Ground truth annotations for UCSD dataset

Antoni Chan (abchan - AT - ucsd.edu)
Department of Computer Science, City University of Hong Kong

Nuno Vasconcelos (nuno - AT - ece.ucsd.edu)
Statistical Visual Computing Lab, University of California, San Diego

Feb 28, 2013

This is the ground-truth people annotations of the UCSD pedestrian dataset used in [1, 2, 3]. If you use this data, pleace cite [3].

1 Ground truth Annotations

Ground-truth annotations are provided for the first 4000 frames of the `vidf` and `vidd` scenes, i.e., the clips `vidf1_33_000.y` to `vidf1_33_019.y`, and `vidd1_33_000.y` to `vidd1_33_019.y`. These correspond to the Peds1 and Peds2 dataset in [3].

The ground-truth pedestrian locations were marked in every 5 frames of the video and interpolated in between. The data is provided in a "person-centric" format, with a list of people and their tracks, or a "frame-centric" format, with the pedestrian locations listed per frame. The region-of-interest (ROI) and perspective map used in [1, 2, 3] are also provided. Finally, the ground-truth counts of people within the ROI are also provided. This is the counting data used in [3] as Peds1 and Peds2. The ground-truth data are saved in the following MATLAB files in the `gt/vidf` and `gt/vidd` directories:

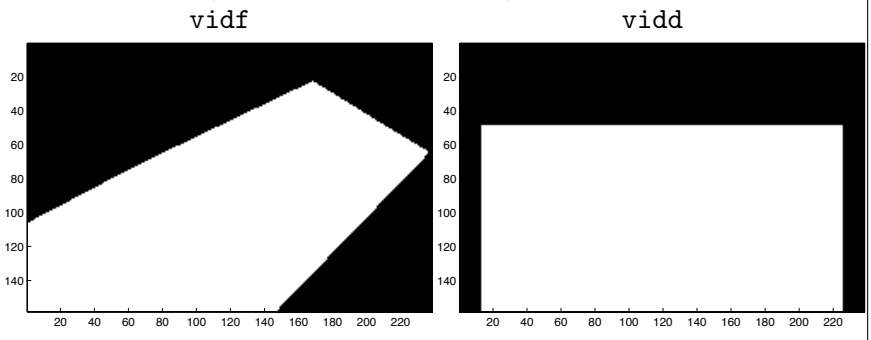
- **vidX1_33_ZZZ_people_full.mat**: location annotations listed by person for video `vidX1_33_ZZZ.y`, where `X` is the video name (d or f), and `ZZZ` is clip number `000` to `019`. The file contains a cell-array of people and their locations in the video clip:

`people{i}.id`	the unique ID for the i-th person. Each person has a unique ID throughout all the video.
`people{i}.loc`	the ground-truth locations of the i-th person in this video, where each row is the location `[x, y, frame]`.
`people{i}.num_pts`	the number of annotations for the i-th person in this video.
`people{i}.ldir`	the instantaneous velocity of the i-th person in this video, where each row is the vector `[dx, dy, frame]`.
`people{i}.tdir`	the traveling direction of the i-th person. For `vidf`, the values are "`l`" for left (towards the camera), "`r`" for right (away from the camera), "`n`" for no motion (stationary). For `vidd`, the values are "`rs`", "`ls`", "`rf`", "`lf`", "`n`" for right-slow, left-slow, right-fast, left-fast, and no motion, respectively.

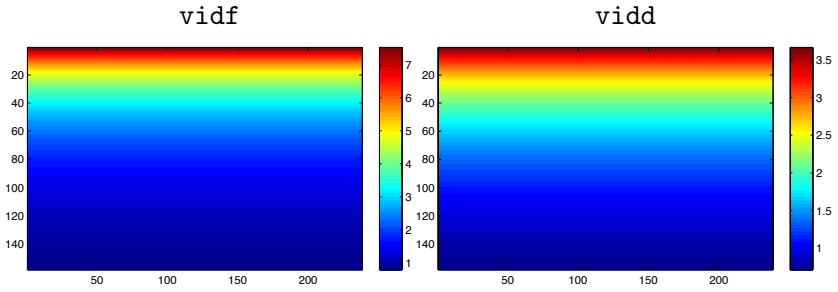
- **vidX1_33_ZZZ_frame_full.mat**: location annotations listed by frame for video clip `vidX1_33_ZZZ.y`. The file contains a cell-array of frames, containing the locations of people in that frame:

`frame{t}.id(i)`	the unique ID of the i-th person in the t-th frame of this video.
`frame{t}.loc(i,:)`	the ground-truth locations of the i-th person in the t-th frame in this video, where the location is `[x, y, frame]`.
`frame{i}.ldir(i,:)`	the instantaneous velocity of the i-th person in the t-th frame in this video, where the vector `[dx, dy, frame]`.
`frame{i}.tdir`	the traveling direction of the i-th person in the t-th frame in this video. See `people{i}.tdir` for possible values.

