Visual Tilt Correction for Vehicle-mounted Cameras

Firas Kastantin and Levente Hajder
Department of Algorithms and their Applications, Eötvös Loránd University,
Pázmány Péter stny. 1/C, Budapest 1117, Hungary

Keywords: Camera Tilt Correction, Planar Motion, Vehicle-mounted Camera.

Abstract: Assuming planar motion for vehicle-mounted cameras is very beneficial as the visual estimation problems are simplified. Planar algorithms assume that the image plane is parallel to the gravity direction. This paper proposes two methods to correct camera images if this parallelity constraint does not hold. The first one utilizes that the ground is visible in the camera images, therefore its normal gives the gravity vector; the second method uses vanishing points to estimate the relative correction homography. The accuracy of the novel methods is tested on both synthetic and real data. It is demonstrated in real images of a vehicle-mounted camera that the methods work well in practice.

1 INTRODUCTION

Nowadays, more and more mobile robots, powered with different types of sensors have been appeared, which can be used for localization, navigation, planning, and other applications. 3D computer vision and robotic vision are the research areas to study algorithms and applications of agent-mounted camera(s) which include and not limited to: SLAM (Stachniss et al., 2016), relative motion estimation (Hartley and Zisserman, 2003), and structure from motion (B. Triggs and P. McLauchlan and R. I. Hartley and A. Fitzgibbon, 2000).

The 3D motion between two sensor views can be represented by three translation and another three rotation parameters, therefore the degrees of Freedom (DoFs) are six. In the case of stereo vision, the 3D motion can be estimated using the stereo vision geometry or the so-called Epipolar Geometry. Epipolar geometry estimates the translation only up to scale (the translation direction is estimated), which reduces the motion estimation problem to 5-DoFs.

Different algorithms have been proposed to estimate the motion parameters from the corresponding points of two views using a monocular camera: (i) for uncalibrated cameras, at least seven points are required (Hartley and Zisserman, 2003), however, the 8-points algorithm (Hartley, 1997) is more popular; (ii) for calibrated cameras, five points are adequate (Nister, 2004; Li and Hartley, 2006; Kukelova et al., 2008) to estimate the epipolar parameters.

Special epipolar constraints with 2 parameters can replace the general motion in case of planar motion (e.g on-road vehicles), which leads to simpler constraint, less computation, and higher accuracy. For planar algorithms, it is assumed that the camera images are perpendicular to the ground. A linear 3-points solver (Ortin and Montiel, 2001) and minimal 2-points solvers (Choi and Kim, 2018; Hajder and Barath, 2020a) were proposed to estimate the motion components under planar motion. The semi-calibrated case, when only the common focal length out of the intrinsic parameters are unknown, with planar motion is also very interesting: this special problems (Hajder and Barath, 2020b) have 3 DOFs.

The DoFs for the estimation are crucial for several reasons, one of those is the time demand of the algorithms. As the input data for extrinsic parameter estimation is usually contaminated, robust statistics are needed. The following table shows how many iterations are required if RANSAC (Fischler and Bolles, 1981) is applied to filter out outliers.

Outlier ratio (%)	20%	35%	50%	65%
2 DoFs	3	6	11	23
3 DoFs	5	10	23	69
5 DoFs	8	25	95	569
7 DoFs	13	60	382	4,655
8 DoFs	17	93	766	13,303

Kastantin, F. and Hajder, L.
Visual Tilt Correction for Vehicle-mounted Cameras.
DOI: 10.5220/0010904800003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 5: VISAPP*, pages 863-870
ISBN: 978-989-758-555-5; ISSN: 2184-4321

It is clearly seen that the time demand is significantly determined by the DOFs of the problem. For fast methods, especially if they are applied in real-time systems, fewer DOFs are preferred. Therefore, it is very beneficial if the planar motion assumption can be assumed for visual motion estimation.

Planar Motion. The planar motion constraint assumes perfect camera mounting and completely flat ground plane, however, the assumption is not always true, non-flatten ground plane can cause rotation around x and/or z-axes breaking the planar motion assumption, thus leads to questions about the reliability and long-term stability of these algorithms.

The detection of road non-planarity is challenging but well researched problem. If an Inertial Measurement Unit (IMU) is mounted to the vehicle, the detection of the gravity vector that is perpendicular to the ground plane is a simple task (Saurer et al., 2017; Guan et al., 2019). However, as an IMU measures the forces, when e.g a vehicle accelerates/brakes or takes a blend, the accurate estimation of the gravity direction is more challenging. A sensor fusion method (Weiss and Siegwart, 2011) is required to apply both visual and inertial sensors.

