

Model Ensemble to Fuse Geometric and Learning Solutions for Camera Rotation Estimation

Supplementary Material

1. Additional implementation details

We train ResNet-50 model on GSV dataset. We take batch size of 32 images, learning rate of 1e-4 and train the model for 100 epochs using Adam optimizer. We use average of MSE and MAE as loss between network predictions and ground truth values in radians. We use same GSV trained DNN model for all our experimentation on other datasets.

For Geometric approach, we use LSD in OpenCV and Clustering algorithm in [1] with default settings, and no special tuning or engineering was needed. Since all datasets have manmade structures with several vanishing lines, LSD and clustering works quite well. We consider only top 3 clusters returned by [1]. Fig.3d in the main document shows failure cases where spurious clusters are detected and estimated angles have high errors. Such failure cases are addressed by comparing the deviation with DNN estimates. For filtering incorrect geometric estimates, we calculate the deviation of geometric pitch estimates using $|\theta_{dnn} - \theta_{h/v}|$. We ignore geometric estimates when the deviation of $\theta_{h/v}$ from θ_{dnn} is large. For GSV dataset, we take the threshold as 1° and for other datasets it is 5° . We use large threshold for other datasets since our DNN is not trained on these datasets and hence, its own results can deviate from ground truth values.

2. Selecting pairs of lines for parallelity constraints

We show ablation on different ways to form the constraints using vanishing lines and show results in Table.1. First we use brute force approach to form $\binom{N}{k}$ constraints. We next form the constraints by analyzing the slopes of lines. First, we select $N - 1$ pairs by considering adjacent lines sorted as per slopes. Since adjacent lines have similar slopes, it results in an ill-conditioned problem and performs poorly. Next, we select $\frac{N}{2}$ pairs by considering lines with maximum slope difference, *i.e.* we form the pair between i^{th} and $N - i^{th}$ line. We also select i^{th} and $\frac{N}{2} + i^{th}$ lines to form pairs. Both the approaches have $\frac{N}{2}$ constraints and have good performance with $L_i, L_{\frac{N}{2}+i}$ giving best results.

3. Derivation of $m(\theta)$

We present the detailed derivation for change in the slope as a function of camera pitch angle. As discussed, the homography transformation for a point $p = [x, y, 1]^T$ is given

Method	Mean	Med.
All $\binom{N}{2}$ constraints	0.85	0.74
$(N - 1)$ constraints: L_i, L_{i+1}	0.94	0.83
$\frac{N}{2}$ constraints: L_i, L_{N-1}	0.87	0.76
$\frac{N}{2}$ constraints: $L_i, L_{\frac{N}{2}+i}$	0.84	0.70

Table 1. Pitch estimation errors for different methods for generating constraints.

by

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ 1 \end{pmatrix} = K.R.K^{-1} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

To understand the transformation better, it is better to examine the matrix operations individually. Let us consider the slope intercept equation of a line in the perspective view

$$y = mx + c \quad (2)$$

where, m and c are original parameters defining the line. When the K^{-1} operator is applied on the image, the transformed image line can be represented as

$$y' = m'x' + c' \quad (3)$$

where,

$$m' = \frac{mf_x}{f_y} \quad (4)$$

$$c' = \frac{c + mc_x - c_y}{f_y} \quad (5)$$

Next, the Rotation matrix R is applied on Eq. (3) to get

$$y'' = m''x'' + c'' \quad (6)$$

where

$$m'' = \frac{m'}{\cos \theta + c' \sin \theta} \quad (7)$$

$$c'' = \frac{c' \cos \theta - \sin \theta}{\cos \theta + c' \sin \theta} \quad (8)$$

Finally, when matrix K is applied we get,

$$\bar{y} = \bar{m}\bar{x} + \bar{c} \quad (9)$$

where,

$$\bar{m} = \frac{m'' f_y}{f_x} \quad (10)$$

$$\bar{c} = c'' f_y - m'' c_x \frac{f_y}{f_x} + c_y \quad (11)$$

Substituting values from Eqs. (4) to (5) and Eqs. (7) to (8) we get

$$\bar{m} = \frac{m}{\cos \theta + \left(\frac{c + m c_x - c_y}{f_y} \right) \sin \theta} \quad (12)$$

$$(13)$$

4. Robustness of Eigen vector sol and Numerical stability

We note that Eigen vector solution is not robust to the outliers. However, it is more robust to the noise in the data than simple pseudo inverse based approach. To validate this claim, we conduct an experiment where we simulate several synthetic equations and solve them using Pseudo-inverse (Eq.21/22) and eigen vector (Eq.16/17) approach. We add different amount of noise in the data. One can clearly see that eigen vector sol. performs consistently better in presence of noise. We report average of rel. error in solution (in %).

