Rolling Shutter Correction in Manhattan World

Pulak Purkait
Toshiba Research Europe
Cambridge, UK
pulak.isi@gmail.com

Christopher Zach
Toshiba Research Europe
Cambridge, UK
christopher.m.zach@gmail.com

Ales Leonardis
University of Birmingham
Birmingham, UK
a.leonardis@cs.bham.ac.uk

Abstract

A vast majority of consumer cameras operate the rolling shutter mechanism, which often produces distorted images due to inter-row delay while capturing an image. Recent methods for monocular rolling shutter compensation utilize blur kernel, straightness of line segments, as well as angle and length preservation. However, they do not incorporate scene geometry explicitly for rolling shutter correction, therefore, information about the 3D scene geometry is often distorted by the correction process. In this paper we propose a novel method which leverages geometric properties of the scene—in particular vanishing directions—to estimate the camera motion during rolling shutter exposure from a single distorted image. The proposed method jointly estimates the orthogonal vanishing directions and the rolling shutter camera motion. We performed extensive experiments on synthetic and real datasets which demonstrate the benefits of our approach both in terms of qualitative and quantitative results (in terms of a geometric structure fitting) as well as with respect to computation time.

1. Introduction

People largely share knowledge and experiences through visual photographs, often captured by low-budget commercial devices. These devices are generally built upon CMOS sensors, which possess a prevalent mechanism widely known as *rolling shutter* (RS). In contrast to *global shutter* (GS), it captures the scene in a row-wise manner from top to bottom with a constant inter-row delay. The RS imaging acquires apparent camera motion for different rows and violates the properties of the perspective camera model. This causes noticeable distortions—straight line segments can become arc segments, which are very prominent for the images in urban areas. This distortion needs to be corrected for aesthetically pleasing visualization and further geometric analysis [14] of the scene.

In this work, we address the RS compensation from a *single* distorted image. This problem has been addressed in recent methods [30, 27], however, no scene geometry was incorporated utilizing only a single image while compensating the RS effects. We observe that most of the images taken in man-made environments (such as urban areas) feature at least two orthogonal vanishing directions. Consequently, we believe that the Manhattan world assumption is satisfied especially when the rolling shutter effect is most prominent in images. In this work, we propose an RS correction method utilizing these orthogonal vanishing directions, therefore the corrected image without RS distortions is not only visually more appealing, but also geometrically more meaningful. Our proposed method demonstrates better performance qualitatively and computationally. We also evaluate proposed method quantitatively by fitting a geometric structure (*e.g.*, rotational homography, epipolar geometry [14]). In Figure 1, we display our result on a real RS distorted image. Notice that this example is not of a typical urban image and proposed method still produces more geometrically consistent results than the baseline.

(a) A distorted image (b) Result by [27] (c) Our Result

Figure 1: (a) A real rolling shutter distorted image. (b) Rectified by Rengarajan *et al*. [27]. (c) Proposed joint estimation of orthogonal vanishing directions and rolling shutter motion. The colors *red*, *green* and *blue* are employed for the orthogonal vanishing directions, while *yellow* is used to mark the outliers (lines that are not associated with the vanishing directions). Sign-post and roads are more geometrically consistent by the proposed method.

1.1. Related Work

Recent works on RS compensation can be grouped into three categories—(i) external sensors based methods, (ii)

multi-frame methods and (iii) single-frame methods.

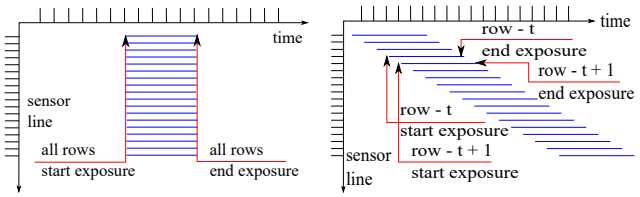
(i) *External sensors (*e.g.*, gyroscopes)* have been utilized [15, 17, 26] to acquire camera motion directly in videos. However, the low acquisition rate does not allow performing RS correction for a single image.

(ii) *Multi-frame methods* study the geometry of an RS camera [3, 8, 10, 28], utilizing multiple RS images or video sequences. A number of interest points are tracked over the frames and then those tracked points are utilized to estimate the camera motion. The camera poses for the other rows are then interpolated in order to correct the RS effect. Grundmann *et al*. [13] utilize a mixture of homographics, estimated from the tracked key points, to compensate rolling shutter effect. None of these methods can directly be applied to Single-frame RS correction.

(iii) *Single-frame* RS correction from a single image, without the help of external sensors, goes back to [2], but [27, 30] are most related to our approach. Su *et al*. [30] propose to utilize motion blur to extract information about the camera motion. They employ a global model of the camera motion trajectory, whose parameters are estimated from the blur kernel. Rengarajan *et al*. [27] detect line segments (LSs) and then group them into the horizontal and vertical arc segments. Straightness of the detected arcs, line length constancy and line angle constancy are incorporated to estimate the motion. However, the works [27, 30] suffer from a number of drawbacks:

- Primarily, no scene geometry is incorporated in [27, 30] for RS compensation, but obtaining correct geometric relations is the primary objective in the first place.
- The method presented in [30] is only applicable for blurred images. Although, in some cases motion blur and RS distortions occur simultaneously, these are very different phenomenon and can appear exclusively.
- Bending of straight lines is not guaranteed for every RS camera motion, *e.g*. if the camera motion only leads to (anisotropic) scaling of image content. In such cases the method of [27] cannot be used to rectify RS distortions.
- The work of [27] assumes that all arc segments (including natural curves) are induced by straight lines and take place in camera motion estimation. Thus, the estimates may be distorted if this assumption is violated.

In this work, we utilize the underlying scene geometry, which we assume is mostly generated by a Manhattan-type world. Orthogonal vanishing directions and the camera motion of an RS image are jointly estimated via an appropriate cost function. While we do not explicitly utilize the straightness property of line segments (as it is done in [27]), our estimated motion parameters are nevertheless sufficiently accurate to obtain straight lines in the generated GS image. Moreover, our method is free from the aforementioned drawbacks. Our contributions are summarized as follows:



(a) Global shutter exposure (b) Rolling shutter exposure

Figure 2: (a) A global shutter opens to allow light to strike the entire sensor surface all at once. (b) In contrast, a rolling shutter exposes the image line-by-line.