Goals. This paper deals with *pure visual estimation of the gravity vector*. We show here that accurate estimation of that is possible even if only camera images are given. It is assumed that the cameras are calibrated, therefore its intrinsic parameters are known. The final goal of gravity direction estimation is to correct the camera orientation. As it is well-known in geometric computer vision (Hartley and Zisserman, 2003), the camera orientation can be synthetically modified by applying the related homography. This homography can be determined if the gravity vector is known. Moreover, image-based gravity direction estimation can be utilized to calibrate digital cameras and IMU sensors to each other.

Contribution. We propose two approaches to correct the tilt of the camera: (i) the first one estimates the ground plane normal assuming dominant ground plane and recovers it to its original alignment, (ii) while the second approach corrects the tilt by estimating the rotation between the two views using far feature points. To the best of our knowledge, these are the first pure visual methods to correct the camera orientation. As a side effect, the algorithms can also be applied to detect non-planar regions within a video sequence taken by a vehicle-mounted camera.

2 PROPOSED METHODS

The camera tilt refers to the orientation of the camera about the x and z-axes, which often are denoted as **pitch** and **roll** angles, respectively. In the case of planar motion, it is assumed no pitch or roll rotation is feasible, the vehicle can rotate only about the y-axis, often is denoted as **yaw** rotation. When the planar motion constraint is violated the camera tilt correction should be considered to compensate the planar motion violation. Fig. 1 shows the camera and image coordinate systems.

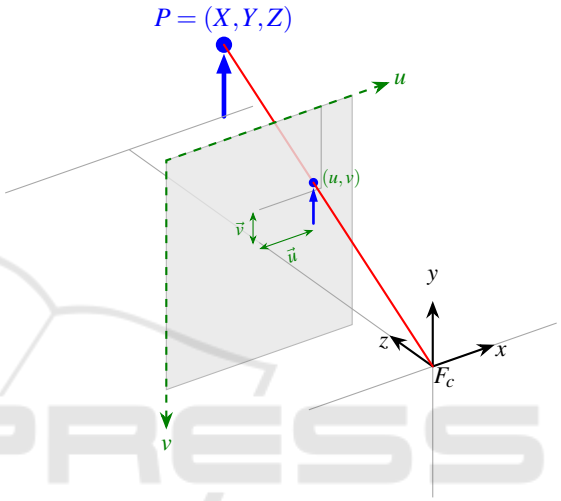


Figure 1: The camera and image coordinate systems.

We propose two different approaches to solve the camera tilt correction problem.

2.1 Tilt Correction using Ground Plane Normal

In a camera coordinates system, assuming a frontal facing camera mounted on top of a moving vehicle, the ground plane normal has to be parallel to y-axis. The nonperfect mounting or nonflat ground can break the ground plane and y-axis alignment. The proposed method estimates the ground plane normal and find the compensation rotation matrix which recovers it to the correct alignment.

2.1.1 Ground Plane Homography

Assuming dominant ground plane, the ground plane homography matrix is estimated after establishing the corresponding features between two views of the scene. For this reason, two consecutive camera images of a moving agent are selected.

The feature points located out of the ground are filtered out by considering only the points located in the lower-half of the image.

The homography general equation is given as follows:

$$\mathrm{q_i} \sim \mathbf{H}\mathrm{p_i} \tag{1}$$

where operator $\sim$ denotes equality up to a scale, $\mathrm{p_i} = \begin{bmatrix} p_{x_i} & p_{y_i} & 1 \end{bmatrix}^T$, $\mathrm{q_i} = \begin{bmatrix} q_{x_i} & q_{y_i} & 1 \end{bmatrix}^T$ are the normalized homogeneous image coordinates of the two views which can be recovered by multiplying the image coordinates with the inverse of the camera matrix, and $\mathbf{H}$ is the homography matrix.

Homography $\mathbf{H}$ is a 3×3 matrix but it has 8 DoFs since it is estimated up to scale. Four point correspondences are necessary to find the homography parameters as each correspondence gives two equations.

The popular 4-points method is used to estimate the homography matrix. RANSAC (Fischler and Bolles, 1981) robust statistics are used to robustly estimate the homograph matrix and find the homography that maximizes the number of inliers, in case of the presence of outliers.

