Trapezoidal Segmented Regression: A Novel Continuous-scale Real-time Annotation Approximation Algorithm

Brandon M. Booth
Signal Analysis and Interpretation Lab (SAIL)
University of Southern California
brandon.m.booth@gmail.com

Shrikanth S. Narayanan
Signal Analysis and Interpretation Lab (SAIL)
University of Southern California
shri@sipi.usc.edu

***Abstract*—Accurate ground truth representations of human behavior and experiences are essential for furthering our understanding of the complex relationships between everyday events and interactions and their effects on people. Producing accurate ground truth signals for subjective or latent experiences is difficult because it requires human annotation and is subject to annotator bias, distraction artifacts, valuation errors, among others. We build on previous work aiming to produce highly accurate continuous-scale ground truth labels for human experiences which advocates using supplemental human observations to warp the continuous-scale annotations to correct these errors. We propose a new method, *trapezoidal segmented regression*, for optimally approximating fused human-produced continuous-scale annotations to simplify its segmentation into intervals of low and high confidence in valuation. We evaluate this algorithm as an alternative to the *total variation denoising* method used in prior work by comparing the ground truths that both methods produce in experiments where the true annotation target signal is known *a priori*. Results show that the proposed signal approximation technique performs on par with the prior method, producing ground truth signals in close alignment with the true target, but with the added advantages of being more easily tuned and intuitive. We conclude that the proposed algorithm enables accurate and more robust ground truth generation.**

***Index Terms*—Annotation fusion, continuous annotation, segmented regression, trapezoidal signals**

I. Introduction

Human experience and behavior modeling efforts that involve data, whether they employ machine learning techniques or statistics and correlation analysis, often require a set of human-produced annotations or labels. Modeling affective states, for example, typically involves the collection of individual-relevant environmental and physiologic data and associating it with dimensional or categorical labels of affect with the aid of human annotators. The accuracy of human behavior inference models trained using these labels is impacted not only by data noise but also the quality of the annotations. In this paper, we build on prior work [1] showing that additional human feedback can be used to modify human-based annotations to reduce biases and artifacts and produce a set of labels well-aligned with the underlying construct of interest.

Though many annotation schemes are possible and have been used in existing human behavior data sets, such as continuous-scale labels for discrete data [2], discrete labels for sampled real-time data [3, 4, 5], and continuous-scale labels for real-time data [6, 7], in this work we focus on the continuous-scale and real-time annotation strategy. We use the term *real-time* to refer to setups where annotators provide ratings simultaneously while viewing the stimulus. *Continuous-scale* refers to the continuous range of values annotators can provide at any given time as opposed to discrete Likert-scale annotations for instance. A typical example of a continuous-scale and real-time annotation scheme is using a user interface slider widget ranging from zero to one to annotate the emotional valence of a movie as it is being watched. We are interested in this annotation scheme primarily for three reasons: (1) it is the most natural way to present real-time data to annotators and avoids any aliasing of the data via discretization, (2) it allows for subtle variations to be annotated, and (3) it enables annotators to fully absorb and incorporate the temporal context of the stimulus into the annotation value.

In order to generate a single set of labels for use as ground truth in behavior modeling, several annotations from different annotators are typically gathered and aggregated to help reduce the impact of unique biases that individuals may introduce. Many methods for the fusion of continuous-scale annotations have been proposed. Correlated spaces regression [8] and canonical correlation analysis [9] (and variants [10, 11]) attempt to produce a ground truth via fusion of the annotations by minimizing the distance between some linear combination of annotations and a combination of features extracted from the associated data stream (e.g. facial expression features). These approaches can improve the ability of learning models to map features to ground truth labels, but do so by presuming some combinations of features and also some combination of annotations are already capable of representing the underlying target signal. When common biases appear in the annotations or when the feature set is not rich enough to capture the relevant details, then this may not be the case. Other approaches such as time-shift and average [12] and dynamic time warping [13] attempt to correct varying perception-response temporal lags introduced by each annotator before combining the annotations, but these methods by themselves do not offer any means to correct annotator perception or valuation biases and artifacts. Some efforts combine the benefits of both approaches [14, 15] but still presume the available features contain information about the construct being annotated. We contend approaches of this sort are unlikely to lead to accurate human behavior models because neither the correct features nor the true labels are known in advance. Furthermore, in situations where fatigue or external distractors interfere with annotators, artifacts can be introduced that may not be representable with the feature set which can diminish the accuracy of the resulting ground truth. In our view, it is therefore preferable for any method producing ground truth labels to do so independently of the collected data.

