Surveillance Camera Autocalibration based on Pedestrian Height Distributions

Jingchen Liu
jingchen@cse.psu.edu
Robert T. Collins
rcollins@cse.psu.edu
Yanxi Liu
yanxi@cse.psu.edu

Laboratory for Perception, Action and Cognition
Pennsylvania State University
State College, PA, USA
http://vision.cse.psu.edu

We propose a new framework for automatic surveillance camera calibration by observing videos of pedestrians walking through the scene. Existing methods require accurate pedestrian detection and tracking [1, 2, 3] and have limited practical use in more challenging real-world environments (Fig. 1). Our method takes noisy foreground masks as input with no need to track or label correspondences of the same person between frames. Instead, we look for a camera model that recovers the most realistic explanation of the relative pedestrian heights. This is motivated by strong prior knowledge that 90% of human heights fall within a very small range of ±7.6% from the mean [4].

Let f and a be the intrinsic parameters of focal length and (known) aspect ratio, respectively. The extrinsic parameters are specified by a vertical camera translation h_c above the ground plane, a tilt θ and roll angle ρ. The vertical vanishing point $v_0 = (v_x, v_y, 1)^T$, calibration parameters and horizon line are related by

$$\rho = \mathrm{atan}(-av_x/v_y) \tag{1}$$

$$\theta = \mathrm{atan2}(\sqrt{a^2v_x^2+v_y^2}, -af) \tag{2}$$

$$v_x x + \frac{v_y}{a^2}y + f^2 = 0, \tag{3}$$

Our approach first estimates the vertical vanishing point, then searches for a value of f that generates a relative pedestrian height distribution that most resembles our prior knowledge about how pedestrian heights are distributed in the real world.

The vanishing point estimation process is illustrated in Fig. 2. Ellipsis fit to foreground blobs are analyzed by RANSAC to find the largest set of major axes that intersect (approximately) at a single vanishing point. The final estimate of v_0 is computed as a minimum mean-squared angle solution using the entire set of inlier axes.

We define the *relative height* of a pedestrian h_i as their actual 3D height h_i^{3D} divided by the unknown camera height h_c, computed as an invariant using the cross ratio [2] of distances between four points

$$h_i = \frac{h_i^{3D}}{h_c} = 1 - \frac{d(p_h, v_l)\cdot d(p_f, v_0)}{d(p_f, v_l)\cdot d(p_h, v_0)}. \tag{4}$$

where p_f and p_h are feet and head position of the pedestrian, v_0 is the vertical vanishing point, and v_l is the intersection of a line passing through these points and the horizon line. Relative height is thus implicitly a function of the unknown focal length f.

We then define a likelihood function to evaluate the similarity of a 1D relative height distribution computed from a hypothesized value for f, with respect to known characteristics of human height distribution in the real world, namely that heights clusters closely about the mean [4]:

$$\mathcal{L}(O|f, \mu, v_0) \sim \sum_i \frac{1}{\mu} r(h_i, \mu)^2, \tag{5}$$

where μ is average computed pedestrian height and the robust distance metric is defined as

$$r(h_i, \mu) = \max\{\tau - \frac{|h_i - \mu|}{\mu}, 0\}, \tag{6}$$

where τ is the relative distance threshold. A more general form of Eqn. 5 incorporating height uncertainty via bootstrap sampling can be rewritten as:

$$\begin{aligned} \mathcal{L}(O|f, \mu, v_0) &= \sum_i \int \frac{1}{\mu} r(h_i, \mu)^2 p(h_i) \mathrm{d}h_i & (7) \\ &\sim \frac{1}{\mu M} \sum_i \sum_m r(h_i^{(m)}, \mu)^2, & (8) \end{aligned}$$

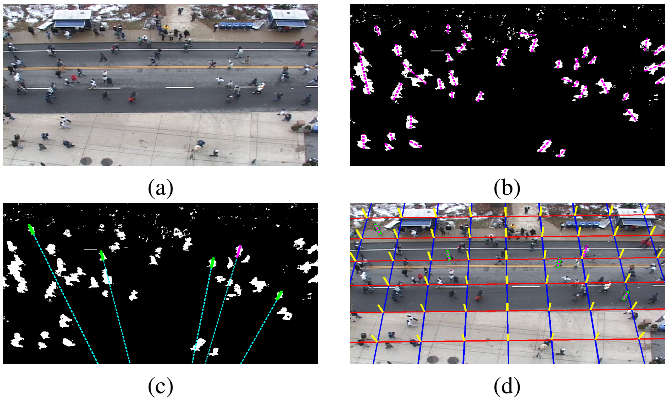
(a)
(b)
(c)
(d)

Figure 1: (a) one video frame; (b) foreground masks; (c) Inliers after RANSAC vanishing point estimation (magenta) and height distribution analysis (green); (d) final calibration results.

where $h_i^{(m)}$ is a random sample of h_i assuming Gaussian noise on pixel locations.

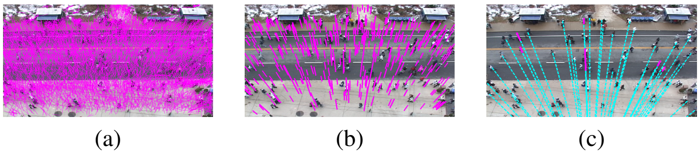
(a)
(b)
(c)

Figure 2: Vertical vanishing point estimation: (a) Major axis of blobs collected over a short sequence. (b) Inlier axes found by RANSAC converge to a single vanishing point. (c) Lines connecting blob centroids to the computed vertical vanishing point.

Experimental results on both synthetic and real data show the robustness of our method to camera pose and noisy foreground detection. Please refer to the paper for detailed qualitative and quantitative evaluation.

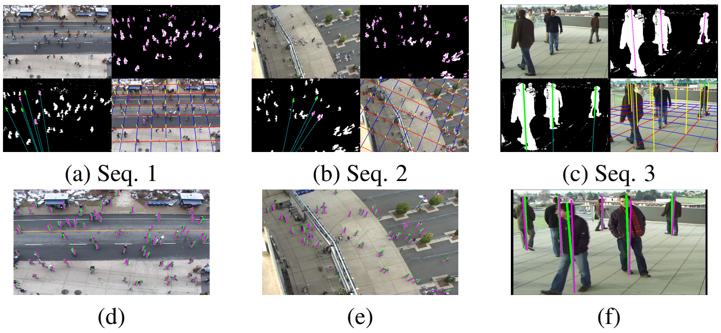
(a) Seq. 1
(b) Seq. 2
(c) Seq. 3
(d)
(e)
(f)

Figure 3: Calibration results with inlier blobs on sample frames (a,b,c); Estimated feet-head projection (magenta) versus groundtruth labels (green) (d,e,f)

[1] Nils Krahnstoever and Paulo R.S. Mendonca. Bayesian autocalibration for surveillance. In *Proc. ICCV*, pages 1858–1865, 2005.

[2] Fengjun Lv, Tao Zhao, and Ramakant Nevatia. Camera calibration from video of a walking human. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28(9):1513–1518, 2006.

[3] Branislav Micusik and Tomas Pajdla. Simultaneous surveillance camera calibration and foot-head homology estimation from human detections. In *Proc. CVPR*, pages 1–8, 2010.

[4] Peter M. Visscher. Sizing up human height variation. *Nature Genetics*, 40(5): 489–490, 2006.