

MOTION VECTOR FIELD ESTIMATION USING BRIGHTNESS CONSTANCY ASSUMPTION AND EPIPOLAR GEOMETRY CONSTRAINT

S. Hosseinyalamdary^a, A. Yilmaz^a,

^a Photogrammetric Computer Vision, Civil Engineering Department, The Ohio State University (hosseinyalamdary.1, yilmaz.15)@osu.edu

Commission VI, WG VI/4

KEY WORDS: IMU, Camera-IMU Integration, Visual-Inertial, Epipolar Geometry Constraint, Brightness Constancy Assumption, Optical Flow

ABSTRACT:

In most Photogrammetry and computer vision tasks, finding the corresponding points among images is required. Among many, the Lucas-Kanade optical flow estimation has been employed for tracking interest points as well as motion vector field estimation. This paper uses the IMU measurements to reconstruct the epipolar geometry and it integrates the epipolar geometry constraint with the brightness constancy assumption in the Lucas-Kanade method. The proposed method has been tested using the KITTI dataset. The results show the improvement in motion vector field estimation in comparison to the Lucas-Kanade optical flow estimation. The same approach has been used in the KLT tracker and it has been shown that using epipolar geometry constraint can improve the KLT tracker. It is recommended that the epipolar geometry constraint is used in advanced variational optical flow estimation methods.

1. INTRODUCTION

Without the loss of generality, the motion vector field computed by most optical flow estimation methods provides dense pixel correspondences between consecutive images, which is required in applications such as 3D recovery, activity analysis, image-based rendering and modeling (Szeliski, n.d.). It is also a pre-processing step for the higher level missions like scene understanding. Despite its common applicability, precise optical flow estimation remains an unresolved problem in computer vision. While advances have been made, unaccounted variations in the scene content and depth discontinuities pose challenges to optical flow estimation.

Optical flow was introduced in the 1980s, and is based on the brightness constancy assumption between matching pixels in consecutive images. However, the brightness constancy does not provide adequate constraint to estimate the flow vector and spatial smoothness constraint has been utilized by Lucas and Kanade (Lucas and Kanade, 1981) and Horn and Schunck (Horn and Schunck, 1981) to overcome this limitation. The spatial smoothness constraint, however, blurs the object boundaries in the estimation motion vector field. Researchers have attempted to address this shortcoming by extending the Horn and Schunck's approach. Brox et al. (Brox et al., 2004) introduced a gradient constancy term, a spatio-temporal smoothness term in the cost function that was in the form of L1 norm. In a follow up study, the authors employed a descriptor matching strategy to estimate the optical flow in the case of large displacements (Brox and Malik, 2011). While improves the results, such temporal smoothness heuristics do not truly reflect the changes in the optical flow appropriately. Total variation has been employed to estimate the optical flow estimation (Zach et al., 2007, Javier Sanchez et al., 2013).

Due to the complex camera and object motions, modeling temporal changes in optical flow using only the image information is a difficult task. In contrast to extensive studies on image based estimation of the optical flow, few studies has been conducted on using other complimentary sensors to improve the results. Hwangbo et al. 2008 have used homography transformation based on the

rotation measurements collected by gyroscope and approximated the neighborhood of the corresponding pixels. They showed that the KLT tracker initialized by rotational homography works better in handling the fast motion and rotation. Wedel et al. (Wedel and Cremers, 2011) have combined the epipolar geometry and optical flow to estimate the position, velocity and orientation of the moving objects in the scene. IMU measurements has been employed to cancel out the camera motion and recover the pure object motion. Slesareva et al. (Slesareva et al., 2005) used the epipolar geometry constraint as a hard constraint and Valgaerts et al. (Valgaerts et al., 2008) has minimized the epipolar geometry constraint associated with the brightness constancy assumption.

In another study, the sensed gravity in IMU has been used as the vertical reference to find the horizon line and vertical vanishing point in the images (Corke et al., 2007). The integration of the IMU measurements and the estimated pose using image sequences has been extensively studied (Jones and Soatto, 2010, Kelly and Sukhatme, 2011, Li and Mourikis, 2013). Leung and Medioni (Leung and Medioni, 2014) have used the IMU measurements to detect the ground plane and use the points on the ground plane to estimation the pose of the camera. The IMU measurements has been used to compensate the camera motion blur (Joshi et al., 2010).

In this paper, the motion vector field is estimated using epipolar geometry constraint. The epipolar geometry based on the IMU measurements are used to formulate the optical flow between consecutive images with respect to both rotation and translation of the camera. The KITTI dataset is employed to evaluate the proposed method and the results are compared to a brightness constancy based the Lucas-Kanade optical flow estimation. The results show significant improvement in optical flow estimation using the proposed approach.