The authors of [1] present a technique for reducing the effect of annotation artifacts and biases by warping the continuous-scale signal produced using any of the aforementioned annotation fusion algorithms. This method segments the fused signal into regions of low and high confidence in annotator valuation, then uses supplemental information collected from humans to warp the signal and generate a ground truth that is more consistent in construct rating over time. That work shows that the enhanced consistency also produces ground truth labels in better alignment with the underlying construct in a few experiments where the true value is known *a priori*.

In that work, *total variation denoising* (TVD) is employed to approximate the fused signal and enable segmentation into low and high valuation confidence regions. TVD suffers both by being difficult to tune and through its inability to approximate regions around all extrema for certain signal inputs. In this paper, we present an alternative method, *trapezoidal segmented regression* (TSR), which is tuned intuitively and always successfully estimates regions associated with extrema near the most pronounced annotated value changes. A brief summary of the prior work's post-fusion signal correction technique is provided in Section II-A followed by a description and analysis of this framework using our proposed *trapezoidal segmented regression* algorithm.

Our unique contributions in this work include an algorithm and publicly available code for calculating the optimum trapezoidal segmented regression of a one-dimensional time series. We also produce results showing that TSR approximates fused annotations as least as well as the prior TVD method in simulated experiments. Finally we provide analysis illustrating why TSR is easier to tune than TVD and therefore could more easily be utilized to produce an accurate ground truth when used as part of the pipeline in [1].

II. Overview of Prior Work

To provide some context for the TSR algorithm, we briefly summarize the rank-based signal warping method proposed in [1] used to produce more accurate ground truth representations from raw annotations. The role of the TVD algorithm used in this framework is described followed by an examination of its drawbacks. These pitfalls are remedied by our proposed TSR algorithm presented in Section III.

A. Rank-based Signal Warping

Fig. 1 gives an overview of the steps in the signal warping algorithm. The pipeline is provided with a set of raw continuous-scale annotations and outputs a single continuous-scale signal suitable for use as ground truth. The method achieves this by correcting the valuation errors introduced in the raw annotations using ordinal embedding and supplemental triplet comparisons gathered from other annotators or crowd sourcing. Per prior observations in [1], only the nearly constant regions of the annotations need to be reevaluated because annotators tend to capture changes well. A signal approximation step is used to estimate an unrefined version of the fused annotations to facilitate the identification of these nearly constant regions. The original framework uses TVD in this step to approximate the signal as a piecewise constant step function, which makes the identification of nearly constant regions easier. Readers are referred to [1] for more details on the signal warping framework; in this paper we focus on improving the signal approximation and constant interval segmentation steps.

B. Total Variation Denoising and Pitfalls

In [1], TVD is used to obtain a piecewise-constant signal approximation $\hat{w}$ of an input signal v by optimizing the following:

$$\hat{w} = \arg\min_{w} \Big[\sum_t \|v_t - w_t\|_{\ell_2}^2 + \lambda \sum_t \|w_{t+1} - w_t\|_{\ell_1} \Big]$$

The λ coefficient is tunable and weighs the relative importance of the total variation term (ℓ_1-norm), which makes the solution more piecewise constant, against the signal error term (ℓ_2-norm). Given a particular signal v, λ needs to be adjusted so the optimization produces a signal approximation $\hat{w}$ similar to v but with nearly constant regions at the extrema. This process can be tedious requiring either many heuristics or several iterations of hand tuning with a human examining the result each time. Some smooth functions with large peaks and shallow bump features are highly sensitive to perturbations of λ and more difficult to tune. Furthermore, λ needs to be configured separately for every signal, even for those with similar structure but varied scales. In practice it is common for TVD to produce piecewise signals where the segments have at least a slight curvature as illustrated in Fig. 2, which hampers the constant interval segmentation stage in the signal warping method. For these reasons, using TVD to approximate the fused annotation complicates the signal warping procedure and hinders its ability to provide an accurate ground truth.