Method	Noise (σ) in data		
	0.0	0.5	1.0
Pseudo-Inv.	0.0	4.06	8.47
Eigen Vector	0.0	0.67	1.76

5. Approximation involved in residual roll correction using geometric method

As mentioned in Sec.4.3 in paper, we first apply roll by ϕ_{dnn} , then apply pitch correction θ_v and then estimate residual roll ϕ_{res} . The overall sequence of operations is given by $R_z(\phi_{res})R_x(\theta_v)R_z(\phi_{dnn})$ which need to be approximated by $R_x(\theta_v)R_z(\phi_{final})$ in order to make final estimate of ϕ_{final} . We make an approximation here giving $\phi_{final} = \phi_{dnn} + \phi_{res} \cos(\theta_v) \approx \phi_{dnn} + \phi_{res}$ as our estimate. We note that above approximation is simple and quite accurate when ϕ_{res} & θ_v are close to 0. Notwithstanding the above, our empirical experiments show that overall estimate ϕ_{final} is better than ϕ_{dnn} .

6. Visual Results

In Fig. 1 and Fig. 2, we demonstrate our visual results. We transform images using estimated pitch angles using both

horizontal and vertical vanishing lines, *i.e.* the estimated θ_h and θ_v . In Fig. 1 first two columns, we show vertical line clusters and images transformed using $\frac{\pi}{2} - \theta_v$ to get the upright view which transforms the lines to be parallel. In next two columns we show transformed view using estimated θ_h which generates BEV and we can see parallel vanishing lines in this view.

Next, in Fig. 2, we show failure case in the first row where both transformations are incorrect. The second row shows cases where only vertical vanishing lines are detected and corresponding transformed views using $\frac{\pi}{2} - \theta_v$ are shown.