2.1.2 Ground Plane Normal

The ground plane normal can be recovered by decomposing the homography matrix related to the ground plane. The homography matrix can be formulated as follows:

$$\mathbf{H} = \mathbf{R}_h + t_h n^T \tag{2}$$

where $\mathbf{R}_h$, t_h are the ground plane transformation components (rotation and translation) from the first to the second view, and n is the ground plane normal.

The first analytical method of Malis and Vargas for homography decomposition (Malis and Vargas, 2007) is used to decompose the homography matrix which returns up to 4 solutions. The solution has the normal vector where y-coordinate is positive and the one closest to y-axis is selected, which retrieved by:

$$n = \arg\min_{n_i} \left\| n_i \times \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \right\|$$

n_i is a normalized normal, thus $\|n_i\| = 1$.

2.1.3 Correct Ground Normal

The rotation matrix which corrects the tilt of the camera is the rotation matrix which rotates the ground plane normal to be parallel to y-axis. Therefore,

$$\mathbf{R}_{tc} n = \begin{bmatrix} 0 \\ v \\ 0 \end{bmatrix},$$

where $\mathbf{R}_{tc}$ is the correction matrix, n is the ground plane normal, and v is a positive real number. There are infinite rotation matrices satisfying the equation, however, the correction matrix should not change the vertical direction, thus the correction matrix includes only rotation around x and/or z-axes.

The correction matrix can be formulated as follows:

$$\mathbf{R}_{tc} = \mathbf{R_z}\mathbf{R_x} \tag{3}$$

where $\mathbf{R_z}$ and $\mathbf{R_x}$ represent the rotation about z and x-axes, respectively.

Let us start with the former rotation:

$$\begin{aligned} \mathbf{R_x} n &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = n' \\ &= \begin{bmatrix} n_x \\ n_y\cos(\alpha) - n_z\sin(\alpha) \\ n_y\sin(\alpha) + n_z\cos(\alpha) \end{bmatrix} = \begin{bmatrix} n'_x \\ n'_y \\ 0 \end{bmatrix}, \end{aligned}$$

where α is the rotation angle about the x-axis. After rotation, the last coordinate must be zero. Therefore,

$$\tan(\alpha) = -\frac{n_z}{n_y}.$$

For angle α, there are two solutions:

$$\begin{aligned} \alpha_1 &= \arctan(-\frac{n_z}{n_y}), \\ \alpha_2 &= \arctan(-\frac{n_z}{n_y}) + \pi, \end{aligned}$$

where $\alpha_1, \alpha_2 \in [0, 2\pi[$.

The rotation about the z-axis transforms the modified normals again. The first coordinate must be zero:

$$\mathbf{R_z} n' = \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0 \\ \sin(\beta) & \cos(\beta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} n'_x \\ n'_y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ v \\ 0 \end{bmatrix}$$

From the first coordinate, and the expression $n'_y = n_y\cos(\alpha) - n_z\sin(\alpha)$, the following formula is obtained:

$$\tan(\beta) = \frac{n_x}{n_y\cos(\alpha) - n_z\sin(\alpha)}$$

For angle β, there are 4 solutions:

$$\begin{aligned} \beta_1 &= \arctan(\frac{n_x}{n_y\cos(\alpha_1) - n_z\sin(\alpha_1)}), \\ \beta_2 &= \arctan(\frac{n_x}{n_y\cos(\alpha_1) - n_z\sin(\alpha_1)}) + \pi, \\ \beta_3 &= \arctan(\frac{n_x}{n_y\cos(\alpha_2) - n_z\sin(\alpha_2)}), \\ \beta_4 &= \arctan(\frac{n_x}{n_y\cos(\alpha_2) - n_z\sin(\alpha_2)}) + \pi, \end{aligned}$$

where $\beta_1, \beta_2, \beta_3, \beta_4 \in [0, 2\pi[$.

Thus, four solutions for $\mathbf{R}_{tc}$ are in total. However, only one solution can be used for tilt correction, which returns $v > 0$, and $\alpha, \beta \in [0, \pi/2] \cup [\frac{3}{2}\pi, 2\pi[$.