In the next section, the brightness constancy based optical flow estimation and the epipolar geometry is briefly explained. An approach has been proposed in section 3 which solve the motion vector field using brightness constancy and epipolar geometry. Section 4 explains evaluation of the proposed approach and provide the results for the KITTI dataset. In final section, it is drawn

conclusion for this paper.

2. BACKGROUND

In spite of great improvements in optical flow estimation over the last two decades, exploiting complimentary sensory data for optical flow estimation has not extensively been studied. Hence, in order to facilitate the proposed discussion we first introduce some basic concepts of optical flow and IMU measurements, which will be followed by a discussion on how the fusion of the two sensory data works to aid optical flow estimation.

2.1 brightness constancy assumption

Optical flow estimation is based on the brightness (or color) constancy assumption. It means that two corresponding pixels in consecutive images at times t and $t + 1$ should have the same brightness:

$$I(x, y, t) = I(x + u, y + v, t + 1). \quad (1)$$

where x, y = pixel coordinates
 u, v = motion vector
 $I(x, y, t)$ = brightness of the pixel which is a non linear function of pixel coordinates and time

Equation (1) expresses a non-linear functional which can be linearized using Taylor series expansion

$$I_x u + I_y v + I_t = 0. \quad (2)$$

where $I_x = \frac{\partial I(x+u, y+v, t+1)}{\partial x}$
 $I_y = \frac{\partial I(x+u, y+v, t+1)}{\partial y}$
 $I_t = \frac{\partial I(x+u, y+v, t+1)}{\partial t}$

The brightness constancy, however, provides a single constraint for estimation of motion vector which has two unknowns. A typical constraint used many is to assume that the neighboring pixels share the same motion, such that an overdetermined system of equations can estimate the motion parameters. This effectiveness comes at the expense of blurring the object and depth boundaries and creates incorrect results. Assuming that the neighboring pixels move together, the motion vector can be estimated

$$\mathbf{l} = \mathbf{A}\mathbf{u}. \quad (3)$$

where $\mathbf{A} = \begin{bmatrix} I_x(1) & I_y(1) \\ I_x(2) & I_y(2) \\ \vdots & \vdots \\ I_x(N) & I_y(N) \end{bmatrix}$
 $\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix}$
 $\mathbf{l} = \begin{bmatrix} I_t(1) \\ I_t(2) \\ \vdots \\ I_t(N) \end{bmatrix}$

the numbers in parenthesis is the number of each pixel in a neighborhood

Equation (3) can be estimated using least squares

$$\mathbf{u} = (\mathbf{N})^{-1}\mathbf{c}. \quad (4)$$

where $\mathbf{N} = \mathbf{A}^T \mathbf{A}$
 $\mathbf{c} = \mathbf{A}^T \mathbf{l}$

The covariance matrix of the estimation motion vector field is

$$\Sigma_{\mathbf{u}} = \sigma_0^2 \mathbf{N}^{-1} \quad (5)$$

where σ_0^2 = reference variance
 $\Sigma_{\mathbf{u}}$ = the covariance matrix of the estimated motion vector field

The Lucas-Kanade optical flow can be improved if a filter is used in the neighborhood. First, the gaussian filter which the central pixel of the neighborhood has more contribution in the optical flow estimation improves the motion vector field estimation. Second, the anisotropic filter which suppresses the contribution of the pixels with different brightness values and therefore, the similar pixels have more impact on the motion vector field estimation. The bilateral filter, which has been used in this paper, combines the gaussian filter and anisotropic filter

$$w(i, j, k, l) = e^{-\frac{(i-k)^2 + (j-l)^2}{2\sigma_d^2} - \frac{||I(i, j) - I(k, l)||}{2\sigma_r^2}}. \quad (6)$$

where i, j = coordinates of the central pixel in the neighborhood
 j, k = coordinates of the each pixel in the neighborhood
 $I(i, j) - I(k, l)$ = the brightness difference of the central pixel and its neighborhood pixel
 σ_d = the standard deviation of the gaussian filter
 σ_r = the standard deviation of the anisotropic filter

2.2 Epipolar geometry constraint

If the translation and rotation of the camera are known from external sensors, the geometry of the camera exposure stations can be constructed. This dictates a constraint that the corresponding pixel of a pixel in the first image lies on a line in the second image, known as epiline. This line is

$$\mathbf{l}' = \mathbf{F}\mathbf{x} \quad (7)$$

where \mathbf{l}' = a vector which indicates the epiline in the second image
 \mathbf{x} = the inhomogeneous pixel coordinates in the first image
 \mathbf{F} = the 3×3 matrix known as fundamental matrix

The corresponding pixel lies on the epiline in the second image and therefore

$$\mathbf{x}'^T \mathbf{l}' = \mathbf{x}'^T \mathbf{F}\mathbf{x} = 0 \quad (8)$$

where \mathbf{x}' = the inhomogeneous coordinates of the corresponding pixel of the first image in the second image

