Geometric-based Line Segment Tracking for HDR Stereo Sequences

Ruben Gomez-Ojeda, Javier Gonzalez-Jimenez

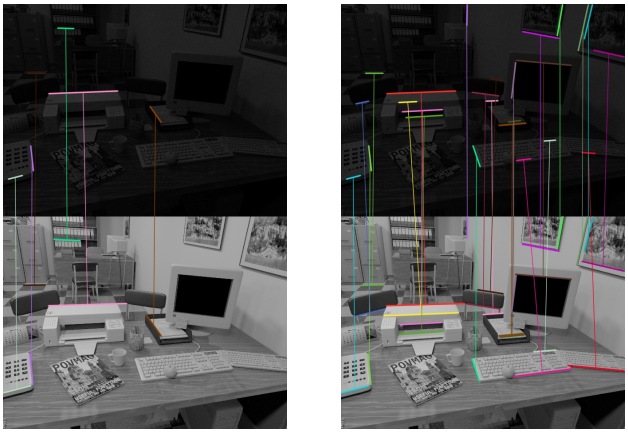
***Abstract*— In this work, we propose a purely geometrical approach for the robust matching of line segments for challenging stereo streams with severe illumination changes or High Dynamic Range (HDR) environments. To that purpose, we exploit the univocal nature of the matching problem, i.e. every observation must be corresponded with a single feature or not corresponded at all. We state the problem as a sparse, convex, ℓ_1-minimization of the matching vector regularized by the geometric constraints. This formulation allows for the robust tracking of line segments along sequences where traditional appearance-based matching techniques tend to fail due to dynamic changes in illumination conditions. Moreover, the proposed matching algorithm also results in a considerable speed-up of previous state of the art techniques making it suitable for real-time applications such as Visual Odometry (VO). This, of course, comes at expense of a slightly lower number of matches in comparison with appearance-based methods, and also limits its application to continuous video sequences, as it is rather constrained to small pose increments between consecutive frames. We validate the claimed advantages by first evaluating the matching performance in challenging video sequences, and then testing the method in a benchmarked point and line based VO algorithm.**

I. Introduction

Although appearance-based tracking has reached a high maturity for *feature-based* motion estimation, its robustness in real-world scenarios is still an open challenge. In this work, we are particularly interested in improving the robustness of visual feature tracking in sequences including severe illumination changes or High Dynamic Range (HDR) environments (see Figure 1). Under these circumstances, traditional descriptors based on local appearance, such as ORB [1] and LBD [2] for points and line segments, respectively, tend to provide many outliers and a low number matches, and hence jeopardizing the performance of the visual tracker.

We claim that line segments can be successfully tracked along video sequences by only considering their geometric consistency along consecutive frames, namely, the oriented direction in the image, the overlap between them, and the epipolar constraints. To achieve robust matches from this reduced segment description we need to introduce some mechanism to deal with the ambiguity associated to such purely geometrical line matching.

For that, we state the problem as a *sparse*, *convex* ℓ_1-minimization of the geometrical constraints from any line segment in the first image over all the candidates in the second one, within a *one-to-many* scheme. This formulation allows for the successful tracking of line segments as it only accepts matches that are guaranteed to be globally unique. In addition, the proposed method results in a considerable speed-up of the tracking process in comparison with traditional appearance based methods. For this reason we believe this method can be a suitable choice for motion estimation algorithms intended to work in challenging environments, even as a recovery stage when traditional descriptor-based matching fails or does not provide enough correspondences.



(a) LSD [3] + LBD [2] (b) LSD [3] + Our proposal

Fig. 1. Pair of consecutive frames extracted from the sequence *hdr/flicker1* from the dataset in [4] under challenging illumination changes. Our approach allows for the robust tracking of line segments in this type of environments, where traditional appearance-based matching techniques tend to fail. changes in illumination conditions or in HDR scenarios where traditional appearance-based matching techniques tend to fail.