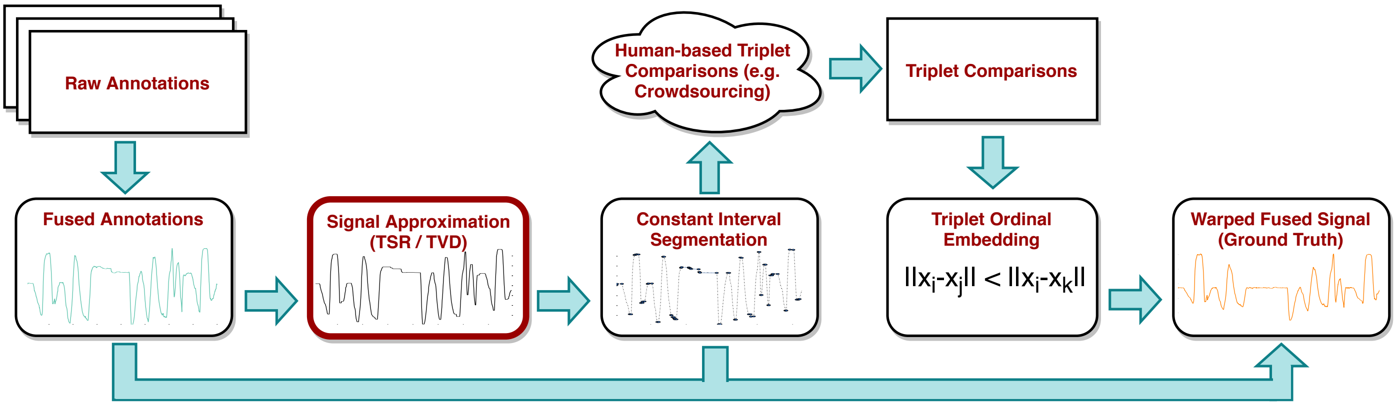


Fig. 1. Overview of signal warping algorithm first proposed in [1]. In this work, we explore using trapezoidal segmented regression in the signal approximation stage (highlighted) instead of total variation denoising.

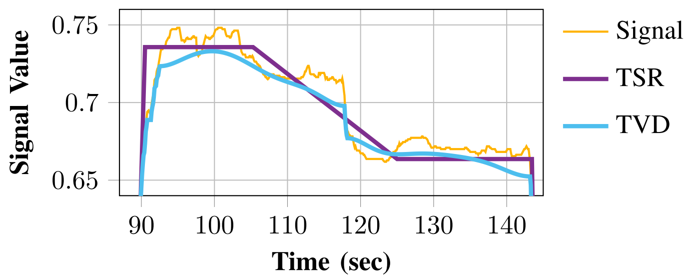


Fig. 2. Illustration of the curvature produced by TVD when approximating nearly constant regions of a signal. The proposed TSR algorithm approximates these regions with constant segments making the extraction of near-constant intervals easier.

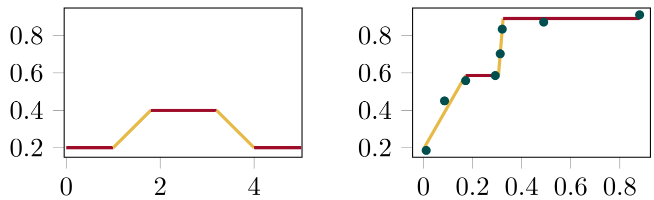


Fig. 3. Two trapezoidal signal examples. The left shows a prototypical trapezoidal signal shape. The right contains the optimum four-segment trapezoidal signal fit to the sample points. We use a broader definition of a trapezoidal signal: continuous alternating sloped and constant line segments. Each line type is colored differently for clarity.

III. Method

We propose a new algorithm called *trapezoidal segmented regression* that can be used for signal approximation to better enable constant interval segmentation compared to TVD as used originally in [1]. We begin by providing some background on segmented regression and formal definitions.

A. Segmented Regression Background and Definitions

In general, segmented regression (SR) involves the approximation of a sampled signal by means of fitting (typically) simpler functions over T segments of the domain. When the segment break points are known *a priori* this problem reduces to T local best-fit optimization problems. SR becomes harder when the segments are not known in advance and the locations of the break points, also known as *knots*, need to be calculated as well. Generalized SR for points in a variable d-dimensional space with a fixed number of T segments ($T > 1$) is a NP-complete problem as shown in [16]. For one-dimensional signals, however, the problem is solvable in polynomial time. Variants of SR limit the domain for knot point selection to the set of sampled points to reduce computational complexity. In this work, we are interested in finding the optimum segmented regression for a fixed T, and we allow segment boundaries to occur at arbitrary real-valued locations in time. We are further interested in continuous signal approximations so we only consider SR solutions where adjacent trapezoidal line segments intersect at their shared boundaries.

