Merging Pose Estimates Across Space and Time: Appendices

Xavier P. Burgos-Artizzu[1]
xpburgos@caltech.edu
David Hall[1]
dhall@caltech.edu
Pietro Perona[1]
perona@caltech.edu
Piotr Dollár[2]
pdollar@microsoft.com

[1] California Institute of Technology
Pasadena, CA, USA
[2] Microsoft Research
Redmond, WA, USA

Appendix A: Extension to Angular Data

Earlier we assumed estimates x lived in $\mathbb{R}^D$ and that (squared) Euclidean distance was a meaningful distance measure. However, for many quantities of interest, such as angles, Euclidean distance is not appropriate (*e.g.*, given two angles $\theta_1 = 5^\circ$ and $\theta_2 = 355^\circ$ the difference between them is 10° not 350°). Below we extend our approach to angular data.

We begin by introducing a distance measure between angles:

$$d_\theta^2(\theta_1, \theta_2) \doteq \|\theta_1 - \theta_2\|_\theta^2 \doteq 2 - 2\cos(\theta_1 - \theta_2). \tag{1}$$

All angles are given in radians unless otherwise specified. The distance $d_\theta^2(\theta_1, \theta_2)$ is equal to the squared Euclidean distance between the two points on the unit circle associated with θ_1 and θ_2. That is given the points $p_1 = (\cos\theta_1, \sin\theta_1)$ and $p_2 = (\cos\theta_2, \sin\theta_2)$, the squared Euclidean distance between them is:

$$d(p_1, p_2) = 2 - 2(\cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2) = 2 - 2\cos(\theta_1 - \theta_2). \tag{2}$$

Using the second order approximation of $\cos\theta \approx 1 - \theta^2/2$ near $\theta = 0$ gives $\|\theta_1\|_\theta \approx \sqrt{2 - 2(1 - \theta_1^2/2)} = |\theta_1|$, which simply states that for small angles straight line distance between two points on a unit circle is approximately the same as the arc length. In addition, $\|2\theta_1\|_\theta \approx 2\min(\|\theta_1\|_\theta, \|\theta_1 + \pi\|_\theta)$ holds for all θ_1 with a maximum error of $\sqrt{2}$ and also proves useful (proof omitted).

The above distance is convenient because it allows for computing the mean of a set of circular quantities in closed form. Specifically (proof omitted):

$$\theta^* = \arg\min_\theta \sum_{i=1}^n d_\theta^2(\theta, \theta_i) = \text{atan2}\left(\sum_{i=1}^n \sin\theta_i, \sum_{i=1}^n \cos\theta_i\right) \tag{3}$$

The quantity θ^* is known as the *circular mean*. If all angles are represented by their associated point on the unit circle, the point (on the circle) that minimizes the sum of distances to the remaining points is the arithmetic mean of the points projected back onto the unit circle. The circular mean is undefined if the arithmetic mean is at the origin. Finally, given two angles θ_1 and θ_2, their circular mean is either $\theta^* = (\theta_1 + \theta_2)/2$ or $\theta^* = (\theta_1 + \theta_2)/2 + \pi$.

Given the above, it is straightforward to extend k-means clustering to angles. Suppose all angles are represented by their associated points on the unit circle. As in standard k-means one can iterate between assigning points to cluster centers and then recomputing the centers. The only difference is that centers need to be re-projected onto the unit circle after being computed using the arithmetic mean in each phase.

Appendix B: Identity Assignment

Until now, we have assumed that the objects being tracked are identical and that in a given frame t any detection x_i^t can be explained by any track y_k^t. Often, however, we may wish to enforce a given detection to belong to a specific trajectory or subset of trajectories. This is useful if we have object specific detectors or some other source of information that associates detections with object identity, such as human labeled frames.

We can formalize this as follows. In addition to X^t and S^t, for each frame t we have $R^t = \{r_1^t, \dots, r_{n_t}^t\}$ where $r_i^t \subseteq \{1, \dots, K\}$ denotes possible identity assignments for detection x_i^t. Specifically, $k \in r_i^t$ means x_i^t can be assigned to trajectory k. Given no information about identity then $r_i^t = \{1, \dots, K\}$, otherwise r_i^t is a subset according to the possible identity assignments for detection x_i^t. The corresponding modification to $L_{space}(Y^t)$ is then:

$$L_{space}(Y^t) = \frac{1}{s^t} \sum_{i=1}^{n_t} \min_{k \in r_i^t} d_{bd}(x_i^t, y_k^t) s_i^t. \tag{4}$$

The only difference between the above and Eqn. (2) in the main text is that k is constrained to belong to r_i^t. Note that if no identity information is given, that is $r_i^t = \{1, \dots, K\}$ for all i and t, the above reduces to Eqn. (2) in the main text. Only minor modifications to the optimization procedure need to be made to incorporate this term.