This equation is known as epipolar geometry constraint. Replacing $\mathbf{x}' = [x + u, y + v, 1]^T$ into equation (8), it can be written as

$$u(xf_{11} + yf_{12} + f_{13}) + v(xf_{21} + yf_{22} + f_{23}) + \mathbf{x}^T \mathbf{F} \mathbf{x} = 0 \quad (9)$$

where f_{ij} = the element of the fundamental matrix at row i and column j

The fundamental matrix indicates the geometry of the camera exposure stations. It is a rank deficient 3×3 matrix

$$\mathbf{F} = \mathbf{K}^{-T} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{K}^{-1} \quad (10)$$

where \mathbf{K} = the calibration matrix of the camera
 \mathbf{R}, \mathbf{t} = rotation and translation of the camera between two camera exposure stations
 $[\cdot]_{\times}$ = the skew symmetric matrix corresponding to the cross product.

Inertial measurement unit (IMU) is a collection of accelerometers and gyroscopes which measures the acceleration and the angular rate of the IMU frame with respect to the inertial (space-fixed) frame independent from the environment. The translation, velocity and rotation can be calculated from the IMU measurement and this process is called IMU mechanization (Jekeli 2007). In order to integrate IMU and camera information, the measurements should be synchronized and should be converted into the same coordinate system. The sensors can be synchronized by using hardware and software triggering for real time purposes and can be interpolated in post processing mode.

Since the IMU measurements have various error sources, ie. bias and scale error, its error accumulates over time; hence, it is customary to integrate the IMU measurements with GPS to maintain the navigation accuracy over time. For this purpose, Kalman filter and its variations are used to provide the optimal navigation solution.

The camera calibration matrix, K , can be calculated from calibration operation. The rotation and translation can be estimated from the external sensors such as IMU accelerometers and gyros. Therefore, the geometry of the camera exposure stations can be reconstructed and the fundamental matrix is known. Hence, the epipolar geometry constraint can be used to improve the motion vector field estimation. This constraint does not provide a one to one correspondence between two images and therefore, other constraints such as brightness constancy assumption should be employed in conjunction with the epipolar geometry constraint to estimate the motion vector field.

It should be noted that epipolar geometry constraint holds if the back projection of the pixels in the images do not lie on a plane and the motion of the camera does include the translational motion. In this paper, the camera mounted on a car records the images in an urban environment. In this scenario, there are many objects on the road and the back projection of the pixels lie on multiple planes. In addition, the dynamic of the car dictates that translational motion is unavoidable even at turns. Therefore, we justify that there is no violation of the epipolar geometry assumption.

2.3 Sensor fusion

Each sensor has its own frame, an origin and orientation. For the camera, its origin is at the camera center and x and y coordinates define a plane parallel to the image plane and the z coordinate (principal axis) is perpendicular to image plane toward the scene. The origin of IMU is its center of mass and its orientation depends on the IMU. GPS origin is the carrier phase center of receiver and it does not have any orientation.

If the sensors are mounted on a rigid platform, different sensor measurements can be transformed to another sensor frame and the measurements of different sensors can be integrated. In order to find the transformation between sensors, their location and orientation are measured before the operation. The displacement vector between two sensor origins (IMU center of mass and the camera perspective center) is called the lever-arm, and the rotation matrix between two frames (IMU orientation and camera alignment) is called boresight. In the following text, we use the subscript to denote the source reference frame and the superscript as the destination sensor frame. Also, the letter "i" and "c" in subscript or superscript stand for IMU frame and camera frame respectively.

In the optical flow estimation, the motion vector field is estimated with respect to the previous image. Therefore, the rotation and translation of the camera should be calculated with respect to the previous epoch. Since the IMU and camera frames are moving with the vehicle, the measurement of these sensors should be transformed to the reference frame. The navigation frame is defined as an intermediate frame to facilitate this transformation. the origin of the navigation frame is assumed at the IMU center of mass and it is aligned toward East, North, and Up, respectively. The axes of navigation frame is assumed to be fixed since the platform movement does not travel more than few hundred meters in this dataset. It should be noted that the subscript or superscript "n" indicates the navigation frame. The camera rotation with respect to reference frame is given as follows:

$$\mathbf{C}_{c,k-1}^{c,k} = \mathbf{C}_{i,k}^{c,k} \mathbf{C}_n^{i,k} \mathbf{C}_{i,k-1}^n \mathbf{C}_{i,k-1}^{c,k-1} \quad (11)$$

where i, k = first letter indicates the frame and the second letter is the epoch number.