Trapezoidal functions are fundamental and widely used in signal processing and electrical engineering [17]. Fig. 3 shows a prototypical trapezoidal function on the left containing two sloped line segments and a constant line segment connecting them. In this work, we use a slightly broader definition of trapezoidal signals: piecewise linear signals with every other line segment having a slope of exactly zero. On the right, Fig. 3 also displays an example optimum four-segment trapezoidal signal approximation of the sample points using this relaxed definition.

To the best of our knowledge, trapezoidal signal regression has not been studied in the annotation modeling literature. Piecewise linear signal approximations have been studied a great deal, however, and form the basis of our approach in this paper.

B. Trapezoidal Segmented Regression

We formulate our optimum TSR algorithm leveraging the work of [18] which provides algorithms for a variety of piecewise-linear SR problems. The optimum TSR algorithm solves the following constrained minimization problem:

$$\underset{a_{1:T}, b_{1:T}, k_{1:T+1}}{\arg\min} \min_{c \in \{0,1\}} \sum_{t=1}^{T} \sum_{k_t \leq i \leq k_{t+1}} (\boldsymbol{a}_{c,t} x_i + b_t - y_i)^2 \quad \text{s.t.:}$$

$$\min\{x\} = k_1 \leq k_2 \leq ... \leq k_T \leq k_{T+1} = \max\{x\}$$

$$\forall t \in \{1, ..., T\}: \quad \boldsymbol{a}_{c,t} k_{t+1} + b_t = \boldsymbol{a}_{c,t+1} k_{t+1} + b_{t+1}$$

$$\boldsymbol{a}_{c,t} = \begin{cases} a_t & c = t \bmod 2 \\ 0 & \text{o.w.} \end{cases}$$

The x and y parameters are vectors whose elements contain the sample point data (x_i, y_i) for $i \in \{1, ..., n\}$. In our annotation fusion application, the x values are sampled times and the y values are sampled ratings. T is the desired number of segments and $c \in \{0, 1\}$ denotes whether the first segment of the regression should be constant (zero slope) or not (non-zero slope). The parameters a and b are coefficients of the line segments between knot points and each $k_{1:T+1}$ are the real-valued knot boundaries along the domain.

Algorithm 1 contains the pseudo-code for optimum continuous TSR borrowing some notation from [18]. A proof of correctness for this algorithm is outside the scope of this work. The optimum solution is produced by calling TRAPEZOIDALSEGREG twice, once with $c = 0$ and once with $c = 1$, then keeping the result with the smallest total cost ($F_{n,T}$).

The algorithm proceeds as follows assuming the sample points $\{(x_1, y_1), ..., (x_n, y_n)\}$ are ordered according to the x values:

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$$

Start with the point (x_1, y_1) and iteratively add the next point x_j (on the first iteration $j = 2$) maintaining a set of linear function coefficients $A_{j,t}$ and $B_{j,t}$ of the right-most line segment that minimizes the total squared error cost ($F_{j,t}$) of fitting $t \in \{1, 2, ..., T\}$ line segments over all points $\{(x_i, y_i)\}_{i=1}^{j}$. Matrix $K_{j,t}$ holds the exact real-valued knot point in the domain (x values) and matrix $I_{j,t}$ holds the indices of the sample points to the left of each knot (used for book-keeping). Once the last point has been considered, reconstruct the signal approximation for the minimum cost set of T line segments and return it. This *dynamic programming* approach reduces the overall run-time complexity.

The FITLINE function returns the coefficients of a line ($y = ax + b$) that minimizes the sum squared error of the x and y sample points passed in. FITCONST returns the coefficient b that minimizes the sum squared error over a constant line ($y = b$). FITLINEX and FITCONSTX return the same linear coefficients that minimize the sum squared error subject to the constraint that the new line/constant segment fitted to points $(x_{i:j}, y_{i:j})$ intersects the right-most line segment from the set of best-fit coefficients for points $(x_{1:i-1}, y_{1:i-1})$. In order to enforce the continuity of the piecewise trapezoidal function, the new line must intersect $y = A_{i,t-1}x + B_{i,t-1}$ somewhere between x_{i-1} and x_{i+1} (for $2 \leq i \leq n - 1$). Both of these methods can be expressed as convex quadratic programs (QPs) and solved optimally in polynomial time.