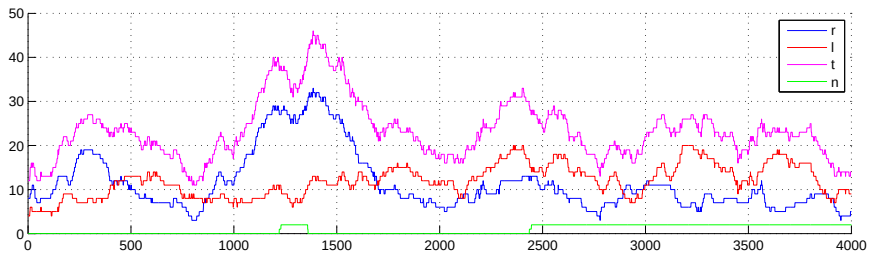
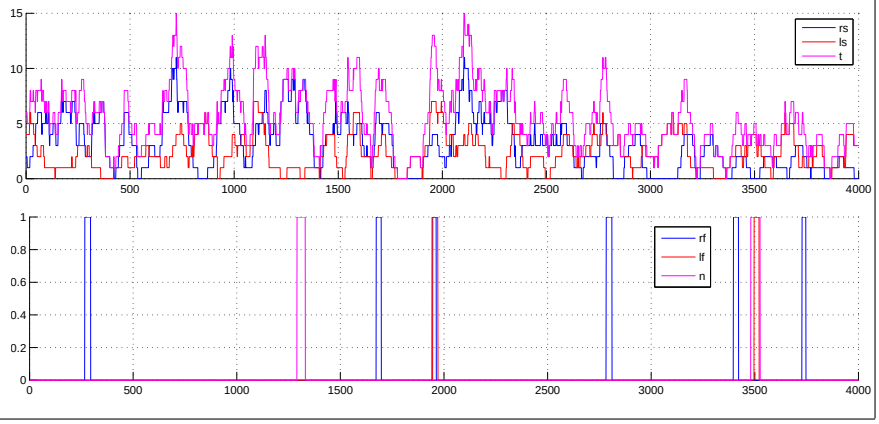
- **vidX1_33_roi*.mat**: the region-of-interest (ROI) for the counting data. The file contains the following variables:

`roi.mask`	mask of the ROI (1 is ROI, 0 background). The ROIs looks like this: 
`roi.xi`	x-coordinates of the ROI polygon.
`roi.yi`	y-coordinates of the ROI polygon.

- **vidX1_33_dmap3.mat**: the perspective map of the scene. The perspective map weighs pixels that originate from objects closer to the camera less than pixels from objects further from the camera. See [3] for more details. The file contains the following variables:

`dmap.pmapxy`	the perspective map, a 2d image where each location `dmap.pmapxy(y,x)` is the weight for pixel (y,x). The map looks like this: 
`dmap.pmapx`	the perspective map for the x-direction (width) only
`dmap.pmapy`	the perspective map for the y-direction (height) only.

- **vidX1_33_ZZZ_count_YK_roi*.mat**: the pedestrian count over the region-of-interest in video `vidX_33_ZZZ.y`. Y is the number of motion classes, either 2 or 4 for `vidf` or `vidd`. The file contains the following variables:

`dirs{j}`	the name of the j-th direction. See `people{i}.tdir` for motion directions. "`t`" means all moving people.
`count{j}(t)`	the number of people moving in the j-th direction, in the t-th frame of this video. Here is a plot of the counts over all frames: `vidf`  `vidd` 

2 Acknowledgments

The collection of this database was funded by US National Science Foundation Award IIS-0534985, NSF IGERT Award DGE-0333451, and by the Research Grants Council of Hong Kong Special Administrative Region of China under Grant CityU 110610 (9041552). Part of the ground-truth was annotated by Jeffrey Cuenco and Zhang-Sheng John Liang.

3 History

- 2013/02/28 - added "vidd" and "vidf", moved experiment results to another README file.
- 2013/02/07 - added "scene" results, updated websites
- 2008/09/14 - initial version

References

[1] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking," In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[2] A. B. Chan and N. Vasconcelos, "Bayesian Poisson Regression for Crowd Counting," In *IEEE Intl Conf. on Computer Vision*, Kyoto, Sept 2009.

[3] A. B. Chan and N. Vasconcelos, "Counting People with Low-Level Features and Bayesian Regression," *IEEE Trans. on Image Processing*, vol. 21(4), pp. 2160-77, April 2012.

[4] http://www.svcl.ucsd.edu/projects/crowds

[5] http://www.svcl.ucsd.edu/projects/peoplecnt

[6] http://visal.cs.cityu.edu.hk/research/peoplecnt/