Obviously, $\mathbf{C}_{i,k}^{c,k}$ and $\mathbf{C}_{i,k-1}^{c,k-1}$ are boresight and it is does not change when the vehicle travels. $\mathbf{C}_{i,k-1}^n$ and $\mathbf{C}_n^{i,k}$ can be estimated using IMU measurements at reference epoch and epoch k . Similarly, the translation vector from the camera frame to the reference frame, $\mathbf{t}_{c,k}^{c,k-1}$, is:

$$\mathbf{t}_{c,k}^{c,k-1} = \mathbf{C}_{i,k-1}^{c,k-1} \mathbf{C}_n^{i,k-1} (\mathbf{t}_{i,k}^n - \mathbf{t}_{i,k-1}^n + \mathbf{C}_{i,k-1}^n \mathbf{C}_{c,k-1}^{i,k-1} \mathbf{d}). \quad (12)$$

where \mathbf{d} = lever arm

3. OPTICAL FLOW ESTIMATION USING EPIPOLAR GEOMETRY CONSTRAINT

In this paper, we provide an approach to improve the Lucas-Kanade optical flow and KLT tracker using epipolar geometry.

In the Lucas-Kanade optical flow, the motion vector field is initialized with zero. This brightness constancy assumption is used to converge the motion vector in an iterative scheme. In order to improve the Lucas-Kanade approach using epilines, epilines can be redefined using equation (6)

$$a_1 u + a_2 v = a_3 \quad (13)$$

where

$$\begin{aligned} a_1 &= x f_{11} + y f_{12} + f_{13} \\ a_2 &= x f_{21} + y f_{22} + f_{23} \\ a_3 &= -\mathbf{x}^T \mathbf{F} \mathbf{x} \end{aligned}$$

It should be noted that epilines should be defined in the same coordinate system as the coordinate system of brightness constancy assumption. Knowing that the corresponding pixels lie on the epilines, the pixels of the first image can be projected into the epilines in the second image and the projected point can be used as the initial motion vector. It improves the initial motion vector field which lead to faster convergence. Figure (1) presents the scheme of the proposed approach.

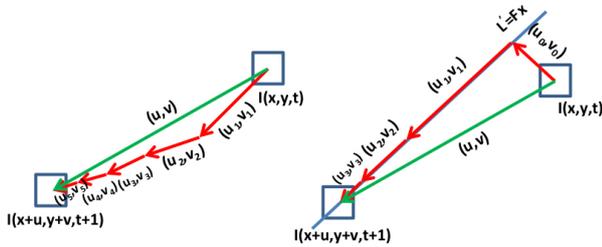


Figure 1: Optical flow estimation: Left: the Lucas-Kanade; Right: the Lucas-Kanade aided by epipolar geometry

After motion vector field initialization using the epilines, the epipolar geometry constraint is employed as a hard constraint and the brightness constancy assumption is used to estimate the motion vector which lies on the epilines. In other words, we transform the 2D motion vector field estimation into 1D motion estimation problem

$$\begin{cases} \mathbf{A} \mathbf{u} = \mathbf{1} \\ \mathbf{b} \mathbf{u} = \mathbf{z} \end{cases} \quad (14)$$

where

$$\mathbf{b} = \begin{bmatrix} a & b \end{bmatrix}$$

$$\mathbf{z} = [a_3]$$

This equation is known as least squares with constraint. it can be estimated as follows

$$\mathbf{u} = \mathbf{N}^{-1} \mathbf{c} + \mathbf{N}^{-1} \mathbf{b}^T (\mathbf{b} \mathbf{N}^{-1} \mathbf{b}^T)^{-1} (\mathbf{c} - \mathbf{b} \mathbf{N}^{-1} \mathbf{c}) \quad (15)$$

the term $\mathbf{c} - \mathbf{b} \mathbf{N}^{-1} \mathbf{c}$ is called vector of discrepancies which states how different are the brightness constancy and the epipolar geometry constraint. In other words, it shows the brightness constancy assumption follows the epipolar geometry. The covariance of the optical flow estimation using epipolar geometry is

$$\Sigma_{\mathbf{u}} = \sigma_0^2 [\mathbf{N}^{-1} - \mathbf{N}^{-1} \mathbf{b}^T (\mathbf{b} \mathbf{N}^{-1} \mathbf{b}^T)^{-1} \mathbf{b} \mathbf{N}^{-1}] \quad (16)$$

The first term is brightness constancy assumption covariance matrix and using epipolar geometry decreases the covariance matrix and improves the precision of the optical flow estimation.

4. EXPERIMENT

In this section, the dataset which is used in this paper is described, and the evaluation procedure and the results of the proposed method is presented.

4.1 KITTI dataset

In this paper, the KITTI dataset #5 collected on September 26, 2011 is used to evaluate the IMU based optical flow estimation. The dataset includes the two monochromic and two color cameras, laser scanner, and GPS/IMU sensors. The left monochromic camera image sequence and GPS/IMU integration has been used in this paper. The GPS/IMU solution has been provided by OXTS RT 3003 navigation system. This system can provide 2 centimeter position accuracy using L1/L2 carrier phase ambiguity resolution. It can also maintain 0.1 degree orientation accuracy (Geiger et al., 2012, Geiger et al., 2013).