Algorithm 1 Optimum trapezoidal segmented regression

```
procedure TRAPEZOIDALSEGREG(x, y, n, T, c)
  for j = 1 to n do            ▷ Iterate over all points in sorted order
    for t = j to T do          ▷ Initialize costs, invalidate knot indices
      F_{j,t} ← ∞
      I_{j,t} ← 0
      K_{j,t} ← 0
    if c ≠ 0 then              ▷ Fit one line segment over all points so far
      a ← 0
      b, err ← FITCONST(x_{1:j}, y_{1:j})
    else
      a, b, err ← FITLINE(x_{1:j}, y_{1:j})
    A_{j,1} ← a
    B_{j,1} ← b
    F_{j,1} ← err
    I_{j,1} ← 1
    K_{j,1} ← 0
    for t = 2 to min{j − 1, T} do   ▷ Consider fitting between 2 and T segments to all
                                      points between the first point and current one
      F_{j,t} ← ∞
      I_{j,t} ← 0
      K_{j,1} ← 0
      for i = t to j − 1 do    ▷ For t target segments, find the best break point reusing the
                                  optimum fit for t − 1 segments over all points up to point i
        h ← I_{i,t−1}
        if k ≠ 0 and A_{h,i} ≠ A_{i,j} then
          u ← A_{i,t−1}
          v ← B_{i,t−1}
          x_min ← max{x_1, x_i}
          x_max ← min{x_{i+1}, x_n}
          if A_{i,t−1} ≠ 0 then
            a, b, x_int, err ←
              FITCONSTX(x_{i:j}, y_{i:j}, u, v, x_min, x_max)
          else
            a, b, x_int, err ←
              FITLINEX(x_{i:j}, y_{i:j}, u, v, x_min, x_max)
          if F_{j,t} > F_{i,t−1} + err then   ▷ Update the cost, linear coefficients, and break points
            F_{j,t} ← F_{i,t−1} + err
            A_{j,t} ← a
            B_{j,t} ← b
            I_{j,t} ← i
            K_{j,t} ← x_int
  z_T ← n                      ▷ Recover the optimum TSR
  x̂_T ← x_n
  for t = T to 2 do
    z_{t−1} ← I_{z_t,t}
    x̂_{t−1} ← K_{z_t,t}
  ŷ_1 ← A_{z_2,1} x̂_1 + B_{z_2,1}
  for t = 2 to T do
    ŷ_t ← A_{z_t,t−1} x̂_t + B_{z_t,t−1}
  return x̂, ŷ, F_{n,T}
```

FITCONSTX should return the solution to the following convex QP:

$$\begin{aligned} &\min_{b_{i,j}} \sum_{k=i}^{j} (b_{i,j} - y_k)^2 \quad \text{s.t.:} \\ &b_{i,j} \leq B_{i,t-1} + x_{\max\{1,i-1\}} A_{i,t-1} \\ &b_{i,j} \geq B_{i,t-1} + x_{\min\{i+1,n\}} A_{i,t-1} \end{aligned} \tag{1}$$

FITLINEX should return the solution to this convex QP:

$$\begin{aligned} &\min_{a_{i,j}, b_{i,j}} \sum_{k=i}^{j} (a_{i,j} x_k + b_{i,j} - y_k)^2 \quad \text{s.t.:} \\ &b_{i,j} + x_{\max\{1,i-1\}} a_{i,j} \leq B_{i,t-1} + x_{\max\{1,i-1\}} A_{i,t-1} \\ &b_{i,j} + x_{\min\{i+1,n\}} a_{i,j} \geq B_{i,t-1} + x_{\min\{i+1,n\}} A_{i,t-1} \\ &a_{i,j} > A_{i,t-1} \end{aligned} \tag{2}$$

In either of the two minimization formulations, it is possible the inequality constraints needs be flipped to achieve a smaller minimum (e.g. $a_{i,j} < A_{i,t-1}$ for the FITLINEX procedure) which still satisfies the intersection point restriction. For each optimization problem both constraint cases need to be handled, thus two QPs must be solved for each FITCONSTX and FITLINEX call corresponding to the normal and flipped constraints and the optimum taken as the minimum.

This algorithm has the same run-time complexity as its linear segmented regression counterpart from [18]; it requires $\mathcal{O}(n^2 T g)$ operations assuming $T < n$ and where g is an upper bound on the steps required for the convex QP to converge (for reference, TVD has a practical runtime of $\mathcal{O}(n)$ [19]). The TSR algorithm also requires $\mathcal{O}(n^2)$ memory. A Python implementation of TSR is publicly available on GitHub [20].