- We utilize those parts of the 3D scene geometry captured in an RS image conforming to the Manhattan-world assumption (MWA), and we formulate a robust objective to simultaneously estimate the underlying vanishing directions and camera motion parameters.
- Extensive experiments show that the proposed approach is computationally efficient and qualitatively more accurate than earlier works. The joint optimization for all parameters is done in a fraction of a second, which is about two orders of magnitude faster than the baselines [27, 30].

This paper is organized as follows. In Sections 2 & 3, we provide a brief introduction to the RS camera and estimation of the vanishing directions, which is extended in Section 4 for the joint estimation. The efficiency of the proposed method is presented in Section 5. We conclude and indicate future extensions in Section 6.

2. Rolling Shutter Cameras

Global shutter and rolling shutter cameras differ in how light incoming at the imaging sensor is gathered. In Figure 2, we display the image capture process with different sensors. In the case of GS camera all the rows of the image sensor are exposed simultaneously for a constant duration of time. A point $\mathbf{P} \in \mathbb{R}^3$ in the scene, observed at the pixel $(p,\ q)$ in the GS camera, satisfies [14]

$$s\mathbf{p} = K\mathbf{P}, \tag{1}$$

where $\mathbf{p} = [p,\ q,\ 1]^\intercal$ is the homogeneous coordinate of the pixel $(p,\ q)$, s is the scene depth, and K is the intrinsic camera matrix.

In the case of RS camera, sensors in each of the rows are exposed for a regular interval of time (same exposure and integration time), while the camera potentially undergoes an (small) amount of motion. The translation of the camera, during capturing different rows of an image, is assumed to be negligible compared to the depth of the scene. Thus, the

projection of the point $\mathbf{P}$ onto RS camera reads as

$$s\mathbf{p}^{rs} = KR(\mathbf{r}^t)\mathbf{P}, \tag{2}$$

where $R(\mathbf{r}^t)$ is the rotation matrix corresponding to the rotation $\mathbf{r}^t$ at time $t = \tau p^{rs}$, where τ is the time delay between two successive rows. The geometric relation between the GS pixels and RS pixels (eliminating $\mathbf{P}$ from 1 and 2) is therefore given by

$$\mathbf{p} \propto KR(\mathbf{r}^t)^\intercal K^{-1}\mathbf{p}^{rs}, \tag{3}$$

where rotation $R(\mathbf{r}^t)$in above depends on the p^{rs}th row of the RS image. Note that the above relation holds up to a scale. For readability, in the rest of the paper, we consider $\mathbf{p}^{rs}$ is on the image plane, *i.e.*, pre-multiplied by K^{-1}, thus,

$$\mathbf{p} = KR(\mathbf{r}^t)^\intercal \mathbf{p}^{rs}. \tag{4}$$

2.1. Motion Modelling

Independent estimation of camera poses for each of the rows of an RS camera is extremely ill-posed. Therefore, similar to [27, 30], we utilize a global parametric motion model, where the rotation parameters are considered to be polynomials in time t. However, as the RS camera takes uniform time τ to capture a row, rotations in turn become polynomials in row number p. More explicitly, for $\zeta = (p-1)/M$,

$$\begin{cases} r_x = \alpha + a_1\zeta + \ldots + a_n\zeta^n \\ r_y = \beta + b_1\zeta + \ldots + b_n\zeta^n \\ r_z = \gamma + c_1\zeta + \ldots + c_n\zeta^n \,, \end{cases} \tag{5}$$

where M is the number of rows in the image, $\mathbf{r}_\mathcal{A} = [r_x,\ r_y,\ r_z]^\intercal$ are the Rodrigues parameterization [25] of the rotation, and we use the Cayley transform [11] to obtain the corresponding rotations matrix

$$R(\mathbf{r}^t_\mathcal{A}) = \frac{1}{Z}\begin{bmatrix} 1+r_x^2-r_y^2-r_z^2 & 2r_xr_y-2r_z & 2r_y+2r_xr_z \\ 2r_z+2r_xr_y & 1-r_x^2+r_y^2-r_z^2 & 2r_yr_z-2r_x \\ 2r_xr_z-2r_y & 2r_x+2r_yr_z & 1-r_x^2-r_y^2+r_z^2 \end{bmatrix} \tag{6}$$

where $Z = 1+r_x^2+r_y^2+r_z^2$. This transformation is chosen due to its numerical simplicity [3]. Note that $\mathbf{r}_\mathcal{A}$ is the unit axis of rotation scaled by $\tan(\frac{\theta}{2})$ where θ is the angle of rotation. Thus, $180°$ rotations, hardly relevant to the rolling shutter case, are automatically excluded. Moreover, under the choice of ζ, there will only be a global rotation $[\alpha, \beta, \gamma]^\intercal$ at the first row. In summary, estimation of the RS motion is equivalent to the estimation of $3(n+1)$ motion parameters $\mathcal{A} = ([\alpha,\ a_1,\ \ldots,\ a_n;\ \beta,\ b_1,\ \ldots,\ b_n;\ \gamma,\ c_1,\ \ldots,\ c_n])$. Note that quartic splines may provide a better fit for a more complex motion [18, 26]. However, the polynomial model (5) is expressive enough to capture natural camera motions. The choice of polynomial motion model is justified further in [27, 30].

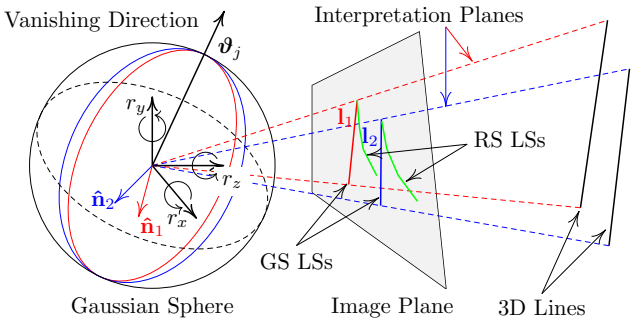


Figure 3: The 3D parallel lines in the world space are projected into the concurrent LSs (green) on a GS camera and arc segments (red) for RS camera.