From Eq. 3 the correction is given as:

$$\mathbf{R}_{tc} = \mathbf{R_z}\mathbf{R_x} = \begin{bmatrix} \cos\beta & -\sin\beta\cos\alpha & \sin\beta\sin\alpha \\ \sin\beta & \cos\beta\cos\alpha & -\cos\beta\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

The point located in the first view is corrected by multiplying the correction matrix $\mathbf{R}_{tc}$ with the homogeneous point coordinates x_i:

$$x_i' \sim \mathbf{R}_{tc} x_i. \tag{4}$$

2.2 Tilt Correction using Vanishing Points

In this section, we propose another method to compute the correction of the camera images. The goal is the same, the vertical direction in the camera images should be perpendicular to the gravity vector.

Based on the camera pinhole model, the spatial point projection to the image plane formula is written as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

where u, v are the point image coordinates, x, y, z are the spatial point coordinates, f_x, f_y are the focal lengths scaled by the pixel size on x and y-axes, and p_x, p_y are the principle point coordinates. Applying the homogenous division to eliminate the similarity operator gives:

$$u = p_x + \frac{x f_x}{z}, \quad v = p_y + \frac{y f_y}{z}$$

Translating the camera with the vector $\begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$, and projecting the spatial point to the camera image yields:

$$u' = p_x + \frac{(x+t_x) f_x}{z+t_z}, \quad v' = p_y + \frac{(y+t_y) f_y}{z+t_z}$$

Assuming very far point ($z \gg x, y, t_x, t_y, t_z$), we can write:

$$\frac{(x+t_x) f_x}{z+t_z} \approx \frac{x f_x}{z}, \quad \frac{(y+t_y) f_y}{z+t_z} \approx \frac{y f_y}{z},$$

which yields that $u \approx u', v \approx v'$. *We can conclude that the very far points are not affected by the translation of the camera, such points are called vanishing points.*

Unlike the ground plane normal correction approach which corrects the tilt globally, the vanishing points approach corrects the tilt relatively between two views by estimating the rotation matrix considering the vanishing points.

2.2.1 Rotation Matrix Estimation

From the general homography formula in Eq. 1 and its decomposition in Eq. 2, the relation between the corresponding points of the two views is as follows:

$$\mathrm{q_i} \sim (\mathbf{R}_h + t_h n^T)\mathrm{p_i},$$

where $\mathrm{p_i} = \begin{bmatrix} p_{u_i} & p_{v_i} & 1 \end{bmatrix}^T$ and $\mathrm{q_i} = \begin{bmatrix} q_{u_i} & q_{v_i} & 1 \end{bmatrix}^T$ are the normalized homogeneous coordinates in the first and second views, respectively. Considering only the vanishing points of the scene, one can eliminate the translation term as

$$\mathrm{q_i} \sim \mathbf{R}_h \mathrm{p_i}.$$

The previous equation can be rewritten as follows:

$$\mathrm{q_i} = \lambda \mathbf{R}_h \mathrm{p_i},$$

where λ is a positive real scalar value, thus $\lambda > 0$. If several points are given, the estimation of the parameters can be reformulated as a least squares minimization:

$$\arg\min_{\lambda, \mathbf{R}_h} J(\lambda, \mathbf{R}_h)$$

It is a so-called point registration problem with rotation and scaling transformations. Here we show how this problem can be optimally solved in the least-squares sense. The proof here is based on the algorithm of Arun et al. (Arun et al., 1987).

The function to be minimized is as follows:

$$\begin{aligned} J(\lambda, \mathbf{R}_h) &= \sum_{i=1}^{N} \|\lambda \mathbf{R}_h \mathrm{p_i} - \mathrm{q_i}\| \\ &= \sum_{i=1}^{N} (\lambda \mathbf{R}_h \mathrm{p_i} - \mathrm{q_i})^T (\lambda \mathbf{R}_h \mathrm{p_i} - \mathrm{q_i}) \\ &= \lambda^2 \sum_{i=1}^{N} \mathrm{p_i}^T \mathrm{p_i} + \sum_{i=1}^{N} \mathrm{q_i}^T \mathrm{q_i}^T - 2\lambda \sum_{i=1}^{N} \mathrm{q_i}^T \mathbf{R}_h \mathrm{p_i}. \end{aligned}$$

Since $\lambda > 0$, minimize the cost function J with respect to rotation $\mathbf{R}_h$ is equivalent to maximize the term $\sum_{i=1}^{N} \mathrm{q_i}^T \mathbf{R}_h \mathrm{p_i}$. The former term can be rewritten as:

$$\sum_{i=1}^{N} \mathrm{q_i}^T \mathbf{R}_h \mathrm{p_i} = trace(\mathbf{Q}\mathbf{R}_h\mathbf{P}) = trace(\mathbf{R}_h\mathbf{P}\mathbf{Q}),$$

where:

$$\mathbf{Q} = \begin{bmatrix} \mathrm{q_1}^T \\ \mathrm{q_2}^T \\ \vdots \\ \mathrm{q_N}^T \end{bmatrix}, \mathbf{P} = \begin{bmatrix} \mathrm{p_1} & \mathrm{p_2} & \cdots & \mathrm{p_N} \end{bmatrix}.$$