C. *Benefits*

Once the trapezoidal signal approximation is computed, the constant interval segmentation step in the next stage of the signal warping method becomes simpler: the constant interval set is taken directly from the constant line segments. More importantly, TSR has one parameter T required for tuning which should be roughly proportional to the number of extrema in the fused annotation signal. Given that T expresses the desired number of segments, the resulting signal approximation is unaffected by the scale of the fused signal. Humans can easily inspect the fused annotation to count the number of extrema and provide a good initial guess for the value T, unlike TVD which requires several human-in-the-loop iterations to obtain a reasonable initial guess for λ. Though this approach does not completely eliminate human oversight from the pipeline, it succeeds in removing much of the ambiguity.

Additionally, the choice of T in TSR is less sensitive to under- and overestimation than λ in TVD. Underestimating T causes the resulting trapezoidal approximation to fail to capture only the most subtle extrema but still provides a good approximation of the overall structure of the signal. Overestimating T creates more constant intervals than necessary but still achieves an accurate signal approximation. More details on this observation are presented in the next section.

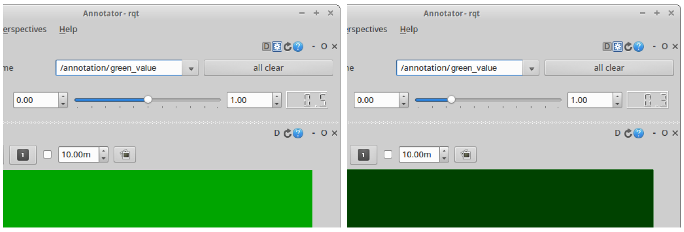
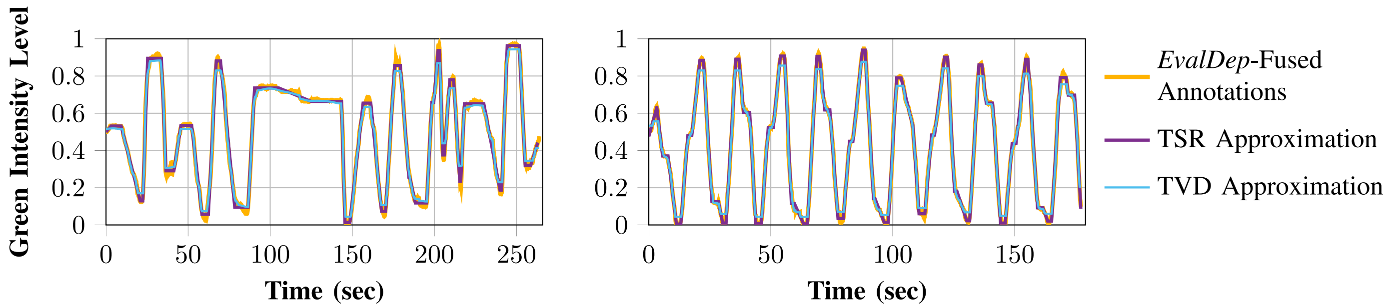


Fig. 4. Two snapshots of the annotation interface at different times used for the green intensity annotation experiments. Annotators used a slider widget to rate the color intensity on a continuous scale and in real-time as the green video changed over time.

IV. EXPERIMENTS AND RESULTS

We evaluate TSR as an alternative for TVD in the rank-based signal warping method for ground truth creation by comparing the two in an annotation fusion scenario where the true underlying signal is known *a priori*. In [1], the authors employ their signal warping method in a continuous real-time human annotation task where ten annotators are asked to rate the intensity of a solid shade of green that appears in a video. We reuse this data set for validating TSR. Fig. 4 shows snapshots of the annotation interface and the green color at two different times during playback. Each annotator independently produces separate annotations then a relatively simple annotation fusion method from [12] called *EvalDep* is used to time align the annotations and fuse them via averaging. This process is performed for two distinct annotation tasks, TaskA and TaskB, and the resulting fused annotations are shown in Fig. 5. TaskA exhibits random changes in intensity at a rate manageable by annotators while TaskB oscillates consistently over time. Both tasks test the capabilities of the annotators to produce accurate annotations and also provide a means to compare TSR and TVD as annotation signal approximation techniques. Fig. 5 also shows the best resulting hand-tuned signal approximations using both methods. The λ parameter in TVD is tuned using a log-scale grid search between 1^{-4} and 1^3 and TSR is tuned over a grid $\{T-2, T-1, ..., T+2\}$ where T is chosen by counting the number of peaks, valleys, and plateaus in the signal and doubling it. In order to compare the impact of using either TVD or TSR as part of the signal warping algorithm in Fig. 1, an oracle is used to generate triplet comparisons for the extracted constant intervals and produce the final warped ground truth signals. Table I displays different correlation measures between these final warped signals (not pictured) and the corresponding true signals.

