A Novel Method for Human Bias Correction of Continuous-time Annotations

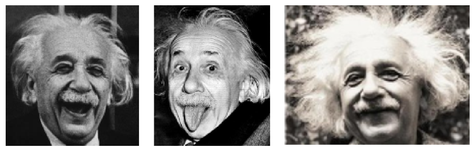
Brandon M. Booth, Karel Mundnich, Shrikanth S. Narayanan

Signal Analysis and Interpretation Lab @ University of Southern California

Motivation

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

Try this annotation challenge:



How silly are these facial expressions on a [0,1] scale?

Why is this hard? Silliness does not have an intuitive scale. Now instead try this: compare the first two images and pick the one with a sillier facial expression.

People are better at ranking than rating[1]

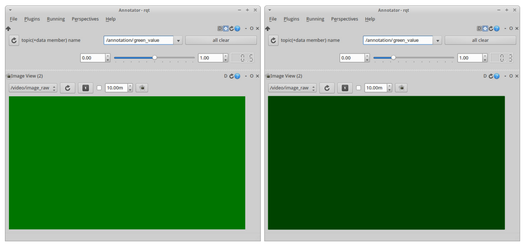
Goal

Can we leverage the improved accuracy of human-based ranking to refine continuous real-time human annotation?

- We propose **rank-based signal warping** to complement existing annotation fusion methods
- We **validate our method** in an experiment **with a known truth**

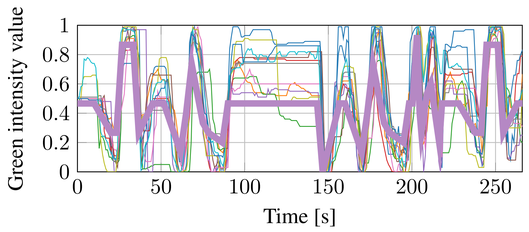
Experiment

Ten annotators rate the intensity of the color green in a video in real-time on a continuous [0,1] scale.

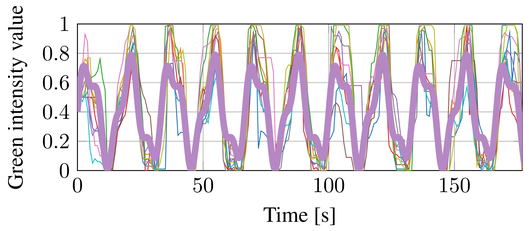


Frame 1 Frame 85

Experiment Results



Task A: Annotations alongside the true value (**bold**)



Task B: Annotations alongside the true value (**bold**)

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

Rank-based Warping

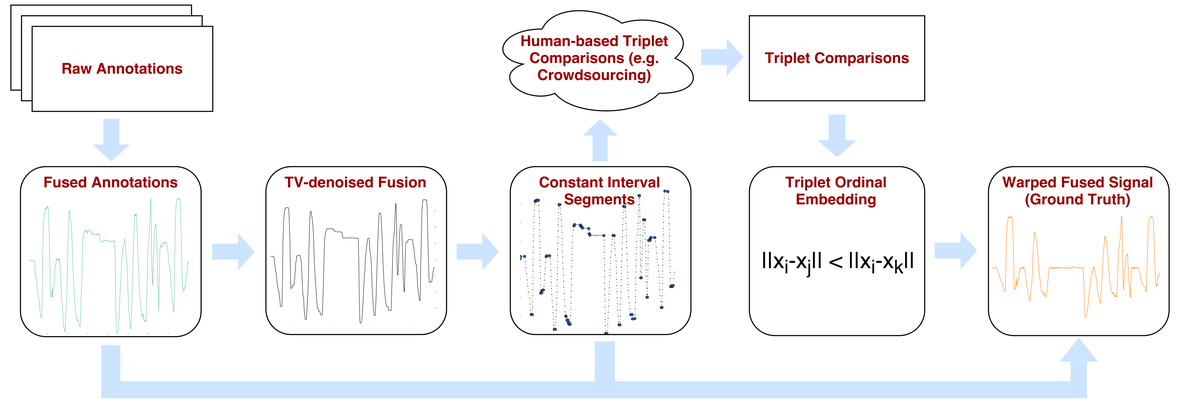
1. Apply any state-of-the-art annotation fusion method
2. Extract nearly constant intervals from fused signal using total variation denoising [2]
3. Collect additional annotations comparing triplets of constant intervals
4. Construct ordinal embedding from constant intervals (using t-STE) [3]
5. Warp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