To deal with outliers we impose some requiring constraints (e.g. small baseline between the two consecutive images) which slightly reduce the effectiveness of line matching in scenes with repetitive structures and also the number of tracked features.

In summary, the contributions of this paper are the following:

- A novel technique for the tracking of line segments along continuous sequences based on a *sparse*, *convex* ℓ_1-minimization of geometrical constraints, hence allowing for robust matching under severe appearance variations (see Figure 1).
- A efficient implementation of the proposed method yielding a less computationally demanding line-segment tracker which reduces one of the major drawbacks of working with these features.
- Its validation in our previous point and line features stereo

Machine Perception and Intelligent Robotics (MAPIR) Group, University of Malaga (email: rubengooj@gmail.com). http://mapir.isa.uma.es/.

This work has been supported by the Spanish Government (project DPI2017-84827-R and grant BES-2015-071606) and by the Andalucian Government (project TEP2012-530).

visual odometry system [5], resulting in a more robust VO system under difficult illumination conditions, and also reducing the computational burden of the algorithm.

These contributions are validated with extensive experimentation in several datasets from a wide variety of environments, where we first compare the accuracy and precision of the proposed tracking technique, and then show its performance alongside a VO framework.

II. Related Work

Feature-based motion reconstruction techniques, e.g. VO, visual SLAM, or SfM, are typically addressed by detecting and tracking several geometrical features (over one or several frames) and then minimizing the reprojection error to recover the camera pose. In this context, several successful approaches have been proposed, such as PTAM [6], a monocular SLAM algorithm that relies on FAST corners and SSD search over a predicted patch in a coarse-to-fine scheme for feature tracking. More recently, ORB-SLAM [7] contributed with a very efficient and accurate SLAM system based on a very robust local bundle adjustment stage thanks to its fast and continuous tracking of keypoints for which they relied on ORB features [1]. Unfortunately, even-though binary descriptors are relatively robust to brightness changes, these techniques suffer dramatically when traversing poorly textured scenarios or severe illumination changes occur (see Figure 1), as the number of tracked features drops.

Some works try to overcome the first situation by combining different types of geometric features, such as edges [8], edgelets [9], lines [10], or planes [11].The emergence of specific line-segment detectors and descriptors, such as LSD [3] and LBD [2] allowed to perform feature tracking in a similar way as traditionally done with keypoints. Among them, in [12] authors proposed a stereo VO algorithm relying on image points and segments for which they implement a stereo matching algorithm to compute the disparity of several points along the line segment, thus dealing with partial occlusions. In [5] we contribute with a stereo VO system (PLVO) that probabilistically combines ORB features and line segments extracted and matched with LSD and LBD by weighting each observation with their inverse covariance. In the SLAM context, the work in [13] proposes two different representations: Plücker line coordinates for the 3D projections, and an orthonormal representation for the motion estimation, however, they track features through an optical flow technique, thus the performance with fast motion sequences deteriorates. Unfortunately, the benefits of employing line segments come at the expense of higher difficulties in dealing with them (and they require a high computational burden in both detection and matching stages), and, more importantly, they still suffer from the same issues as keypoints when working with HDR environments.

A number of methods for dealing with varying illumination conditions have been reported. For example, [14] proposed a direct approach to VO, known as DSO, with a joint optimization of both the model parameters, the camera motion, and the scene structure. They used the photometric model of the camera as well as the affine brightness transfer function to account for the brightness change. In [15] authors contributed a robust gradient metric and adjusted the camera setting according to the metric. They designed their exposure control scheme based on the photometric model of the camera and demonstrated improved performance with a state-of-art VO algorithm [16]. Recently, [17] proposed a deep neural network that embeds images into more informative ones, which are robust to changes in illumination, and showed how the addition of LSTM layers produces more stable results by incorporating temporal information to the network. Although those approaches have proven to be effective to moderate changes in illumination or exposure, they would still suffer in more challenging scenarios such as the one in Figure 1.