We evaluate the robustness of TSR and TVD by comparing the Kendall tau correlations of their resulting warped signals to the true signals for different T and λ parameter settings. In both cases, extracted constant segments with durations shorter than 200ms are ignored because they are shorter than average human perception-touch response times [21]. Fig. 6 plots these results for both tasks and also displays the corresponding number of constant segments produced as the tunable parameters vary for each of the two methods.



(a) Signal approximations for TaskA. $T = 51$ for TSR and $\lambda = 0.001$ for TVD.

(b) Signal approximations for TaskB. $T = 75$ for TSR and $\lambda = 0.001$ for TVD.

Fig. 5. Plots of the TSR and TVD signal approximations of the *EvalDep* fusion in both green intensity annotation tasks.

TABLE I
AGREEMENT MEASURES BETWEEN THE TRUE SIGNAL AND VARIOUS GROUND TRUTH ESTIMATES USING THE SIGNAL WARPING METHOD

Task	Ground Truth Technique	Pearson	Spearman	Kendall's Tau	NMI
A	EvalDep Average	0.906	0.946	0.830	0.807
	Warped EvalDep with TVD	0.967	0.939	0.835	0.818
	Warped EvalDep with TSR	0.974	0.939	0.831	0.818
B	EvalDep Average	0.969	0.969	0.855	0.947
	Warped EvalDep with TVD	0.981	0.980	0.881	0.987
	Warped EvalDep with TSR	0.990	0.988	0.911	0.989

All warped results use a complete set of ordinal comparisons from the oracle. For more details see [1]. NMI = normalized mutual information.

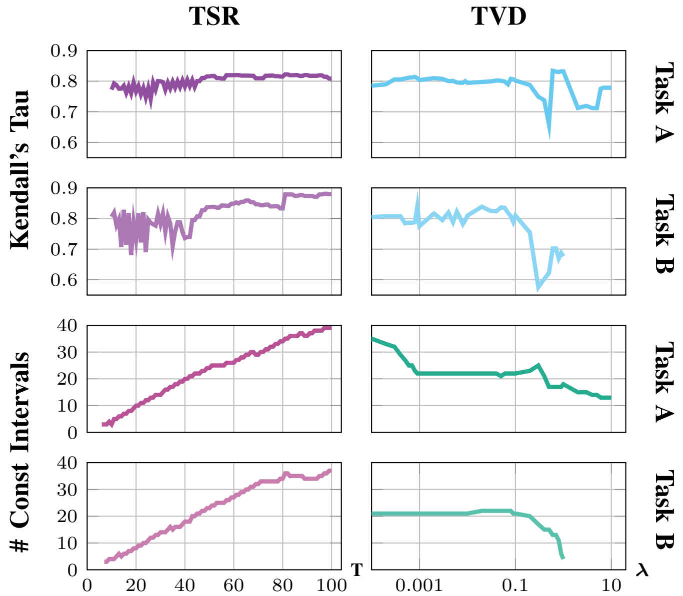


Fig. 6. Differences between TSR and TVD algorithms in two experimental tasks when used for signal approximation as part of the signal warping ground truth method. The number of constant intervals extracted during the segmentation stage of the pipeline are shown at the bottom as functions of each algorithm's tunable parameter (T or λ). Kendall's tau correlations between the final warped signal and the true target signal are shown at the top.

V. DISCUSSION

Results from Table I suggest that the ground truth signals produced using the fused annotation warping method employing TSR for signal approximation are similar and perhaps better than those produced using TVD. Fig. 5 indeed shows the two methods produce very similar approximations to the fused annotation signals in both TaskA and TaskB.

The TSR method does, however, have two other advantages over TVD. The plots of ground truth correlations over a range of parameter settings for TSR and TVD in Fig. 6 reveal that the final warped signal quality is more stable for variations of T (TSR) near the optimum setting compared to the variance in correlation as λ (TVD) varies near its optimum. The interpretablity of T enables humans to provide a good initial estimate and the results in these two task experiments suggest that this estimate will perform comparably to the optimum. The sharp peaks near $\lambda = 1$ in TaskA or $\lambda = 0.001$ in TaskB suggest that finding the optimum when using TVD requires an exhaustive search over a fine grid.