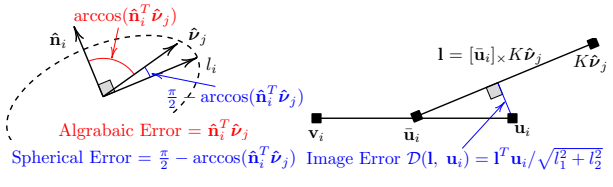
3. Vanishing Directions

The geometry of man-made structures in urban areas has been exploited in a number of works [29, 33]. This geometry possesses predominant linear structures and orthogonal vanishing directions [6, 24]. In this section, we formulate different cost functions for the vanishing directions.

Parallel lines in a 3D scene become concurrent lines, once they are projected onto an image plane. The point of intersection is known as a vanishing point. Most work on vanishing point estimation is carried out on the Gaussian sphere [1, 6, 19, 21, 24, 36], which is a unit sphere in 3D centred at the camera centre. An *interpretation plane*, composed of a single line segment (LS) and the centre of projection [20], crosses over the Gaussian sphere in which a great circle is formed. A vanishing direction (VD) is the intersection of the interpretation planes, *i.e.*, a VD is perpendicular to the normals of the interpretation planes, passes through the intersection of the great circles and points towards a vanishing point in the image plane (Figure 3).

Antunes *et al.* [4] exploited the Facility Location problem, and Bazin *et al.* [6] proposed a branch and bound method, to maximize the number of LSs globally which is consistent with the orthogonal VDs. Tardif [31] exploited J-linkage (a variant of RANSAC) [32] for clustering the LSs. There are also methods for simultaneous tracking and estimation of the VDs in a video [20, 21]. A CNN based approach [36] is exploited to learn the prior knowledge of the cardinal directions. It is then used to guide the sampling for a randomized estimation method. However, the existing methods do not address the orthogonal VDs estimation for an RS camera that we will formulate in the following section.

In a Manhattan world [7], the VDs are orthogonal and can be represented as a rotated canonical bases [4, 6, 24], $\hat{\mathbf{e}}_x = [1,\ 0,\ 0]^\intercal, \hat{\mathbf{e}}_y = [0,\ 1,\ 0]^\intercal$, and $\hat{\mathbf{e}}_z = [0,\ 0,\ 1]^\intercal$. Let $\boldsymbol{\vartheta} = [\theta,\ \phi,\ \psi]^\intercal$ be the Rodrigues parameterization



(a) Algebraic error (b) Geometric error

Figure 4: Different choices of errors utilized for joint estimation of vanishing directions and camera motions.

of the rotation corresponding to the orthogonal VDs $V = [\hat{\boldsymbol{\nu}}_x,\ \hat{\boldsymbol{\nu}}_y,\ \hat{\boldsymbol{\nu}}_z]$. Then $V = R(\boldsymbol{\vartheta})\mathcal{E}$ where $R(\boldsymbol{\vartheta})$ is the rotation matrix (6) corresponding to $\boldsymbol{\vartheta}$ and $\mathcal{E} = [\hat{\mathbf{e}}_x,\ \hat{\mathbf{e}}_y,\ \hat{\mathbf{e}}_z]$.

There are the following natural cost functions to estimate the (orthogonal) VDs from line segments: the algebraic and the geometric error. The algebraic error enables easy reasoning about intrinsic ambiguities in Section 4.2, but the geometric error is closer to the usual noise assumption of image observations.

Algebraic error One way to estimate orthogonal VDs is to minimize the *algebraic error*, which is the absolute sum of the projections of the normals along VDs. *i.e.*,

$$\arg\min_{\boldsymbol{\vartheta}} \sum_{i=1}^{N} \min_{\hat{\mathbf{e}}\in\mathcal{E}} \rho\Big(\hat{\mathbf{n}}_i^\intercal R(\boldsymbol{\vartheta})\hat{\mathbf{e}}\Big), \tag{7}$$

where $\hat{\mathbf{e}} \in \mathcal{E}$, N is the number of LSs and $\hat{\mathbf{n}}_i$ is the unit vector along the normal of the interpretation plane of the i^{th} LS l_i. $\rho(.)$ is a robust M-estimator (see Section 4.3) which is utilized to estimate VDs under outliers. The normal of the interpretation plane at the camera centre is obtained by taking the cross product of homogeneous pixel co-ordinates of the end points of l_i. *i.e.*,

$$\mathbf{n}_i = K^{-1}\mathbf{u}_i \times K^{-1}\mathbf{v}_i. \tag{8}$$

The unit vector along the normal $\hat{\mathbf{n}}_i = \frac{\mathbf{n}_i}{\|\mathbf{n}_i\|_2}$. A vanishing direction $\hat{\boldsymbol{\nu}}_j$ must pass through the great circle induced by the interpretation plane of a line segment corresponding to $\hat{\boldsymbol{\nu}}_j$. Thus, another cost (spherical error), can be defined by the sum of the angles between $\hat{\boldsymbol{\nu}}_j$ and the associated interpretation plane $\hat{\mathbf{n}}_i$ [Figure 4],

$$\arg\min_{\boldsymbol{\vartheta}} \sum_{i=1}^{N} \left(\frac{\pi}{2} - \arccos\Big(\min_{\hat{\mathbf{e}}\in\mathcal{E}} \rho\Big(\hat{\mathbf{n}}_i^\intercal R(\boldsymbol{\vartheta})\hat{\mathbf{e}}\Big)\Big)\right). \tag{9}$$

The algebraic and spherical error are indeed quite similar and return identical results.

Geometric error Since the usual noise model assumes noisy positions of extracted points on the image plane, the most meaningful cost function uses the geometric error in the image plane: given latent variables for ideal 2D lines passing exactly through the corresponding vanishing point, the (squared) point-line distances of the detected line end points and the ideal line are accumulated. The ideal line is given in closed form by also passing through the midpoint [31], leading to the following objective,

$$\arg\min_{\boldsymbol{\vartheta}} \sum_{i=1}^{N} \min_{\hat{\mathbf{e}}\in\mathcal{E}} \rho\Big(\mathcal{D}([\bar{\mathbf{u}}_i]_\times KR(\boldsymbol{\vartheta})\hat{\mathbf{e}},\ \mathbf{u}_i)\Big), \tag{10}$$

where $\bar{\mathbf{u}}_i = 0.5\mathbf{u}_i + 0.5\mathbf{v}_i$ is the midpoint of l_i, $[\cdot]_\times$ denotes the skew-symmetric cross-product matrix, and the distance of a point $\mathbf{u}$ from a line $\mathbf{l} = [l_1,\ l_2,\ l_3]^\intercal$ is computed as

$$\mathcal{D}(\mathbf{l},\mathbf{u}) = \mathbf{l}^\intercal\mathbf{u}/\sqrt{l_1^2+l_2^2}. \tag{11}$$

The perpendicular distances of the end points $\mathbf{u_i}$ and $\mathbf{v_i}$ from the straight line, joining the midpoint $\bar{\mathbf{u}}_i$ and the vanishing point $K\hat{\boldsymbol{\nu}}_j$, are identical [Figure 4]. Thus, choosing any one of the distances is sufficient, and $\mathcal{D}([\bar{\mathbf{u}}_i]_\times KR(\boldsymbol{\vartheta})\hat{\mathbf{e}},\ \mathbf{v}_i)$ was not included in (10) to symmetrize the cost.

