Description of PSU HUB Data Files, Release V1, February 2012

See also examples.m file for concrete examples of how to read and use the data files.

The HUB dataset consists of two 15-minute sequences taken in the Penn State student union building (also known as "The HUB"). The first sequence, called SU2-L has low-density crowds, and the second sequence, SU2-H, has higher crowd density and is thus more difficult. Each sequence is broken into a series of short 20 second clips. All people in the middle frame of each clip were detected and tracked by human volunteers to collect ground truth trajectories. Examples trajectories from two clips are shown in Figure 1. The volunteers also labeled people that they thought were together in a small group. The grouping labels provided by six different volunteers were pooled together to produce a "human consensus" ground truth classification for every person in the scene, into categories of "alone", "in a pair", and "in a group of 3 or more".
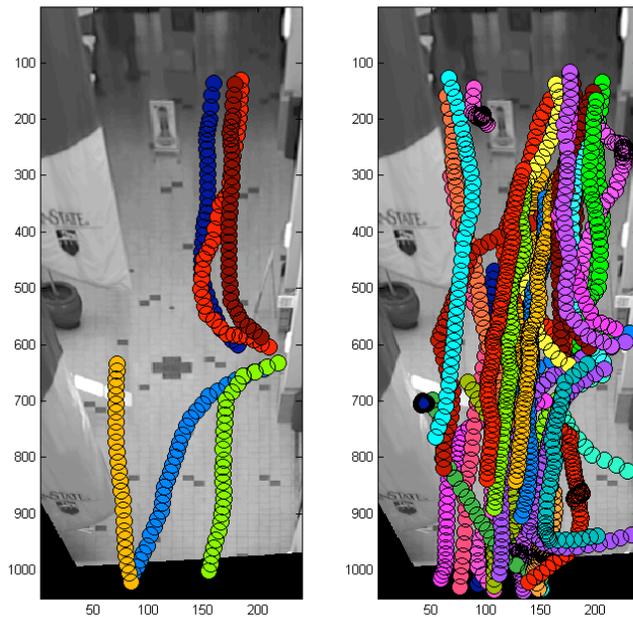

Figure 1: Sample ground truth trajectories. Left: from clip 50 of SU2-L (low density sequence). Right: Clip 20 of SU2-H (high density sequence).

The function read_traj_block.m reads in the trajectories for the people in each clip. See the comments at top of that file for more information. The routine returns arrays of X,Y coordinates for the trajectories of all people in the clip. Since the human volunteers only clicked on each person's location every 10 frames over the 20 second clip, there are 61 points for each trajectory. The X,Y locations are set to -1 for sampled frames where that person was not visible, otherwise they will be a positive numbers representing col and row in a top-down view (look at how the trajectories are displayed in read_traj_block.m for more details). Also returned in a third array are ID numbers for each person. These index into the second ground truth file, described below.

Consensus ground truth classifications are provided in the file
nummembers_all_consensus.dat, which contains lines in the following format

```
510  4 4 3 4 4 2 4 4 3 2
511  4 4 3 4 4 1 4 4 3 2
512  1 1 1 1 1 1 1 6 1 1
513  3 3 3 3 3 3 3 6 3 2
514  3 3 3 3 3 3 3 6 3 2
515  1 0 0 1 0 1 0 3 -1 -1
```

The first column contains a person's ID number. The next 6 columns (columns 2-7)
contain raw labeling results from each of the 6 human coders, estimating the size of the
group that person was with. Column 8 contains the mode of the numbers in columns 2-7,
and column 9 contains a count of the number of coders who contributed to that modal
value. If the mode is not unique (for example, if 3 coders thought the person was alone,
and the other 3 coders thought the person was part of a pair), the smallest modal value is
the one reported in column 8. Finally the two columns of most interest to you are the last
two. Column 10 contains the "trichotomous" ground truth classification label for the
individual, into three categories : alone (value 1), in a pair (value 2), and in a group of
three or more (value 3). Column 11 contains a "dichotomous" ground truth label into
only two categories : alone (value 1) or in a group (value 2). Finally, for individuals
where there was no consensus (no mode, or less than a majority of coders labeled that
individual), the classification labels in these last two columns will be -1 (for example, the
person with ID 515 above), and you should just ignore that individual when training or
testing. Using person ID 511 as an example, we see that 4 coders (1, 2, 4 and 5) think
they are in a group of 4 people, while coder 3 thinks they are in a group of only 3, and
coder 6 thinks they are alone. Despite the subjective disagreement, the consensus is that
511 is in a group of size 4, as determined by 4 coders, and thus the trichotomous class
label for this ID is 3 (group of 3 or more), and dichotomous label is 2 (in a group).

The nummembers_all_consensus.dat file tells us estimates of the group size each person
is with, but it doesn't identify which people were together. The files pairweights1.dat, for
the sparse sequence, and pairweights3.dat, for the dense sequence, help fill in that
information by specifying pairwise links of the form [ID1 ID2 count], for example:

```
502 506 4
502 510 5
502 511 4
506 510 5
506 511 5
510 511 5
503 513 6
503 514 6
513 514 6
```

In these links, ID1and ID2 specify two individuals, and the count tells the number of coders who thought those two people were together in the same group. ID1 is always < ID2, so any pair of people will be listed only once. To save space, only pairs with nonzero counts are included in the file. Counts range from 1 to 6, with 6 telling us that the pairwise link was unanimously thought to exist by all 6 human coders. A count of 1 means only a single coder thought those two people were together. You can think of these counts as weights on the pairwise links, or probabilities if you divide by 6. The transitive closure of high-confidence pairwise links, taken together with group size information from the nummembers_all_consensus.dat file should enable recovering the labels of all people in each group. For example, we saw earlier that IDs 513 and 514 were thought to be in a group of 3 by all coders. From the pair information, we see high confidence pairwise links between 513 and 514, 503 and 513, and 503 and 514. We therefore infer that 503, 513 and 514 are the members of that group of 3. As another example of transitive closure, we can infer that 502, 506, 510, 511 form a group of 4 people, which is also consistent with the nummembers information for 510 and 511.

Cavaet: Because group membership is based on subjective decisions made independently by multiple human coders, there may be some inconsistencies in groups inferred from low-confidence pairwise links. However, if you only stick to high-confidence links that were agreed on by 4 or more coders, those problems should be minimal. Please let me know if you find significant inconsistencies between the pairwise link information and the group count information in nummembers_all_consensus.

A companion paper describes the project this data was collected for (this is called the "SU2" dataset in the paper). We ask that you refer to this paper if you use the data in your own research publications:

Weina Ge, Robert Collins, and Barry Ruback, "Vision-based Analysis of Small Groups in Pedestrian Crowds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear, 2012.

Finally, due to conditions of the PSU Institutional Review Board (IRB) permissions that allowed us to collect this dataset on campus, we are unfortunately not allowed to make the original video or images available to the public. We regret the highly restrictive nature of this IRB, and will seek to negotiate better terms before undertaking any future data collections.

Bob Collins, Penn State University, Feb 2012