III. Geometric-based Line Segment Tracking

A. Problem Statement

The first stage of our segment matching algorithm takes as input a pair of images from a stereo video sequence, I_1 and I_2, which can be either from the stereo pair or two consecutive ones in the sequence. Let us define the sets of line segments $\mathcal{L}_1 = \{\mathbf{s}_i,\ \mathbf{e}_i \mid i \in 1, ..., m\}$ and $\mathcal{L}_2 = \{\mathbf{s}_j,\ \mathbf{e}_j \mid j \in 1, ..., n\}$ in I_1 and I_2, where we represent the line segment k by their endpoints $\mathbf{s}_k$ and $\mathbf{e}_k$ in homogeneous coordinates. We also employ the vector of the line

$$\vec{\mathbf{l}}_k = \frac{\mathbf{s}_k - \mathbf{e}_k}{\|\mathbf{s}_k - \mathbf{e}_k\|_2} \tag{1}$$

estimated from the segment endpoints to compare the geometric features of each of them.

Then, given $\mathcal{L}_1$ and $\mathcal{L}_2$, our aim is to find the subset of corresponding line segments between the two input images (see Figure 3), defined as $\mathcal{M}_{12} = \{(l_i, l_j) \mid l_i \in \mathcal{L}_1 \wedge l_j \in \mathcal{L}_2\}$. For l_i and l_j to be a positive match, they must be parallel, have a sufficient ovelap and be compliant with the epipolar geometry of the two views. In order to impose the lines to be parallel, we consider the angle formed by the two line segments in the image plane, θ_{ij}:

$$\theta_{ij} = atan(\ ||\vec{\mathbf{l}}_i \times \vec{\mathbf{l}}_j||\ /\ \vec{\mathbf{l}}_i \cdot \vec{\mathbf{l}}_j\ ). \tag{2}$$

The above-mentioned expressions, however, might lead to inconsistent results as any line in the image could satisfy Equation (2) without being related to the query one. Therefore, we deal with this phenomena by also defining the *overlap* of two line segments $\rho_{ij} \in [0, 1]$ as the ratio between their common parts, as depicted in Figure 2, where ρ_{ij} equals 0 and 1 when there is none or full overlapping between the line segments, respectively. In addition, we also define the ratio between the line lengths as:

$$\mu_{ij} = \frac{max(L_i, L_j)}{min(L_i, L_j)} \tag{3}$$

where $L_k = \|\mathbf{s}_k - \mathbf{e}_k\|_2$ stands for the length of the k-th line, which discards any likely pair of segments whose lengths are not similar enough (if they are of similar length the value of μ_{ij} is close to one, and bigger than one otherwise).

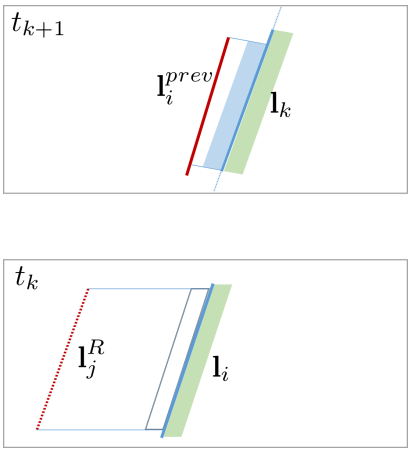


Fig. 2. Scheme of the line segment overlap for both the *stereo* and *frame-to-frame* cases. In the bottom image t_k we plot the stereo overlap between the reference line in the left image $\mathbf{l}_i$ and a match candidate in the right one $\mathbf{l}_j^R$, which is the ratio of the lengths of the shadowed areas (in blue the overlap and in green the line's length). Similarly, the above image t_{k+1} depicts the overlap between the reference line in the second image $\mathbf{l}_k$ and the projected line in the second frame $\mathbf{l}_i^{prev}$.