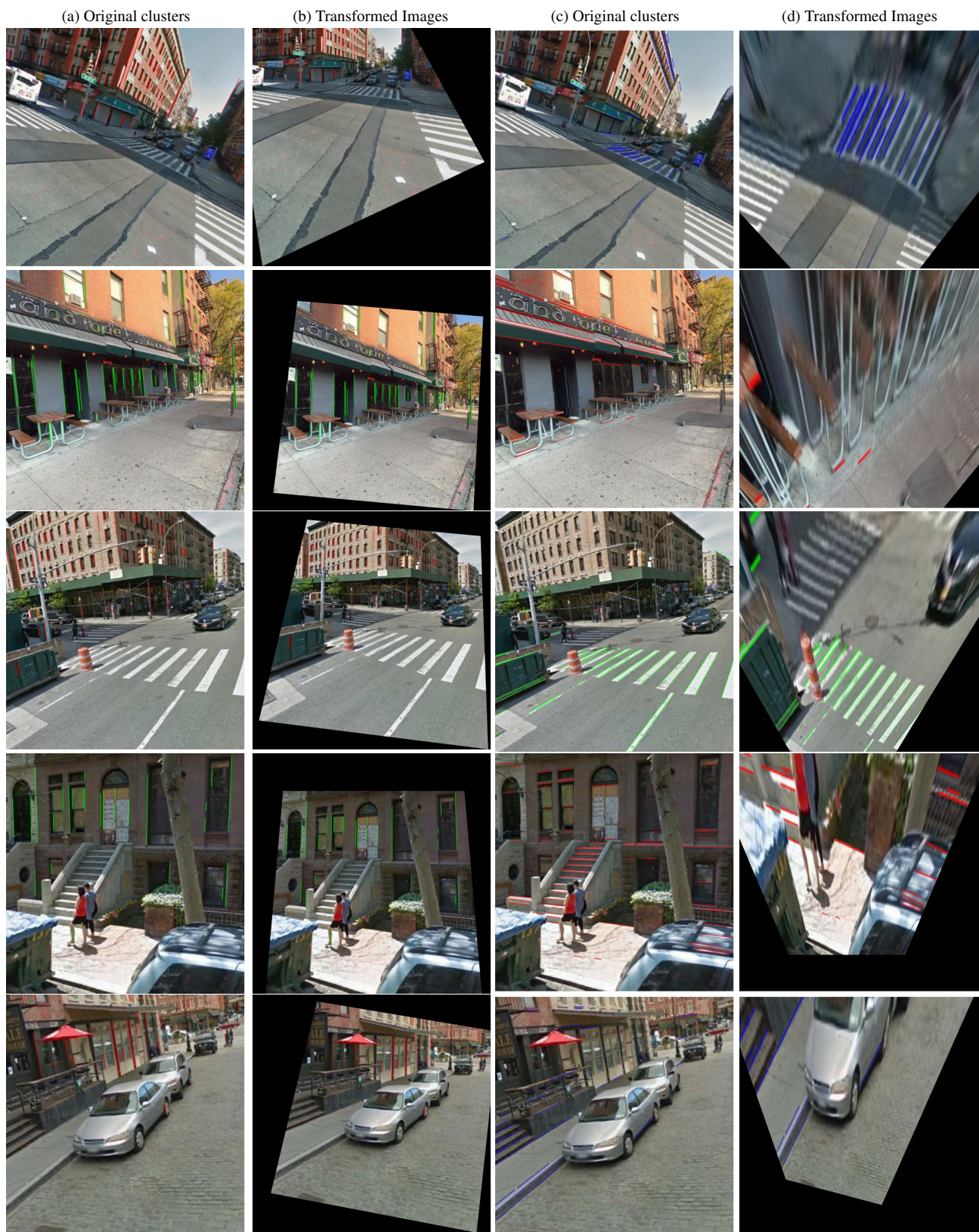


Figure 1. Homography transformation results using estimates of geometric approach.

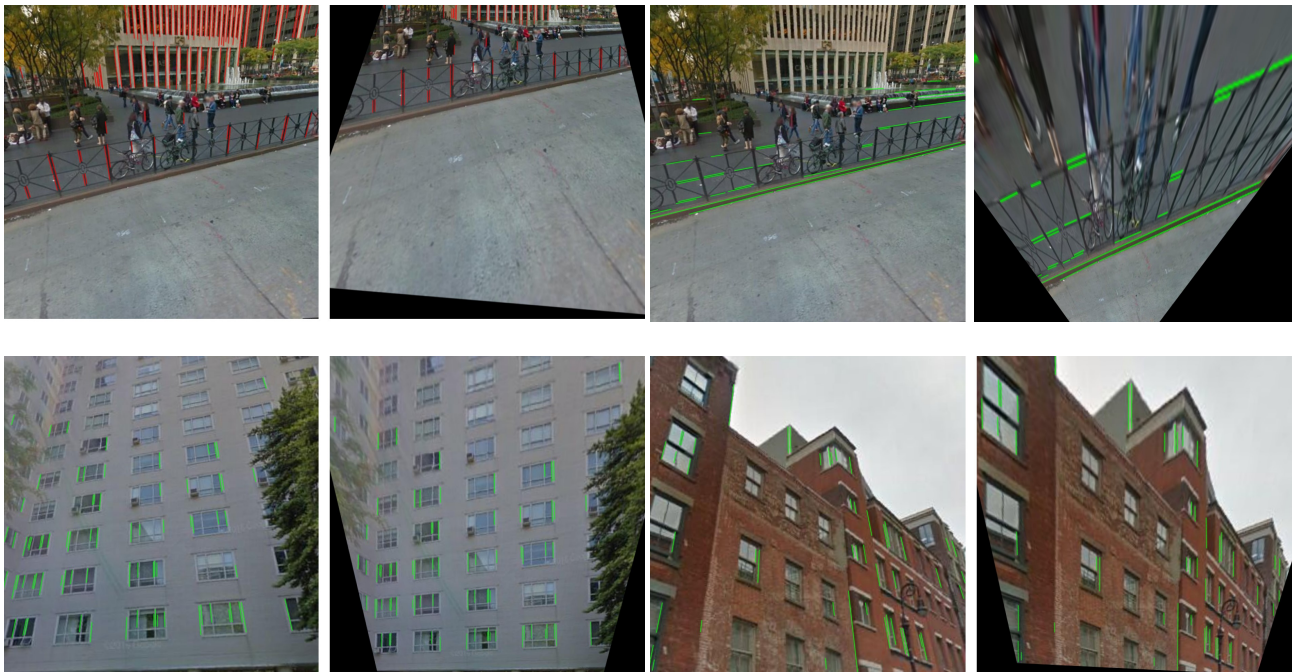


Figure 2. Homography transformation results using estimates of geometric approach.