4. Rolling Shutter Correction

Section 3 addresses global shutter cameras, but for rolling shutter images each row is captured with a separate camera pose (4), thus, line segments become arc segments in general. Hence, significant RS distortions will lead to failure in detecting vanishing directions.

4.1. Joint estimation

Through the RS rectification, we aim to have a distortion free GS image from an input of a single distorted RS image. The main difference to the objectives given in Section 3 is, that the image points defining the interpretation plane $\hat{\mathbf{n}}_i$ have to be motion compensated. Thus, jointly estimating RS motion parameters $\mathcal{A}$ and orthogonal VDs $\boldsymbol{\vartheta}$ using an algebraic error amounts to minimizing

$$\arg\min_{\boldsymbol{\vartheta},\mathcal{A}} \sum_{i=1}^{N} \min_{\hat{\mathbf{e}}\in\mathcal{E}} \rho\Big(\hat{\mathbf{n}}_i^\intercal R(\boldsymbol{\vartheta})\hat{\mathbf{e}}\Big), \tag{12}$$

where $R(\boldsymbol{\vartheta})$ is the rotation matrix of $\boldsymbol{\vartheta} = [\theta,\ \phi,\ \psi]^\intercal$. The unit vector $\hat{\mathbf{n}}_i$ is computed as

$$\hat{\mathbf{n}}_i = \frac{(R(\mathbf{r}_\mathcal{A}^u)^\intercal\mathbf{u}_i^{rs}) \times (R(\mathbf{r}_\mathcal{A}^v)^\intercal\mathbf{v}_i^{rs})}{\|(R(\mathbf{r}_\mathcal{A}^u)^\intercal\mathbf{u}_i^{rs}) \times (R(\mathbf{r}_\mathcal{A}^v)^\intercal\mathbf{v}_i^{rs})\|}$$

where $\mathbf{r}_\mathcal{A}^u$ and $\mathbf{r}_\mathcal{A}^v$ are rotation parameters at the rows of $\mathbf{u}_i$ and $\mathbf{v}_i$ (5); $R(\mathbf{r}_\mathcal{A}^u)$ and $R(\mathbf{r}_\mathcal{A}^v)$ are the rotation matrices (6)

corresponding to $\mathbf{r}^u_{\mathcal{A}}$ and $\mathbf{r}^v_{\mathcal{A}}$ respectively. The spherical and geometric errors are given analogously, and we state the geometric error,

$$\arg\min_{\boldsymbol{\vartheta},\mathcal{A}} \sum_{i=1}^{N} \min_{\hat{\mathbf{e}}\in\mathcal{E}} \rho\Big(\mathcal{D}([\bar{\mathbf{u}}_i]_\times K R(\boldsymbol{\vartheta})\hat{\mathbf{e}},\ \mathbf{u}_i)\Big), \tag{13}$$

where $\bar{\mathbf{u}}_i = 0.5KR(\mathbf{r}^u_{\mathcal{A}})^\intercal \mathbf{u}^{rs}_i + 0.5KR(\mathbf{r}^v_{\mathcal{A}})^\intercal \mathbf{v}^{rs}_i$ is the midpoint and $\mathbf{u}_i = KR(\mathbf{r}^u_{\mathcal{A}})^\intercal \mathbf{u}^{rs}_i$ is one of the end points of l_i in GS coordinates.

4.2. Gauge freedom

Under the motion model (5), we observe that the above joint estimation can not be solved directly due to the presence of rotational gauge freedom. For any rotation matrix Q and some rotation parameters $\mathcal{A}$, the following identities can be established

$$\hat{\mathbf{n}}_i^\intercal R(\boldsymbol{\vartheta})\hat{\mathbf{e}} = \hat{\mathbf{n}}_i^\intercal Q^\intercal Q R(\boldsymbol{\vartheta})\hat{\mathbf{e}} = (Q\hat{\mathbf{n}}_i)^\intercal Q R(\boldsymbol{\vartheta})\hat{\mathbf{e}} \tag{14}$$

$$\begin{aligned} \text{and,}\quad Q\hat{\mathbf{n}}_i &= Q\frac{(R(\mathbf{r}^u_{\mathcal{A}})^\intercal \mathbf{u}^{rs}_i) \times (R(\mathbf{r}^v_{\mathcal{A}})^\intercal \mathbf{v}^{rs}_i)}{\|(R(\mathbf{r}^u_{\mathcal{A}})^\intercal \mathbf{u}^{rs}_i) \times (R(\mathbf{r}^v_{\mathcal{A}})^\intercal \mathbf{v}^{rs}_i)\|} \\ &= \frac{(Q^\intercal R(\mathbf{r}^u_{\mathcal{A}})^\intercal \mathbf{u}^{rs}_i) \times (Q^\intercal R(\mathbf{r}^v_{\mathcal{A}})^\intercal \mathbf{v}^{rs}_i)}{\|(Q^\intercal R(\mathbf{r}^u_{\mathcal{A}})^\intercal \mathbf{u}^{rs}_i) \times (Q^\intercal R(\mathbf{r}^v_{\mathcal{A}})^\intercal \mathbf{v}^{rs}_i)\|} \\ &= \frac{(R(\mathbf{r}^u_{\mathcal{A}'})^\intercal \mathbf{u}^{rs}_i) \times (R(\mathbf{r}^v_{\mathcal{A}'})^\intercal \mathbf{v}^{rs}_i)}{\|(R(\mathbf{r}^u_{\mathcal{A}'})^\intercal \mathbf{u}^{rs}_i) \times (R(\mathbf{r}^v_{\mathcal{A}'})^\intercal \mathbf{v}^{rs}_i)\|} \end{aligned} \tag{15}$$

where we utilize the properties of the cross product, and that rotations preserve the Euclidean norm. $\mathcal{A}'$ is the modified motion parameters with the initial rotation Q. From the above identities, it is clear that the algebraic error and the spherical error have an intrinsic gauge freedom, and hence the optimal camera motion and VDs are only defined up to a global rotation freedom. For the algebraic error this is also easy to see if one has zero error (and the VD therefore perfectly aligned with the interpretation plane), but demonstrating gauge freedom for the geometric error in general is rather involved. The main reason is that the proof is non-constructive: due to non-linearities the relation between a rotation applied on all $\hat{\mathbf{n}}_i$ and the one applied on $R(\boldsymbol{\vartheta})$ is implicit. We cast the gauge invariance for Geometric error as a *conjecture* and provide further discussion in the supplementary material.