Applying singular value decomposition for the product matrix $\mathbf{PQ}$ gives:

$$trace(\mathbf{R}_h\mathbf{P}\mathbf{Q}) = trace(\mathbf{R}_h\mathbf{U}\mathbf{S}\mathbf{V}^T).$$

$\mathbf{R}_h\mathbf{U}$ and $\mathbf{V}^T$ are orthonormal matrices, and $\mathbf{S}$ is a diagonal one. Thus, they can be written as:

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix},$$
$$\mathbf{R}_h\mathbf{U} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}.$$

Then

$$trace(\mathbf{R}_h\mathbf{U}\mathbf{S}\mathbf{V}^T) = trace(\mathbf{S}\mathbf{V}^T\mathbf{R}_h\mathbf{U}) = \sum_{i=1}^{3} s_i a_i^T b_i.$$

The singular values are always positive: $s_i > 0$, thus maximizing $\sum_{i=1}^{3} s_i a_i^T b_i$ is equivalent to maximize $\sum_{i=1}^{3} a_i^T b_i$.

a_i and b_i are vectors of orthonormal matrices, the maximum value of their dot products equals to one, and it can only be possible when $a_i = b_i$. The three vectors are the same if:

$$\mathbf{R}_h\mathbf{U} = \mathbf{V}.$$

The rotation that minimizes the cost function J is:

$$\mathbf{R}_h = \mathbf{V}\mathbf{U}^T. \quad (5)$$

At least three points are necessary to estimate the rotation since 3-DoFs is possible. RANSAC method (Fischler and Bolles, 1981) is used to robustly estimate the rotation and reject outliers. The proposed method estimates the rotation considering only the vanishing points, thus the outlier set after RANSAC includes all scene points except the vanishing points.

2.2.2 Tilt Correction

Tilt correction is performed by decomposing the rotation matrix into the elemental rotations, that are the rotations around the axes of the coordinate system. Then the correction rotation matrix is constructed by rotation around x and z-axes. There are an infinite number of rotation sequences equivalent to the rotation matrix, however rotation sequence where rotation about y-axis is the last step is considered.

Euler angles method with $y-x-z$ convention (Tait-Bryan conventions) is used to decompose the rotation matrix to the elemental rotations angles, the sequence starts with rotation about z-axis, followed by rotation about x-axis, and finally rotation about the y-axis: $\mathbf{R}_h = \mathbf{R}_y\mathbf{R}_x\mathbf{R}_z$. Let us denote the rotation matrix:

$$\mathbf{R}_h = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}.$$

The euler angles α, β, γ are given by:

$$\begin{aligned} \alpha &= atan2\left(r_{13}, r_{33}\right), \\ \beta &= atan2\left(-r_{23}, \sqrt{1-r_{23}^2}\right), \\ \gamma &= atan2\left(r_{21}, r_{22}\right), \end{aligned} \quad (6)$$

where α, β, γ are the rotation angles around the y, x, and z-axes, respectively. $atan2(y,x)$ is the two arguments arctan function which returns the angle between the positive horizontal axis and ray $\begin{bmatrix} x & y \end{bmatrix}$. The correction matrix is $\mathbf{R}_{tc} = \mathbf{R}_x\mathbf{R}_z$, where

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\beta) & -\sin(\beta) \\ 0 & \sin(\beta) & \cos(\beta) \end{bmatrix},$$
$$\mathbf{R}_z = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3 EXPERIMENTAL RESULTS

The proposed approaches are tested on synthetic as well as real data. The latter ones are recorded using a monocular camera mounted on top of a car during a trip in the roads of our city.