The dataset uses PointGray Flea2 grayscale camera. The car's hood and sky are removed from the images. The camera calibration is given which provides focal length and principal point in pixel unit. The camera measurements in pixels should be converted to the metric unit and then it can be integrated with IMU.

The information are provided in each sensor frame; GPS/IMU navigation unit is aligned to front, left and up of vehicle and the camera is oriented toward right, down and front of the car. The lever arm and boresight between camera and IMU can be computed from the provided information. GPS/IMU navigation solution and images are synchronized with a reasonable accuracy. Figure (2) demonstrates the sensor configuration on the vehicle.

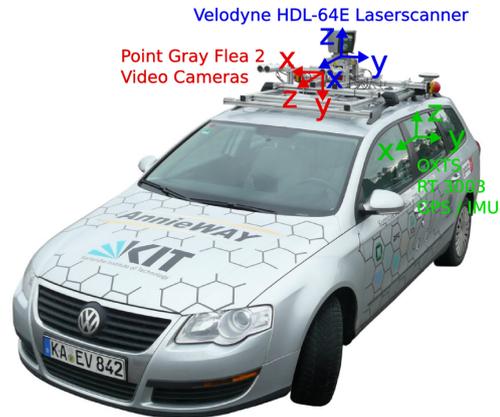


Figure 2: Sensor configuration in KITTI dataset

4.2 Procedure

If the motion vector field is assumed as a solution to partial differential equation problem, the motion vector fields is initialized as zero and it can be considered as Dirichlet boundary value problem. Also, we consider that the gradient of the image is zero on the border of image and out of it which is the Neumann boundary value problem. Since equation (2) is a linearized form of the equation (1) and higher order terms are neglected, it holds when the motion vector is small. Therefore, the larger motion

vector can be estimated in a pyramidal scheme. In pyramidal scheme, the image is smoothed and downsampled to lower level of the pyramid. In the lowest level of the pyramid, the Lucas-Kanade optical flow is employed to estimate the motion vector field. Then, the motion vector field is upsampled and it is used as initial motion vector field in the upper pyramid level. The number of the levels depend on the maximum motion in the image. Here, 5 levels of pyramid has been considered for the images and it has been downsampled and upsampled by 2.

In equation (3), it can be seen that motion vector is estimated within a neighborhood using the Lucas-Kanade optical flow. The size of the neighborhood is of importance, the small neighborhood leads to more precise optical flow estimation and the large neighborhood provides more reliable motion vector field. In this paper, the window size of the optical flow estimation is 17×17 at the finest level (level 0) and it decreases by 2 per level. Therefore, the window size from finest to coarsest levels of pyramid are 17×17 , 15×15 , 13×13 , 11×11 , 9×9 and 7×7 . This helps us to have lower window size when the image size has been reduced.

Also, the bilateral filter is employed to increase the precision of the motion vector field estimation (gaussian filter) and improve the estimate of motion vector field at the object boundaries (isotropic filter). The standard deviation for the gaussian filter and isotropic filter are considered as 4 and 50, respectively.

It also should be noted that the fundamental geometry is different for each level. In other words, the calibration matrix should be downsampled at each level and the fundamental matrix should be constructed for each level. Figure (3) shows the output of GPS/IMU which explains the dynamic of vehicle along the operation.

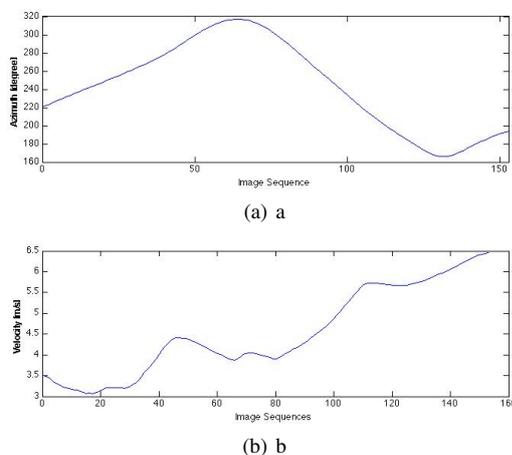


Figure 3: (a) Azimuth (b) Velocity of the vehicle

4.3 Results

The results of the proposed method is shown in Figure (4) and Lucas/Kanade optical flow estimation results are given for comparison purposes. The results are given in four subfigures which shows the motion vector field between image 0 and image 1, image 49 and image 50, image 99 and image 100, and image 149 and image 150 for subfigures (a), (b), (c), and (d), respectively. These subfigures include four component. The first component (top-left) shows the overlaid of the previous image on the current image where the previous image is in green channel and the current image is in red channel. The second component (down-left)

shows the reconstructed the current image from the previous image using the estimated motion vector field known as warped image. The third component illustrates the estimated motion vector field using the proposed method and the fourth component shows the Lucas-Kanade motion vector field. The motion vector field is color-coded, that is the direction of the motion vector is shown by color and the brightness of the color indicates the magnitude of the estimated motion vector.