Finally, we also consider epipolar geometry as a possible constraint for the two different cases of study. In the first case, *stereo* matching, we define the angle formed by the middle point flow vector, $\mathbf{x}_{ij} = \mathbf{m}_i - \mathbf{m}_j$ where the middle point is defined as $\mathbf{m}_k = (\mathbf{s}_k + \mathbf{e}_k)/2$, as:

$$\theta_{ij}^{st} = asin(\|\mathbf{x}_{ij} \times \boldsymbol{\eta}_1\| \,/\, \|\mathbf{x}_{ij}\|) \tag{4}$$

where $\boldsymbol{\eta}_1$ stands for the director vector of the X direction. In contrast, in the *frame-to-frame* case, we assume that images are separated by a small motion and therefore we define the angle formed by $\mathbf{x}_{ij}$ and the Y direction (whose unit vector is given by $\boldsymbol{\eta}_2$), namely:

$$\theta_{ij}^{ff} = asin(\|\mathbf{x}_{ij} \times \boldsymbol{\eta}_2\| \,/\, \|\mathbf{x}_{ij}\|). \tag{5}$$

B. Sparse ℓ_1-Minimization for Line Segment Tracking

In this paper, we formulate line-segment tracking as a sparse minimization problem solely based on the previously introduced geometric constraints. Although this representation has been already employed in computer vision for noise reduction [18], face recognition [19], and loop closure detection [20] (among others), to the best of our knowledge this is the first time it is employed for the geometric tracking of line segment features. For that, we also take advantage of the *1-sparse* nature of the tracking problem, i.e. a single line l_i from the first image should only have at most one match candidate from the $\mathcal{L}_2$ set. It must be noticed that, in the case of detecting divided lines, it is possible for more than one line to match the query one, however, this case is even more likely to occur with appearance based methods, as any locally similar line in the image can be a candidate.

Let us define the n-dimensional *matching* vector $\boldsymbol{\omega}_i$ of the line $l_i \in \mathcal{L}_1$ as:

$$\boldsymbol{\omega}_i = \begin{bmatrix} \omega_{i0} \ldots \omega_{ij} \ldots \omega_{in} \end{bmatrix}^\top \tag{6}$$

where ω_{ij} equals one if l_i and l_j are positive matches and zero otherwise, and n stands for the number of line segments in $\mathcal{L}_2$. Moreover, we define the line segment *error* vectors $\boldsymbol{\beta}_{ij}$ and the objective b for both the *stereo* and *frame-to-frame* cases as:

$$\boldsymbol{\beta}_{ij} = \begin{bmatrix} \theta_{ij} \\ \theta_{ij}^{epip} \\ \rho_{ij} \\ \mu_{ij} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \tag{7}$$

for the line segments $l_i \in \mathcal{L}_1$ and $l_j \in \mathcal{L}_2$, where *epip* refers to the epipolar constraints defined in Equations (4) and (5) for the two cases of study.

Now, by concatenating all line segment error vectors we form the $4 \times n$ matrix $\mathbf{A}_i$:

$$\mathbf{A}_i = \begin{bmatrix} \boldsymbol{\beta}_{i0}, \ \ldots \boldsymbol{\beta}_{ij}, \ \ldots \boldsymbol{\beta}_{in} \end{bmatrix}. \tag{8}$$

that must satisfy the linear constraint $\mathbf{A}_i\boldsymbol{\omega}_i = \mathbf{b}$ if the sum over all the components from the matching vector $\boldsymbol{\omega}_i$ is one (which is our hypothesis). While ℓ_2-norm is usually employed to solve the previous problem with the typical least-squares formulation, it is worth noticing that it leads to a dense representation of the optimal $\boldsymbol{\omega}_i^*$, which contradicts the 1-sparse nature of our solution.