The above gauge rotational invariance introduces a gauge freedom of degree 3. We require fixing this independence [23, 34] to remove the ambiguity. We now describe two options in the following.

Natural choice An obvious choice to fix the Gauge independence could be $\alpha = \beta = \gamma = 0$. [27] suggested similar choices in their formulation. This choice will remove the $3-$fold ambiguity in the solution space. Furthermore, under this choice, the rotation $\mathbf{r}^0_{\mathcal{A}}$ becomes $\mathbf{0}$ at $\zeta = 0$, *i.e.*, $R(\mathbf{r}^0_{\mathcal{A}}) = I$ which implies no motion of the RS camera while capturing the first row. Hence, under this choice of gauge fixing, the motion parameters (5) become

$$\begin{cases} r_x = a_1\zeta + \ldots + a_n\zeta^n \\ r_y = b_1\zeta + \ldots + b_n\zeta^n \\ r_z = c_1\zeta + \ldots + c_n\zeta^n. \end{cases} \tag{16}$$

We consider polynomials of order $n = 2$, thus, the number of parameters is 9 (where $|\mathcal{A}| = 6$ and $|\boldsymbol{\vartheta}| = 3$). Note that ideally we can fix the global rotation $[\alpha, \beta, \gamma]^\intercal$ to any row. We have tried with fixing it to the zero rotation at the mid-row of an RS image and obtained very similar results.

Aesthetic choice Another choice is to fix one of the VDs to be vertical in the rectified image for a visually pleasant output. The other (orthogonal) VDs are unconstrained. Therefore, we allow only in-plane rotation (roll). *i.e.*, fix $\alpha = 0$, $\beta = 0$ but allow γ to have any value. An additional constraint has to be incorporated to enforce one of the VD orthogonal. We set the constraint as the following lemma.

Lemma 1 *Rotating the canonical axis with $\psi = \theta\phi$ is analogous to fixing the vanishing direction $\hat{\boldsymbol{\nu}}_y$ vertical.*

Proof If $R(\boldsymbol{\vartheta})$ is the rotation matrix (6) corresponding to $\boldsymbol{\vartheta} = [\theta,\ \phi,\ \psi]^\intercal$, then $R(\boldsymbol{\vartheta})\hat{\mathbf{e}}_y = \hat{\boldsymbol{\nu}}_y$. *i.e.*, $\hat{\boldsymbol{\nu}}_y$ is just the second column of the rotation matrix (6). Thus, $\hat{\boldsymbol{\nu}}_y$ is vertical if x-component $2\theta\phi - 2\psi$ becomes zero. *i.e.*, $\psi = \theta\phi$.

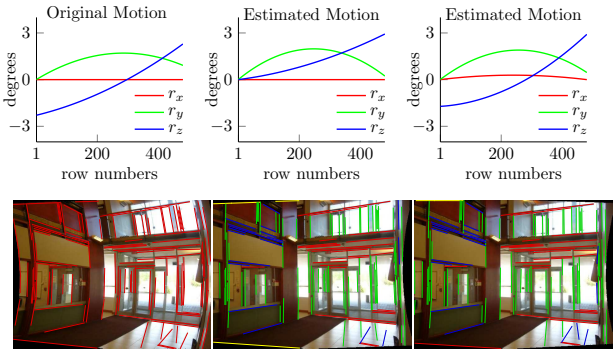
Conversely, if $\psi = \theta\phi$, the x-component of $\hat{\boldsymbol{\nu}}_y$ is zero and hence the VD $\hat{\boldsymbol{\nu}}_y$ is vertical. ∎

According to Lemma 1, under the aesthetic choice of gauge, we need to estimate only 2 parameters $\{\theta,\ \phi\}$ for the orthogonal VDs which are the canonical rotations along X-axis and Y-axis and additionally $3n+1 = 7$ motion parameters $\mathcal{A}$ (yielding again 9 parameters in total) for polynomials of order $n = 2$. In all of our experiments, we employ the aesthetic choice of fixing the gauge (unless stated otherwise).

4.3. Implementation

Our method is implemented in MATLAB. We employ the Levenberg-Marquardt algorithm to optimise (13). The built-in non-linear optimization routine `fmincon` is utilized for this task with user supplied Jacobian which is carried out by applying the chain rule. The details of the optimization steps along with the derivations of the Jacobian are described in supplementary material.

Initialization We initialize VDs along the canonical cardinal direction and the motion parameters are initialized as zeros. We have tried with an elegant initialization of the VDs by adopting a minimal solver [37] for a GS camera



(a) Synthetic Image (b) Natural Choice (c) Aesthetic Choice

Figure 5: Joint estimation of the RS camera motion and the orthogonal VDs: (a) a synthetically generated polynomial motion (5) and the extracted LSs on the synthetic image, (b) joint estimation of RS camera motion and the orthogonal VDs with natural choice of gauge fixing - colors are used to distinguish the VDs, and (c) joint estimation with aesthetic choice of gauge fixing.

under the Manhattan world. However, our simple choice of initialization works well in practice and utilized in all the experiment.

Robust estimator ρ In this work, we utilize the Huber M-estimator [16] defined as follows:

$$\rho(x) = \begin{cases} 0.5x^2 & \text{if } |x| < \delta \\ \delta(|x| - 0.5\delta) & \text{otherwise.} \end{cases} \tag{17}$$

The inlier threshold δ for the above M-estimator defines the maximum deviation attributable to the effect of noise of the LSs. A LS is considered as an outlier if it does not agree with any of the VDs within the error threshold δ. Experimentally, we experienced the best choice as $\delta = 2$ pixels.