It should be noted that the warped image has been estimated using the proposed method. In contrast, the warped image using the lucas-Kanade method is not given in this paper and the length of the paper did not allow us to provide these results. Therefore, the proposed approach and the Lucas-Kanade method has been compared using the color-coded motion vector field in left components of each subfigure.

The black regions in warped images has occurred where the warped image goes beyond the image boundaries and there is no pixel value for those regions of the warped image. It occurs more frequently at borders of image where there are many occlusions. Also, the motion vector field estimation is not reliable at the borders of the image since there is not enough information in the neighborhood. The poor estimation of the motion vector field lead to some artifacts in warped images.

As can be seen in Figure (3), there is not significant rotation in camera motion and the translation motion is very slow in epoch 0. The left side of subfigure (a) shows that the proposed method has smoother motion vector field estimation. However, it introduces incorrect motion vector field estimation at the region of the car. The epipolar geometry assumes that the scene is static and the motion is purely due to the camera motion. Therefore, the motion vector field may not be correctly estimated on the moving object in the scene. As previously motioned, both method show some problems at the border of the image. In the warped image, the poorly estimated motion vector field generates some artifact in the warped image. For instance, the bar of the traffic light is not straight in the warped image and the incorrect motion vector field has generated artifacts there.

In subfigure (b), the motion vector field has been poorly estimated in the right side of the image. The right side of the image is very dark and there is very low contrast there. Lack of rich texture leads to ambiguous motion vector field in this part of the image. However, it is clear that the proposed method can handle the textureless regions much better than the Lucas-Kanade optical flow estimation since it exploits the epipolar geometry constraint and it can resolve the ill-conditioned optical flow estimation problem in low contrast and textureless regions.

In subfigure (c) the translational motion of the camera is faster and motion vector field is larger. Surprisingly, the Lucas-Kanade method has better performance with respect to the proposed method. For instance, the ground motion (lower part of image) has been estimated smoother and more realistic in the Lucas-Kanade. In the warped image, left side of the image has severely distorted. The optical flow estimation is poor in this region due to very large motion vector field.

In subfigure (d), there is a large motion in the right side of the image and the Lucas-Kanade fails to estimated the correct motion, but the proposed method has better estimate of the motion vector field in this region. It can be concluded that the proposed method has superior performance over the Lucas-Kanade in large motion vector field. The warped image displays some artifacts in the right side of the image due to the occlusion in this part of the image.