3.1 Experiments on Synthetic Data

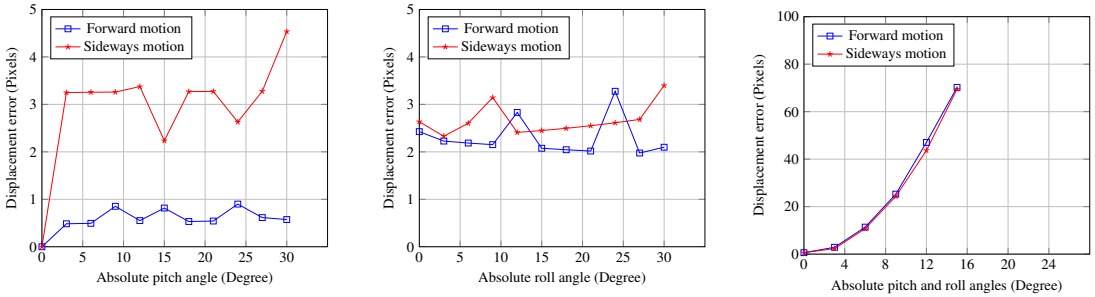
The two approaches are evaluated using two different camera setups, (i) Forward motion: the second camera is given by translating the first camera along the z-axis, (ii) Sideways motion: the second camera is given by rotating the first camera about the y-axis 15°, and translating along the z-axis. The focal length is set to $f = 1000$, translation magnitude between the two cameras is 100.

Ground plane is perpendicular and touches the bottom border of the first camera image. The generated corresponding points are set to lie on the ground plane: (i) 20% very close points where points' depth $d < 1000$, (ii) 70% where $3000 < d < 4000$, and (iii) 10% are very far points $50,000 < d < 100,000$.

To test our approaches, rotation about x-axis (pitch angle), z-axis (roll angle), and both about x and z-axes are applied on the first camera image, and the two approaches are used to recover the first camera to its original orientation. The average displacement error is proposed to evaluate the two methods:

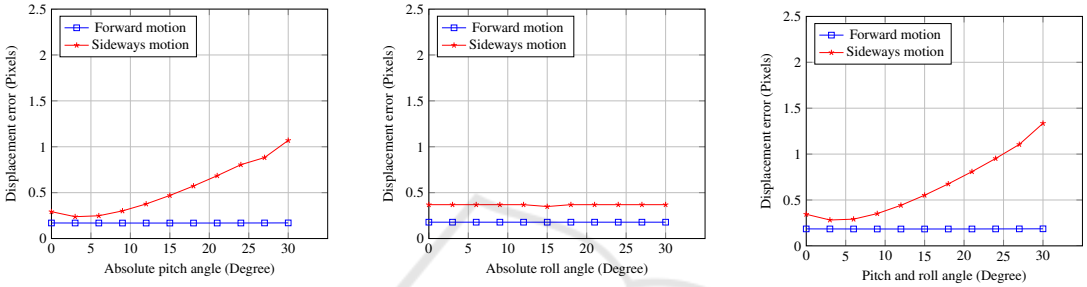
$$err = \frac{1}{n}\sum_{i=1}^{n} \left\| u_{\text{original}_i} - u_{\text{corrected}_i} \right\|$$

where n is the number of the feature points, u_{original_i} is the original image coordinates of the point i, and



(a) (b) (c)

Figure 2: The displacement error of the ground plane normal approach w.r.t the applied rotation, (a): pitch rotation is applied, (b): roll rotation is applied, and (c): both pitch and roll rotations are applied.



(a) (b) (c)

Figure 3: The displacement error of the vanishing points approach w.r.t the applied rotation, (a): pitch rotation is applied, (b): roll rotation is applied, and (c): both pitch and roll rotations are applied.

$u_{\text{corrected}_i}$ is the image coordinates of the point i in the corrected image using our approaches.

Fig. 2 shows the average displacement error in pixels between the original and corrected image using the ground plane normal approach with respect to the rotation applied. The ground plane normal approach returns the camera to its correct orientation with good precision when only a pitch or a roll rotation is applied. It still returns good precision when both pitch and roll rotations contaminate the camera orientation with small angles $< 5°$.

Fig. 3 shows the displacement error of the vanishing points method with respect to the rotation applied. The vanishing points approach recovers the camera to its correct orientation with very accurate precision during the forward motion, it also works efficiently during the sideways motion with error < 1.5 pixels.

3.2 Experiments with Real Data

Tilt correction on real data is evaluated qualitatively, a set of interesting situations are chosen where the planar motion constraint is broken.