LS detector We adopt the LS detector `lsd` [12] in all of our experiments. In an RS camera, some natural lines in a 3D scene become arc segments; `lsd` approximates those low curvature arcs by multiple short line segments. We tune a specific set of parameters in the `lsd` detector for which it can detect near perfect LSs (arcs with very low curvatures). In particular, we set the following parameters:

- the gradient angle tolerance in the region growing algorithm $= 45°$,
- the density of the aligned points of a rectangle is $= 0.5$,
- and the minimum of the lengths of the considered line segments is chosen as 25 pixels.

Image Rectification The corrected image can be obtained by a forward mapping procedure [10] of the RS pixels into the global frame (4) under the estimated motion parameters. The unknown pixels are interpolated linearly. Pixels located outside of the projected frame are placed as intensity 0.

4.4. Limitations

Our proposed method cannot be applied to every image exhibiting rolling shutter artefacts. As with most other methods it comes with several limitations:

- The image content should comply to the MWA to some extent, *i.e.*, two VDs are necessary. However, the majority of the images containing line segments actually satisfy MWA [35]. We further advocate this fact – overall 78.2% (43 out of 55) images in the existing RS datasets [10, 13, 17, 27][1] satisfy MWA, and 93.5% (43 out of 46) of which line segments were present. Our method, being much faster and accurate than [27], can certainly be used to rectify those. Again, the images for which line segments are absent, [27] also fails.
- The depth of the scene is assumed to be sufficiently large for the translational motion to be insignificant.
- We consider only static images at this point where the lens distortions were assumed to be negligible.
- Camera motion is assumed to be smooth during the image exposure period. This is a non-restrictive assumption for hand-held cameras, but may pose problems with cameras mounted on vehicles without vibrations dampening.

However, rolling shutter compensation from a single image is an ill-posed problem in general, therefore, all existing methods need to rely on some prior assumptions. Most limitations above are shared with other works such as [27, 30].

5. Results

We conduct experiments on some synthetic and real images to verify the effectiveness and efficiency of the proposed RS correction method. In particular, we justify our claim that the proposed method is able to restore the geometry of the image more accurately. Certainly, direct pixel-wise measurements (PSNR, *etc.*) are not good choices to evaluate the consistency of the geometry. In addition, there is a global rotational gauge bias in the output. Here we consider rather 3D geometric models [14] to evaluate the baselines quantitatively.

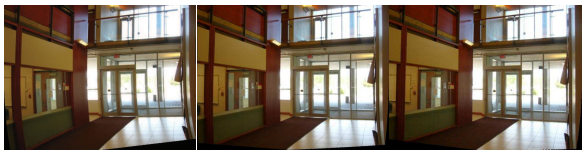
5.1. Synthetic Data

Effectiveness of the proposed method In this section, we perform an experiment on a synthetic data. We choose an image (`P1040850`) from the York Urban Image datasets [9] for this experiment. We synthesize an RS

[1] Images in the paper and in the supplementary material

(a) X-axis only (b) $\mathcal{H}_{mre} = 9.78p$ (c) $\mathcal{H}_{mre} = 3.35p$

(d) Y-axis only (e) $\mathcal{H}_{mre} = 4.90p$ (f) $\mathcal{H}_{mre} = 1.03p$



(g) Z-axis only (h) $\mathcal{H}_{mre} = 1.04p$ (i) $\mathcal{H}_{mre} = 0.70p$

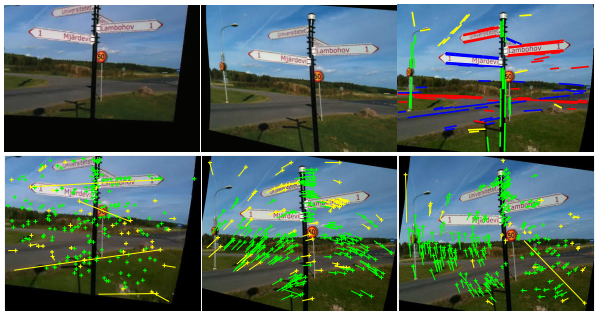
Synthesized Image [27] **[Ours]**

Figure 6: Comparison of the proposed method with [27]: (a), (d) and (g) are the synthesized RS images where the motions are generated only along X-axis, Y-axis and Z-axis respectively. Images (b), (e) and (h) are the corresponding results by [27]. Images (c), (f) and (i) are the results by the proposed method.

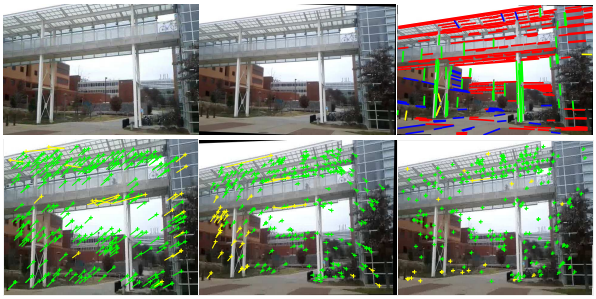
image by randomly generating coefficients $\mathcal{A}$ of the polynomial motion (5) with mean 0 and std 0.02 where α and β were fixed as 0. In Figure 5(a), we display the synthesized RS image. `lsd` detector is applied on the synthetic RS image and the detected LSs are also displayed. Notice that most of the straight LSs has now become arc segments in the synthesized RS image. `lsd` approximates those low curvature arcs by several shorter line segments.

First, the proposed joint estimation (13) is employed under the natural choice of gauge fixing. In Figure 5(b), we display the estimated camera motion, the estimated VDs, and the restored image with LSs-VDs associations respectively[2]. In Figure 5(c), we show the results of the joint estimation under the aesthetic choice of gauge fixing. In the later case, non-zero value of γ enables the corrected image to have an inplane rotation. The mean angular error of the estimated motion (*i.e.* the average absolute differences of the rotations $\angle(R(\mathbf{r}^t_{\mathcal{A}}), R(\mathbf{r}^t_{\mathcal{A}'}))$ for all the rows) were $0.25°$ and $0.18°$ for the natural choice and for the aesthetic choice of the gauge parameters respectively. Note that γ was fixed as zero during the computation of the error.

[2]Similar color scheme as in Figure 1 is utilized throughout.