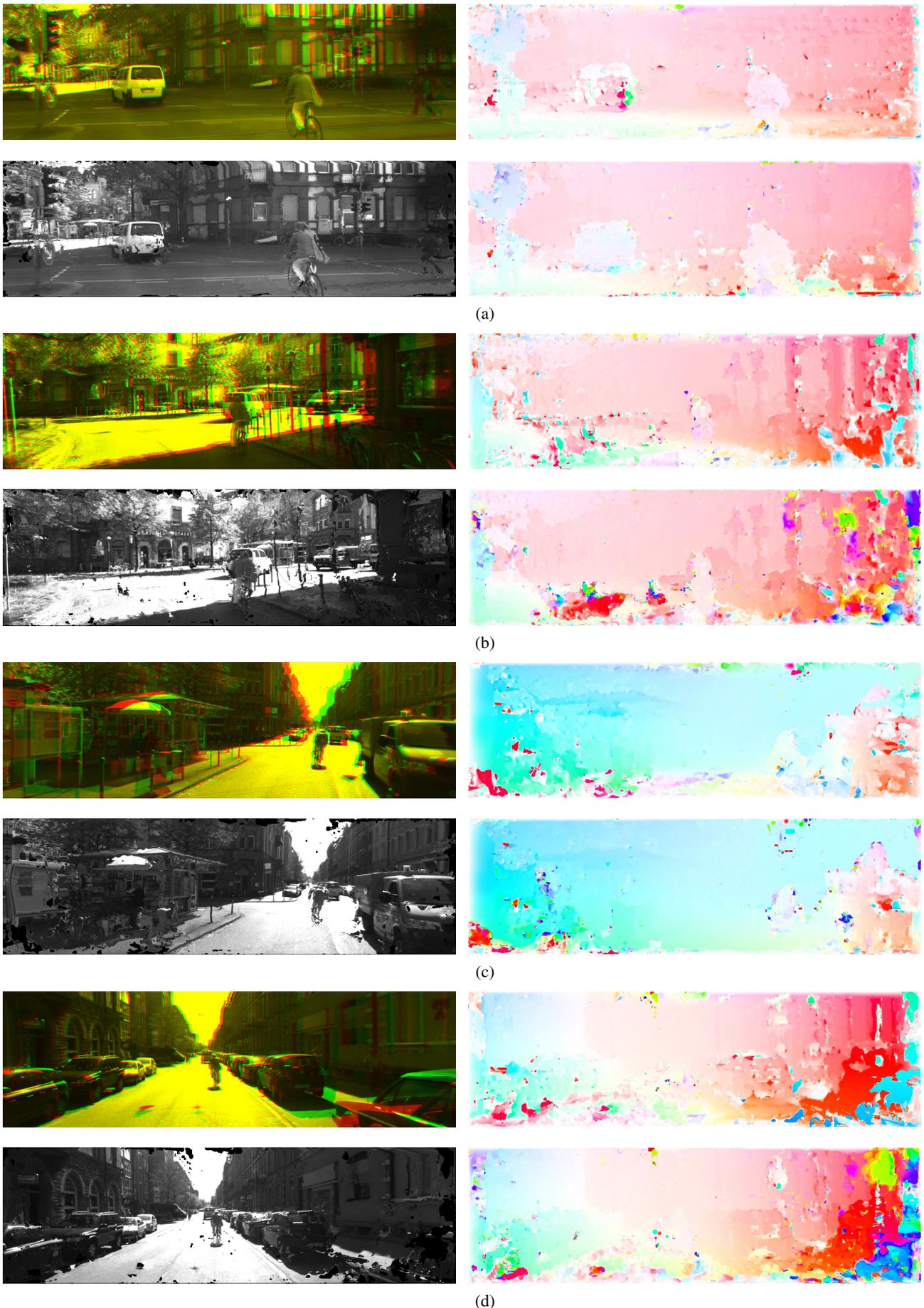


Figure 4: top left: first and second images are overlaid in different colors; top right: the color coded optical flow LK+epipolar geometry constraint; down left: warped image LK+epipolar geometry constraint; down right: warped image LK; a, b, c and d are 1, 50, 100 and 150 frame numbers in the image sequence

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-1-9-2014

The proposed method can be generalized to the KLT tracker. The KLT tracker uses the estimated motion vector for interest points to locate them in the next image. The KLT tracker uses the brightness constancy assumption, but the epipolar geometry can aid the KLT tracker as explained in this paper. We refer to the KLT tracker aided by epipolar geometry constraint as "modified KLT tracker" here. Figure (5) shows the modified KLT tracker between image 80 and image 81. Likewise Figure (4), The first image in green channel is overlaid on the second image in red channel.

Seven points has been tracked in a modified KLT tracker and these points are displayed in small hollow light blue points. The epilines has been estimated for each of these tracked point using IMU measurements and it has been displayed in blue line. The epilines intersect at a point known as epipole. The epipole is the same as focus of expansion if the motion is translational motion and it lies at the center of the image. However, the epilines is not at the center of the image since the camera has rotation too. It can be seen in Figure (3).

The projection of these points on the epipolar line is displayed in solid yellow points. Some of the projected points (yellow) covered the tracked points (blue) since the tracked points were close to the epiline. The proposed method has been used to estimate the corresponding points in the second image which is shown by solid red points. Obviously, the epilines pass trough the corresponding points (red dots) in the second image. The tracking points have been converged to the corresponding points in the second image. Therefore, the proposed method can be used in sparse correspondence to track the interest points faster and make them more reliable.

5. CONCLUSION

In this paper, we proposed a method to improve Lucas/Kanade optical flow using epipolar geometry constraint. The IMU measurement obtains the translation and rotation of the camera exposure stations and if the camera is calibrated, the epipolar geometry can be reconstructed. The brightness constancy assumption in association with the epipolar geometry constraint leads to more reliable motion vector field estimation. In the proposed method, first the original point is projected on the epiline and the projected point is used as an initialization point. Then, the epipolar geometry constraint and the brightness constancy assumption are used in a least squares with constraint scheme to estimate the motion vector field.

The KITTI dataset has been used to verify the proposed method. The GPS/IMU measurements are transformed to the camera co-

ordinate system. The translation and rotation from GPS/IMU and the camera calibration are employed to compose the epipolar geometry. Since the motion vector field is large, it has been estimated in a pyramidal scheme. Using the epipolar geometry and brightness constancy assumption, the motion vector field is estimated for each pyramid level and the estimated motion vector field has been upsample and it has been used as initial motion vector field. The estimated motion vector field in the highest level of pyramid has been demonstrated in this paper.

The results show that the proposed method can improve the estimated motion vector field. The proposed method may fail in the moving object regions since the epipolar geometry constraint assumes that the motion vector field is due to the camera motion and the scene objects are static. On the other hand, the proposed method has superior performance in textureless and low contrast regions.

The proposed approach can be used for sparse correspondence. Once the interest points are detected in an image, the proposed method can be used to track the interest points in the upcoming images. Likewise with the dense correspondence, the interest points are projected on the epiline and then brightness constancy assumption and epipolar geometry constraint are used to estimate the motion vectors and subsequently the corresponding points.

