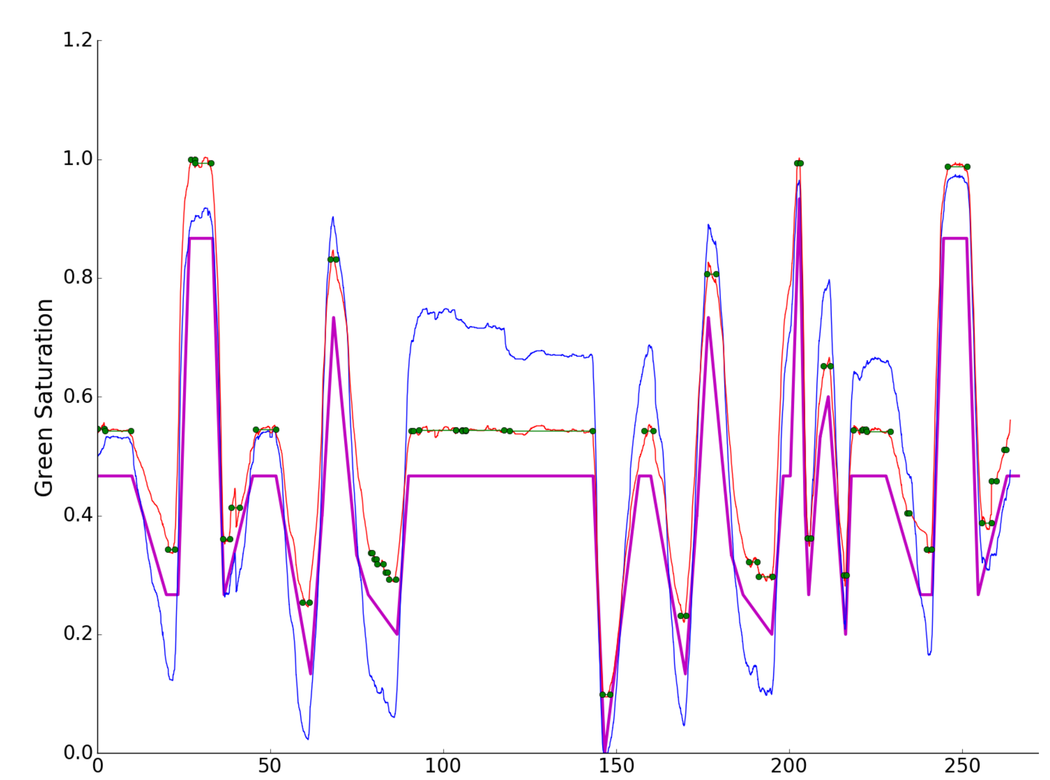
A † denotes significant improvement ($p < 0.005$, Fisher z-transform). Warped results use a complete set of ordinal comparisons. NMI = normalized mutual information.

$$\mathrm{S}_t = \begin{cases} \mathcal{E}_i - \frac{1}{|\mathcal{I}_i|}\sum_{s\in\mathcal{I}_i} y_s & \exists \mathcal{I}_i \in \mathcal{I} : t \subseteq \mathcal{I}_i \\ 0 & o.w. \end{cases}$$

$$y'_t = \begin{cases} y_t + S_t & \exists \mathcal{I}_i \in \mathcal{I} : t \subseteq \mathcal{I}_i \\ \left(\frac{y_t - y_a}{y_b - y_a}\right)(y_b + \mathcal{S}_{b^+}) + \left(\frac{y_b - y_t}{y_b - y_a}\right)(y_a + y'_{a^-}) & \exists [a,b] = \mathcal{J}_j \in \mathcal{J} : t \subseteq \mathcal{J}_j \end{cases}$$

Fig. 1: Let $\mathcal{I}$, $\mathcal{J}$ and $\mathcal{E}$ be the ordered sets of intervals, inter-intervals and embedding values. Let $t \in \{1, 2, ..., T\}$, y_t denote the fused annotation, and y'_t denote the warped signal value. Subscript + or − denote a time just before or after the time index.

Results



Objective truth (magenta), fused annotations (blue), warped signal (red), embedded intervals (green)

Conclusion

We leverage the **natural ability of human annotators to annotate trends** in real-time and

We separately leverage accurate similarity comparisons to achieve accurate ground truth.

References

[1] Stephen R Becker, Emmanuel J Candès, and Michael C Grant. "Templates for convex cone problems with applications to sparse signal recovery". In: *Mathematical programming computation* 3.3 (2011), p. 165.

[2] Laurens Van Der Maaten and Kilian Weinberger. "Stochastic triplet embedding". In: *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on.* IEEE. 2012, pp. 1–6.

[3] Georgios N Yannakakis and Héctor P Martínez. "Ratings are overrated!" In: *Frontiers in ICT* 2 (2015), p. 13.