(a) $|\mathcal{R}_F| = 196.58$, $\sigma_{\mathcal{R}} = 7.60$ [10] (b) $|\mathcal{R}_F| = 186.44$, $\sigma_{\mathcal{R}} = 7.31$ [27] (c) $|\mathcal{R}_F| = 212.39$, $\sigma_{\mathcal{R}} = 8.40$ **[Ours]**



(d) $|\mathcal{R}_F| = 237.27$, $\sigma_{\mathcal{R}} = 3.49$ [13] (e) $|\mathcal{R}_F| = 229.44$, $\sigma_{\mathcal{R}} = 6.63$ [27] (f) $|\mathcal{R}_F| = 239.83$, $\sigma_{\mathcal{R}} = 4.94$ **[Ours]**

Figure 7: Comparison on the image sequences: (a)-(c) Results on `clip03.mov` sequence from [10] captured by an iPhone. (d)-(f) Results on `nxs_wobble_6_dual.mov` sequence from [13] captured by Nexus S. A selected image-pair from each of the sequences is displayed in separate rows for better qualitative comparison. The inliers-outliers are displayed only on the second image (bottom row) of the image pairs along with the mean and *std* of the number of inliers. The estimated VDs are also displayed.

Comparison with the baselines In this section, we compare the proposed method with one of the most relevant baselines [27] on synthesized images where the motions were generated randomly along the individual axes. The evaluation metric considered here is the mean reprojection error $\mathcal{H}_{mre}$ of the original image and the restored image, upto a global rotational homography[14] (or conjugate rotation) due to gauge freedom. The estimation of the rotation and the computation of $\mathcal{H}_{mre}$ are performed on a discrete set of point-correspondences. Let $\{(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}'_i) : i \in \mathcal{I}\}$ is the set of normalized (i.e. premultiplied by K^{-1}) point-correspondences between the original and restored image.The rotation R is estimated [5] as follows

$$[U, S, V] = \text{svd}\Big(\sum_{i\in\mathcal{I}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i'^{\intercal}\Big), \quad R = VU^{\intercal} \tag{18}$$

The $\mathcal{H}_{mre}$ is then computed as the mean of the geometric error (see Section 4.2.2 of [14]) of the point correspon-

dences w.r.t. the rotational homography $H = KRK^{-1}$ in terms of pixel coordinates. The point correspondences are obtained by detecting a number of SIFT key points [22] on the pair of images and then matched across the image pair using VLFeat[3] toolbox. Note that the outliers are discarded and 250 best scoring point-correspondences are chosen. The $\mathcal{H}_{mre}$ (in pixels) are displayed in Figure 6. We observe that the restored image by our proposed method has smaller reprojection errors $\mathcal{H}_{mre}$ than [27][4].

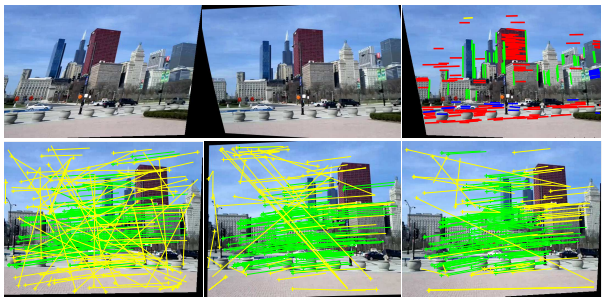
5.2. Real Data

Comparison with the baselines on video We apply our proposed method frame-by-frame on image sequences from the datasets [10][5] and [13][6]. The sequences in the former datasets are more distorted than the latter one. Here, the evaluation is done on image pairs chosen from the rectified frames of a sequence. Hence, the images in the pair are related by epipolar geometry, and we estimate the fundamental matrix between them for evaluation. The second dataset does not come with calibration information, and we estimated the focal length as 0.9 times the maximal image dimension and the principal point as the image center in order to apply our approach. The RANSAC procedure was applied 100 times on each image pair. The inlier threshold is chosen as 0.5 pixels in all the cases. The mean and the standard deviation of the number of found inliers $\mathcal{R}_F$ are reported in Figure 7. Note that [27] and our proposed method utilize only a single frame for the RS correction; in contrast, [10, 13] exploit all the images in the sequence for the rectification. We observe that our method performs better than the single-image baseline [27] and is equivalent with multi-image aproaches [10, 13].

Comparison with the hardware solution The proposed method is also evaluated with the hardware-based solution [17][7]. Note that [17] estimates the camera rotation from the gyroscope readings. The results are displayed in Figure 8. The method [17] failed to synchronise the gyroscope motion precisely and restored the chosen image-pairs inaccurately. Although, [27] performs quite well on the selected image-pairs, the proposed method exhibits even better.

5.3. Runtime comparison

The run-times were computed on an $i7\ CPU\ 2.8GHz$ (using a single core) with $8Gb$ of RAM. On an average, for a 360×520 image, it takes around 0.3 second to correct an image, including LS detection (0.05 second) and rectification (0.1 second), which is $(50-200)\times$ speed-ups over the most recent method [27] (requires ≈ 45 seconds). Note that both the methods were implemented in MATLAB. Therefore, real-time RS correction for videos with our method is naturally possible by an optimized implementation.



(a) $|\mathcal{R}_F| = 148.90$, $\sigma_{\mathcal{R}} = 2.47$ [17] (b) $|\mathcal{R}_F| = 196.44$, $\sigma_{\mathcal{R}} = 5.23$ [27] (c) $|\mathcal{R}_F| = 208.64$, $\sigma_{\mathcal{R}} = 5.67$ **[Ours]**

Figure 8: Comparison of the proposed method with [17] and [27] applied on a video sequence. We display the results on a image-pair of the sequence in separate rows, where inliers-outliers are displayed only on the second image.

6. Conclusion

We proposed an RS camera motion compensation method using vanishing directions of the line segments, extracted from a single view. The geometry in the Manhattan world is exploited for the concurrent estimation of the vanishing directions and the motion parameters. The proposed method is also the first of its kind to estimate orthogonal vanishing directions on an RS image. Extensive experiments demonstrate the computational efficiency and the effectiveness of the proposed approach. Furthermore, our approach is much faster than the existing methods and likely to be accelerated to operate in real time. Further, we argued that the majority of the images of urban areas actually satisfy the MWA, thus, can be corrected by more efficient proposed method.