The epipolar geometry constraint can be integrated with a high performance optical flow estimation. It can be used as an additional term in the optical flow cost function. It has been discussed in the literature, but these papers estimate fundamental matrix in conjunction with the motion vector field and they do not consider the IMU measurements to reconstruct the epipolar geometry. Here, it is suggested that the IMU measurement are employed in motion vector field estimation using variational method.

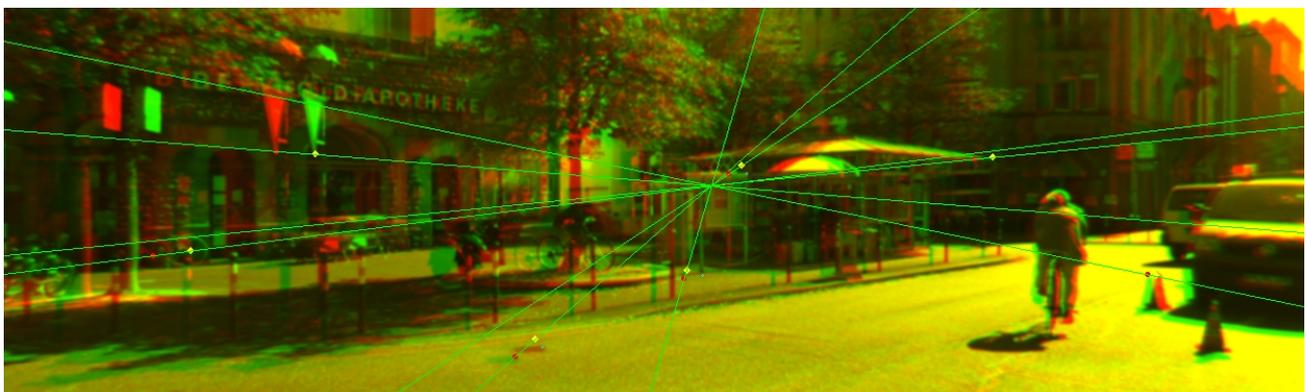


Figure 5: KLT tracker aided by epipolar geometry constraint

REFERENCES

- Brox, T. and Malik, J., 2011. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(3), pp. 500–513.
- Brox, T., Bruhn, A., Papenberg, N. and Weickert, J., 2004. High accuracy optical flow estimation based on a theory for warping. In: *European Conference on Computer Vision (ECCV)*, Springer, pp. 25–36.
- Corke, P., Lobo, J. and Dias, J., 2007. An introduction to inertial and visual sensing. *The International Journal of Robotics* 26, pp. 519–535.
- Geiger, A., Lenz, P. and Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Geiger, A., Lenz, P., Stiller, C. and Urtasun, R., 2013. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*.
- Horn, B. K. P. and Schunck, B. G., 1981. Determining optical flow. *Artif. Intell.* 17(1-3), pp. 185–203.
- Javier Sanchez¹, E. M.-L. J. S., Meinhardt, E. and Facciolo, G., 2013. TV-L1 Optical Flow Estimation. *Image Processing On Line* 3, pp. 137–150.
- Jones, E. S. and Soatto, S., 2010. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research (IJRR)* 30(4), pp. 407430.
- Joshi, N., Kang, S. B., Zitnick, L. and Szeliski, R., 2010. Image deblurring using inertial measurement sensors. In: *ACM SIGGRAPH*.
- Kelly, J. and Sukhatme, G. S., 2011. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *International Journal of Robotics Research* 30(1), pp. 56–79.
- Leung, T. and Medioni, G., 2014. Visual navigation aid for the blind in dynamic environments. In: *The IEEE Conference on Computer Vision and Pattern Recognition*.
- Li, M. and Mourikis, A. I., 2013. High-precision, consistent ekf-based visualinertial odometry. *International Journal of Robotics Research* 32(6), pp. 690–711.
- Lucas, B. D. and Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'81*, pp. 674–679.
- Slesareva, N., Bruhn, A. and Weickert, J., 2005. Optic flow goes stereo: A variational method for estimating discontinuitypreserving dense disparity maps. In: *Proc. 27th DAGM Symposium*, pp. 33–40.
- Szeliski, R., n.d. *Computer vision algorithms and applications*. Springer, London; New York.
- Valgaerts, L., Bruhn, A. and Weickert, J., 2008. A variational approach for the joint recovery of the optical flow and the fundamental matrix. In: *Proc. 27th DAGM Symposium*, pp. 314–324.
- Wedel, A. and Cremers, D., 2011. *Stereo Scene Flow for 3D Motion Analysis*. Springer.
- Zach, C., Pock, T. and Bischof, H., 2007. A duality based approach for realtime tv-l1 optical flow. In: *Ann. Symp. German Association Patt. Recogn.*, pp. 214–223.