Results

Agreement measures for baseline (EvalDep [4]) and warped fused annotation approaches

Task	Signal	Pearson	Spearman	Kendall's Tau	NMI
A	Baseline	0.906	0.946	0.830	0.484
	Warped	0.967	0.939	0.835	0.562
B	Baseline	0.969	0.969	0.855	0.774
	Warped	0.988	0.987	0.906	0.862

Our Approach



Our warping method:

$$\mathcal{I}_i = \begin{cases} \{t : \min(\mathcal{C}_i) \le t \le \max(\mathcal{C}_i)\} & i \in \{1,2,...,|C|\} \\ \{0\} & i = 0 \\ \{T\} & i = |C|+1 \end{cases} \qquad S_i = \begin{cases} \mathcal{E}_i - \frac{1}{|\mathcal{I}_i|}\sum_{t\in\mathcal{I}_i} y_t & i \in \{1,2,...,|C|\} \\ 0 & \text{o.w.} \end{cases}$$

$$y'_t = \begin{cases} y_t + S_i & \exists \mathcal{I}_i : t \in \mathcal{I}_i \\ y_t + \left(\frac{y_t - y_a}{y_b - y_a}\right)\mathcal{S}_{i+1} + \left(\frac{y_b - y_t}{y_b - y_a}\right)\mathcal{S}_i & \exists i : a \le t \le b \end{cases} \quad \text{where} \quad a = \max(\mathcal{I}_i),\ b = \min(\mathcal{I}_{i+1})$$

We let $t \in \{1,2,...,T\}$ be a time index, y_t denote the fused annotation signal, y'_t denote the warped signal value, and let $\mathcal{C}$ be the ordered sequence of non-overlapping time intervals corresponding to the extracted constant intervals. We define $\mathcal{E}$ as the sequence of embedding values in $\mathbb{R}^d$ corresponding to the time interval sequence $\mathcal{C}$. The sequence $\mathcal{I}$ is used instead of $\mathcal{C}$ to handle edge cases. For notational simplicity, we also introduce a new sequence S whose i^{th} element is the difference between interval i's average value and the corresponding embedding value.

Comparison Plot

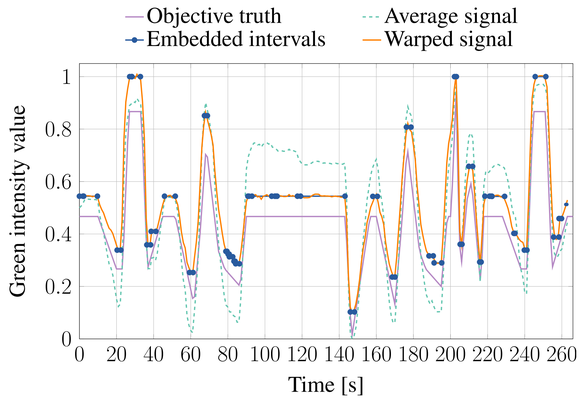


Fig. 1: The spatially warped signal **better approximates the structure** of the objective truth and also achieves **greater self-consistency** over the entire annotation duration.

Conclusion

- We leverage the **natural ability of human annotators to annotate trends** in real-time
- We **separately leverage accurate similarity comparisons** to achieve accurate ground truth

References

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

[2] 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.

[3] 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.

[4] Soroosh Mariooryad and Carlos Busso. "Correcting Time-Continuous Emotional Labels by Modeling the Reaction Lag of Evaluators". In: *IEEE Transactions on Affective Computing* 6.2 (2015), pp. 97–108.

Thanks to our sponsors for supporting this work:

NSF NIH National Institutes of Health