During the RS compensation of a video, individual frames were corrected separately. However, tracking the vanishing directions over the frames while compensating the RS effect can improve the performance and hence there lies a potential future extension.

Acknowledgements

We acknowledge MoD/Dstl and EPSRC for providing the grant to support the UK academics (Ales Leonardis) involvement in a Department of Defense funded MURI project. This work was also funded in part by EPSRC EP/M026477/1.

[3] http://www.vlfeat.org/

[4] The results are supplied by the authors upon request.

[5] https://www.cvl.isy.liu.se/research/datasets/rs-dataset/

[6] http://www.cc.gatech.edu/cpl/projects/rollingshutter/

[7] http://users.ece.utexas.edu/~bevans/projects/dsc/software/rollingShutter/

References

[1] D. G. Aguilera, J. G. Lahoz, and J. F. Codes. A new method for vanishing points detection in 3d reconstruction from a single view. In *ISPRS comission 2*, 2005. 3

[2] O. Ait-Aider, N. Andreff, J. M. Lavest, and P. Martinet. Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In *ECCV*, pages 56–68. Springer, 2006. 2

[3] C. Albl, Z. Kukelova, and T. Pajdla. R6p-rolling shutter absolute camera pose. In *CVPR*, pages 2292–2300. IEEE Computer Society, 2015. 2, 3

[4] M. Antunes and J. P. Barreto. A global approach for the detection of vanishing points and mutually orthogonal vanishing directions. In *CVPR*, pages 1336–1343, 2013. 3

[5] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700, 1987. 7

[6] J.-C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys. Globally optimal line clustering and vanishing point estimation in manhattan world. In *CVPR*, pages 638–645. IEEE Computer Society, 2012. 3

[7] J. M. Coughlan and A. L. Yuille. Manhattan world: Orientation and outlier detection by bayesian inference. *Neural Computation*, 15(5):1063–1088, 2003. 3

[8] Y. Dai, H. Li, and L. Kneip. Rolling shutter camera relative pose: Generalized epipolar geometry. In *CVPR*. IEEE Computer Society, June 2016. 2

[9] P. Denis, J. H. Elder, and F. J. Estrada. Efficient edge-based methods for estimating manhattan frames in urban imagery. In *ECCV*, pages 197–210. Springer, 2008. 6

[10] P.-E. Forssén and E. Ringaby. Rectifying rolling shutter video from hand-held devices. In *CVPR*, pages 507–514. IEEE Computer Society, 2010. 2, 6, 7, 8

[11] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012. 3

[12] v. G. R. Grompone, J. Jakubowicz, J.-M. Morel, and G. Randall. Lsd: a fast line segment detector with a false detection control. *IEEE TPAMI*, 32(4):722–732, 2010. 6

[13] M. Grundmann, V. Kwatra, D. Castro, and I. Essa. Calibration-free rolling shutter removal. In *ICCP*, pages 1–8. IEEE Computer Society, 2012. 2, 6, 7, 8

[14] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1, 2, 6, 7

[15] S. Hee Park and M. Levoy. Gyro-based multi-image deconvolution for removing handshake blur. In *CVPR*, pages 3366–3373. IEEE Computer Society, 2014. 2

[16] P. J. Huber. *Robust statistics*. Springer, 2011. 6

[17] C. Jia and B. L. Evans. Probabilistic 3-d motion estimation for rolling shutter video rectification from visual and inertial measurements. In *MMSP*, 2012. 2, 6, 8

[18] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *ECCV*, pages 27–40. Springer, 2012. 3

[19] J. Košecká and W. Zhang. Video compass. In *ECCV*, pages 476–490. Springer, 2002. 3

[20] T. Kroeger, D. Dai, and L. Van Gool. Joint vanishing point extraction and tracking. In *CVPR*, pages 2449–2457. IEEE Computer Society, 2015. 3

[21] J.-K. Lee and K.-J. Yoon. Real-time joint estimation of camera orientation and vanishing points. In *CVPR*, pages 1866–1874. IEEE Computer Society, 2015. 3

[22] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 8

[23] P. F. McLauchlan. Gauge independence in optimization algorithms for 3d vision. In *International Workshop on Vision Algorithms*, pages 183–199. Springer, 1999. 5

[24] F. M. Mirzaei and S. I. Roumeliotis. Optimal estimation of vanishing points in a manhattan world. In *ICCV*, pages 2454–2461. IEEE Computer Society, 2011. 3

[25] A. Morawiec. *Orientations and rotations*. Springer, 2003. 3

[26] A. Patron-Perez, S. Lovegrove, and G. Sibley. A spline-based trajectory representation for sensor fusion and rolling shutter cameras. *IJCV*, 113(3):208–219, 2015. 2, 3

[27] V. Rengarajan, A. N. Rajagopalan, and R. Aravind. From bows to arrows: Rolling shutter rectification of urban scenes. In *CVPR*, pages 2773–2781, 2016. 1, 2, 3, 5, 6, 7, 8

[28] O. Saurer, K. Koser, J.-Y. Bouguet, and M. Pollefeys. Rolling shutter stereo. In *ICCV*, pages 465–472, 2013. 2

[29] A. Saxena, S. H. Chung, and A. Y. Ng. 3-d depth reconstruction from a single still image. *IJCV*, 76(1):53–69, 2008. 3

[30] S. Su and W. Heidrich. Rolling shutter motion deblurring. In *CVPR*, pages 1529–1537. IEEE Computer Society, June 2015. 1, 2, 3, 6

[31] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, pages 1250–1257. IEEE Computer Society, 2009. 3, 4

[32] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *ECCV*, pages 537–547. Springer, 2008. 3

[33] E. Tretyak, O. Barinova, P. Kohli, and V. Lempitsky. Geometric image parsing in man-made environments. *IJCV*, 97(3):305–321, 2012. 3

[34] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment–a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999. 5

[35] A. Yuille. The manhattan world assumption: Regularities in scene statistics which enable bayesian inference. In *NIPS*, 2000. 6

[36] M. Zhai, S. Workman, and N. Jacobs. Detecting vanishing points using global image context in a non-manhattan world. In *CVPR*, pages 5657–5665. IEEE Computer Society, June 2016. 3

[37] L. Zhang, H. Lu, X. Hu, and R. Koch. Vanishing point estimation and line classification in a manhattan world with a unifying camera model. *International Journal of Computer Vision*, 117(2):111–130, 2016. 5