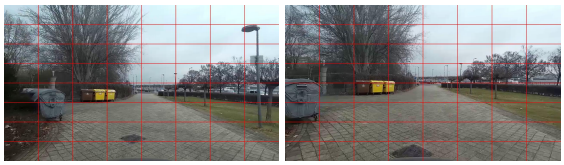
For evaluating the ground plane normal correction method, three consecutive camera images are selected in each interesting situation, the corresponding points located on the ground plane are established manually between the first and second images, and between the second and third images. Remark that the third image is required for correcting the second one as the correcting homographies are not the same for different images. The correction matrices for the first and second images are computed from first-second and second-third pairs, respectively.

For each pair the correction matrix $\mathbf{R}_{tc}$ is estimated. The homography matrix for unnormalized coordinates is $\mathbf{H} = \mathbf{K}\mathbf{R}_{tc}\mathbf{K}^{-1}$, where $\mathbf{K}$ is the intrinsic camera parameters matrix. The homography matrix $\mathbf{H}$ is used to transform the image and retrieve the corrected camera image.

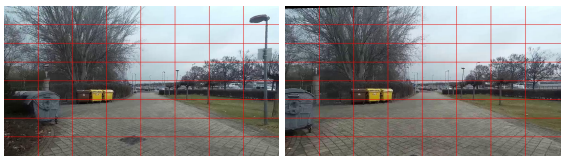
Figures 4, 5, 6 show the original and the cropped corrected camera images using the ground plane normal approach for different situations.

Vanishing points method starts with establishing the corresponding feature points between two consecutive camera images using SURF method (Bay et al., 2006), RANSAC is used with 2 pixels threshold and fixed 100 iterations to robustly estimate the rotation matrix. Unnormalized homography matrix is calculated $\mathbf{H} = \mathbf{K}\mathbf{R}_{tc}\mathbf{K}^{-1}$, and the first image is corrected accordingly.

Figures 7, 8, 9 show the original and corrected image pairs for different situation. The corresponding inlier points are visualized which represent the vanishing points.

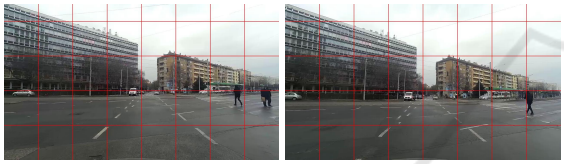


(a) First camera image. (b) Second camera image.

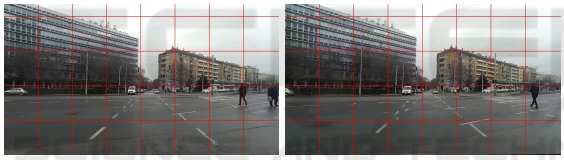


(c) Corrected first image. (d) Corrected second image.

Figure 4: Tilt correction using ground plane normal method, the ground plane is non-flatten causing rotation about x-axis. The two original images in the first row show the vertical displacement caused by the non-flatten ground plane, the second row shows how the displacement is corrected.

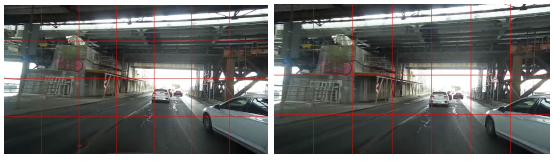


(a) First camera image. (b) Second camera image.



(c) Corrected first image. (d) Corrected second image.

Figure 5: Tilt correction using ground plane normal method, the camera is shaking due to non-perfect camera mounting. The first row shows the horizon line is displaced, the second row shows the corrected camera images.

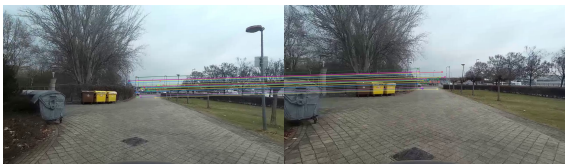


(a) Original camera image. (b) Corrected camera image.

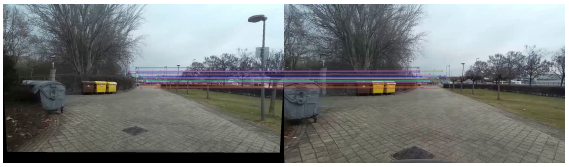
Figure 6: Tilt correction using ground plane normal method, the car is turning right causing camera shaking. The horizontal lines are sloped in the original image (a), the horizontal lines are corrected in (b).

