

# Groundtruth for “vidf” UCSD sequence

Statistical Visual Computing Lab  
University of California, San Diego

Antoni Chan ([abchan - AT - ucsd.edu](mailto:abchan - AT - ucsd.edu))  
Nuno Vasconcelos ([nuno - AT - ece.ucsd.edu](mailto:nuno - AT - ece.ucsd.edu))

Sept 14, 2008

This is the ground-truth data for the “vidf” sequence from the UCSD pedestrian database used in “Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking” [1].

## 1 Groundtruth Data

The ground-truth pedestrian locations were marked in every 5 frames of the video (`vidf1_33_000.y` to `vidf1_33_009.y`), 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] are also provided. Finally, the ground-truth counts of people within the ROI are also provided. This is the counting data used in [1]. The ground-truth data are saved in the following MATLAB files:

- **vidf1\_33\_XXX\_people\_full.mat**: location annotations for video `vidf_33_XXX.y` (where `XXX` is 000 to 009), listed by person. The file contains a cell-array of people and their locations in the video:

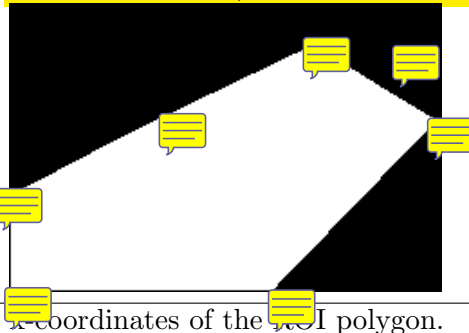
<code>people{i}.id</code>	the unique ID for the $i$ -th person. Each person has a unique ID throughout all the video.
<code>people{i}.loc</code>	the ground-truth locations of the $i$ -th person in this video, where each row is the location <code>[x, y, frame]</code> .
<code>people{i}.num_pts</code>	the number of annotations for the $i$ -th person in this video.
<code>people{i}.ldir</code>	the instantaneous velocity of the $i$ -th person in this video, where each row is the vector <code>[dx, dy, frame]</code> .
<code>people{i}.tdir</code>	the traveling direction of the $i$ -th person, either “l” for left (towards the camera) or “r” for right (away from the camera).

- **vidf1\_33\_XXX\_frame\_full.mat**: location annotations for video `vidf_33_XXX.y` (where `XXX` is 000 to 009), listed by frame. The file contains a cell-array of frames, containing the locations of people in that frame:

<code>frame{t}.id(i)</code>	the unique ID of the i-th person in the t-th frame of this video.
<code>frame{t}.loc(i,:)</code>	the ground-truth locations of the i-th person in the t-th frame in this video, where the location is <code>[x, y, frame]</code> .
<code>frame{i}.ldir(i,:)</code>	the instantaneous velocity of the i-th person in the t-th frame in this video, where the vector <code>[dx, dy, frame]</code> .
<code>frame{i}.tdir</code>	the traveling direction of the i-th person in the t-th frame in this video, either “l” for left (towards the camera) or “r” for right (away from the camera).

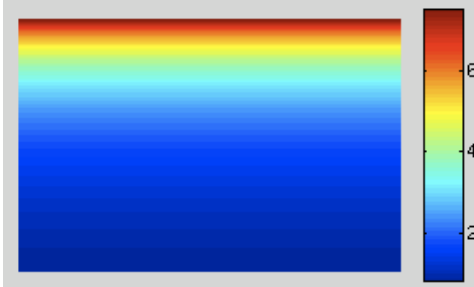
- **vidf1\_33\_roi\_mainwalkway.mat**: the region-of-interest (ROI) for the counting data. The file contains the following variables:

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

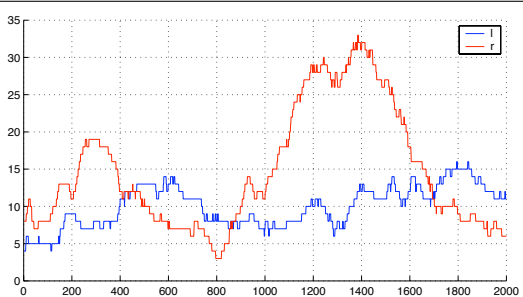


- **vidf1\_33\_dmap3.mat**: the perspective map of the scene. The perspective map weighs pixels that originate from objects closer to the camera less than pixels from objects further from the camera. See [1] for more details. The file contains the following variables:

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



- **vidf1\_33\_XXX\_count\_roi\_mainwalkway.mat**: the pedestrian count in video `vidf_33_XXX.y` (where XXX is 000 to 009), over the region-of-interest. The file contains the following variables:

<div style="border: 1px solid red; padding: 2px; display: inline-block;"> <math>\text{dirs}\{1\}='l'</math>  <math>\text{dirs}\{2\}='r'</math> </div>	
$\text{dirs}\{j\}$	the name of the $j$ -th direction, either “l” or “r”.
$\text{count}\{j\}(t)$	the number of people moving in the $j$ -th direction, in the $t$ -th frame of this video. Here is a plot of the two directions over all videos (000-009):
<div style="border: 1px solid red; padding: 2px; display: inline-block;"> <math>t</math>的范围是1-200.         </div>	

## 2 Experiments

The experiments in [1] use the provided ROI, perspective map, and counting data. Frames 601 to 1400 (corresponding to videos 003 to 006) were used for the training set, and the remaining frames were used as the test set. There was a small bug in the counting data reported in [1], but that is fixed here. You can find more information in the paper [1], or online at [2, 3]. If you use this data, please cite [1].

## 3 History

- 2008/09/14 - initial version

## 4 Acknowledgments

The collection of this database was funded by US National Science Foundation Award IIS-0534985 and NSF IGERT Award DGE-0333451. Part of the ground-truth was annotated by Jeffrey Cuenoco and Zhang-Sheng John Liang.

## 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] <http://www.svcl.ucsd.edu/projects/crowds>
- [3] <http://www.svcl.ucsd.edu/projects/peoplecnt>
- [4] A. B. Chan and N. Vasconcelos, “Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30(5), pp. 909-926, May 2008.