Another benefit to the TSR method is the predictability of the number of constant intervals exceeding the minimum duration of 200ms as shown in Fig. 6. The number produced using TSR is nearly linear and independent of the signal being approximated whereas the number generated by TVD is non-linear and dependent on the structure of the signal. This offers a level of control to users of the signal warping method who may want to minimize the amount of required supplemental annotation during the triplet comparison stage by reducing the number of constant intervals. There is an interesting optimization problem here attempting to maximize the expected ground truth quality while minimizing annotation costs that is worth exploring in future work.

The only downside to using TSR seems to be its computational complexity. We believe the extra time required is worth the cost when it produces a more robust and accurate ground truth. Speeding up TSR to achieve a good sub-optimum trapezoidal signal regression is a compelling subject for future research as well.

VI. Conclusion

We present a new method, *trapezoidal segmented regression*, for optimally approximating fused human-produced continuous-scale annotations as trapezoidal functions. The method is proposed and evaluated as an alternative to *total variation denoising* used as a part of an ordinal signal warping methodology [1] for generating more accurate ground truth labels from human annotations. The two methods and the ground truth labels they help produce are compared in experiments where true annotation target signals are known *a priori*. Results show that the proposed signal approximation method performs on par with the prior approach, producing ground truth signals in better alignment with the true target, but with the added advantages of being more robust, more easily tuned, and more intuitive.

References

[1] Brandon M Booth, Karel Mundnich, and Shrikanth S Narayanan. "A novel method for human bias correction of continuous-time annotations". In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 3091–3095.

[2] Benedek Kurdi, Shayn Lozano, and Mahzarin R Banaji. "Introducing the open affective standardized image set (OASIS)". In: *Behavior research methods* 49.2 (2017), pp. 457–470.

[3] Mohammad Soleymani et al. "A multimodal database for affect recognition and implicit tagging". In: *IEEE Transactions on Affective Computing* 3.1 (2012), pp. 42–55.

[4] Sander Koelstra et al. "Deap: A database for emotion analysis; using physiological signals". In: *IEEE transactions on affective computing* 3.1 (2012), pp. 18–31.

[5] Carlos Busso et al. "IEMOCAP: Interactive emotional dyadic motion capture database". In: *Language resources and evaluation* 42.4 (2008), p. 335.

[6] Fabien Ringeval et al. "Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions". In: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE. 2013, pp. 1–8.

[7] Ellen Douglas-Cowie et al. "The HUMAINE database". In: *Emotion-Oriented Systems*. Springer, 2011, pp. 243–284.

[8] Mihalis A Nicolaou, Stefanos Zafeiriou, and Maja Pantic. "Correlated-spaces regression for learning continuous emotion dimensions". In: *Proceedings of the 21st ACM international conference on Multimedia*. ACM. 2013, pp. 773–776.

[9] Harold Hotelling. "Relations between two sets of variates". In: *Biometrika* 28.3/4 (1936), pp. 321–377.

[10] Mihalis A Nicolaou, Vladimir Pavlovic, and Maja Pantic. "Dynamic probabilistic cca for analysis of affective behavior and fusion of continuous annotations". In: *IEEE transactions on pattern analysis and machine intelligence* 36.7 (2014), pp. 1299–1311.

[11] Galen Andrew et al. "Deep canonical correlation analysis". In: *Proceedings of the International Conference on Machine Learning*. 2013, pp. 1247–1255.

[12] 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.

[13] Meinard Müller. "Dynamic time warping". In: *Information retrieval for music and motion* (2007), pp. 69–84.

[14] Feng Zhou and Fernando De la Torre. "Generalized canonical time warping". In: *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2016), pp. 279–294.

[15] George Trigeorgis et al. "Deep Canonical Time Warping". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 5110–5118.

[16] Fabien Lauer. "On the complexity of piecewise affine system identification". In: *Automatica* 62 (2015), pp. 148–153.

[17] M Govind and TN Ruckmongathan. "Trapezoidal and triangular waveform profiles for reducing power dissipation in liquid crystal displays". In: *Journal of Display Technology* 4.2 (2008), pp. 166–172.

[18] Eduardo Camponogara and Luiz Fernando Nazari. "Models and algorithms for optimal piecewise-linear function approximation". In: *Mathematical Problems in Engineering* 2015 (2015).

[19] Laurent Condat. "A direct algorithm for 1-D total variation denoising". In: *IEEE Signal Processing Letters* 20.11 (2013), pp. 1054–1057.

[20] Brandon M. Booth. *2018 Continuous Annotations*. https://github.com/brandon-m-booth/2018_continuous_annotations. 2019.

[21] Aditya Jain et al. "A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students". In: *International Journal of Applied and Basic Medical Research* 5.2 (2015), p. 124.