4 CONCLUSIONS AND FUTURE WORK

We have addressed visual tilt correction for vehicle-mounted cameras in this paper. For algorithms assuming planar motion, it is essential that the image plane should be perpendicular to the road surface. Two algorithms have been proposed here. The first one ap-

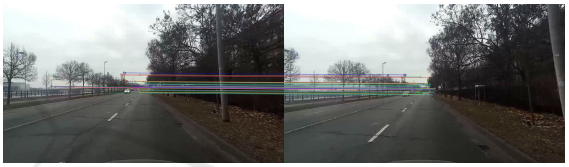


(a) Original camera image pair.



(b) Corrected image pair.

Figure 7: Tilt correction using vanishing points method when ground plane is not-flatten causing rotation about x-axis. The corresponding vanishing points are visualized in the original and corrected pairs.



(a) Original camera image pair.

(b) Corrected image pair.

Figure 8: Tilt correction using vanishing points method, the car moving with high speed causing shaking of the camera.

(a) Original camera image pair.

(b) Corrected image pair.

Figure 9: Tilt correction using vanishing points method, camera shaking situation.

plies the ground plane as its normal coincides with the gravity vector. The second algorithm filters the vanishing points close to the horizon and computes the correction homography.

The accuracy of the algorithms are quantitatively compared with the synthesized data. It is also demonstrated that they are applicable for images of real vehicle-mounted cameras.

Future Work. The proposed methods have been applied in this paper only for tilt correction. However, the estimated vertical directions can be used for other purposes. If the road is non-totally planar, it can be detected. IMU sensors can measure the vertical direction via gravity force, thus the proposed methods can help in camera-IMU calibration. LiDAR-camera calibration is also possible as the ground plane can be estimated from LiDAR point clouds as well. These are our research directions for the future.

ACKNOWLEDGEMENTS

Our work is supported by the project EFOP-3.6.3-VEKOP-16-2017- 00001: Talent Management in Autonomous Vehicle Control Technologies, by the Hungarian Government and co-financed by the European Social Fund. L. Hajder also thanks the support of the "Application Domain Specific Highly Reliable IT Solutions" project that has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the Thematic Excellence Programme TKP2020-NKA-06 (National Challenges Subprogramme) funding scheme.

REFERENCES

Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700.

B. Triggs and P. McLauchlan and R. I. Hartley and A. Fitzgibbon (2000). Bundle Adjustment – A Modern Synthesis. In Triggs, W., Zisserman, A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Computer Vision – ECCV 2006*, pages 404–417, Berlin, Heidelberg. Springer Berlin Heidelberg.

Choi, S. and Kim, J.-H. (2018). Fast and reliable minimal relative pose estimation under planar motion. *Image and Vision Computing*, 69:103–112.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. 24(6):381–395.

Guan, B., Su, A., Li, Z., and Fraundorfer, F. (2019). Rotational alignment of imu-camera systems with 1-point RANSAC. In *Pattern Recognition and Computer Vision - Second Chinese Conference, PRCV 2019, Xi'an, China, November 8-11, 2019, Proceedings, Part III*, pages 172–183.

Hajder, L. and Barath, D. (2020a). Least-squares optimal relative planar motion for vehicle-mounted cameras. In *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*, pages 8644–8650.

Hajder, L. and Barath, D. (2020b). Relative planar motion for vehicle-mounted cameras from a single affine correspondence. In *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*, pages 8651–8657.

Hartley, R. (1997). In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593.

Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Kukelova, Z., Bujnak, M., and Pajdla, T. (2008). Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Proceedings of the British Machine Vision Conference 2008, Leeds, UK, September 2008*, pages 1–10.

Li, H. and Hartley, R. I. (2006). Five-point motion estimation made easy. In *18th International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong, China*, pages 630–633.

Malis, E. and Vargas, M. (2007). *Deeper understanding of the homography decomposition for vision-based control*. PhD thesis, INRIA.

Nister, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770.

Ortin, D. and Montiel, J. M. M. (2001). Indoor robot motion based on monocular images. *Robotica*, 19(3):331–342.

Saurer, O., Vasseur, P., Boutteau, R., Demonceaux, C., Pollefeys, M., and Fraundorfer, F. (2017). Homography based egomotion estimation with a common direction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(2):327–341.

Stachniss, C., Leonard, J. J., and Thrun, S. (2016). Simultaneous localization and mapping. In *Springer Handbook of Robotics*, pages 1153–1176.

Weiss, S. and Siegwart, R. (2011). Real-time metric state estimation for modular vision-inertial systems. In *IEEE International Conference on Robotics and Automation, ICRA 2011, Shanghai, China, 9-13 May 2011*, pages 4531–4537.