In contrast, we can formulate the problem of finding $l_j \in \mathcal{L}_2$ that properly matches $l_i \in \mathcal{L}_1$ as a *convex*, *sparse*, *constrained* ℓ_1-minimization as follows:

$$\min_{\boldsymbol{\omega}_i} \ \|\boldsymbol{\omega}_i\|_1 \ \text{ subject to } \ \|\mathbf{A}_i\boldsymbol{\omega}_i - \mathbf{b}\|_2 \leq \epsilon \tag{9}$$

where the constraint corresponds to the above-mentioned geometrical conditions, and $\epsilon > 0$ is the maximum tolerance for the constraint error. Moreover, the problem in Equation (9) can be also solved with the homotopy approach [21] in the following *unconstrained* manner:

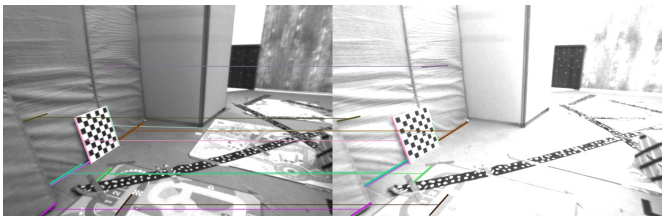
$$\min_{\boldsymbol{\omega}_i} \ \lambda \|\boldsymbol{\omega}_i\|_1 \ + \ \frac{1}{2} \|\mathbf{A}_i\boldsymbol{\omega}_i - \mathbf{b}\|_2 \tag{10}$$

with λ a weighting parameter empirically set to 0.1, resulting in a very effective and fast solver [22].

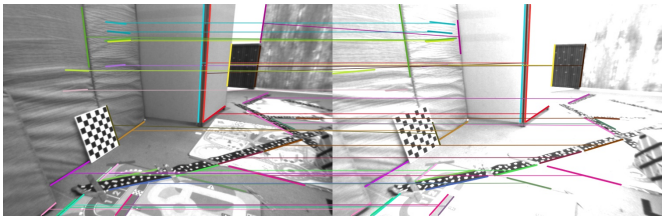
Then, we efficiently solve the problem in Equation (10) for each $l_i \in \mathcal{L}_1$ obtaining the sparse vector $\boldsymbol{\omega}_i$, which after being normalized indicates whether the line segment l_i has a positive match (in the maximum entry j of $\boldsymbol{\omega}_i$). Finally, we guarantee that line segments are uniquely corresponded by only considering the candidate with minimum error, defined as $\|\boldsymbol{\beta}_{ij}\|$, if the error for the second best match is at least 2 times bigger than the best one. For further details on the mathematics of this Section, please refer to [21].

C. Dealing with Outliers

When dealing with repetitive structures a number of outliers can appear. To deal with this problem in the *stereo* case, we implement a filter based on the epipolar constraint, for which we first estimate robustly the normal distribution formed by the angles with the horizontal direction. Then, we discard the matches whose angle with the horizontal direction



(a) LSD [3] + LBD [2]



(b) LSD [3] + Our proposal

Fig. 3. Stereo correspondences between two different images from the EuRoC dataset. Our matching algorithm is capable of finding matches that does not necessarily have similar appearance.

lies above 2 times the standard deviation of the distribution formed by all matches.

In the *frame-to-frame* case, as the camera pose is not known yet, we cannot directly apply epipolar geometry. However, we approximate an epipolar filter, based on the assumption that input images belong to consecutive frames from a sequence, and therefore they are separated by small motions. For that, we discard the matches whose angle with the vertical direction (this is the epipolar constraint in the case of null motion) lies above 2 times the standard deviation of the distribution formed by all matches as they are less likely to fullfil the motion constraints.

IV. Point-Segment Visual Odometry Overview

In this section we briefly describe the PLVO stereo visual odometry system [5] where the proposed matching algorithm has been integrated for line segment tracking. PLVO combines probabilistically both point and line segment features and its C++ implementation is available publicly `https://github.com/rubengooj/StVO-PL`.

1) Point Features: In PLVO points are detected and described with ORB [1] (consisting of a FAST keypoint detector and a BRIEF descriptor) due to its efficiency and good performance. In order to reduce the number of outliers, we only consider the measurements that are mutual best matches, and also check that the two best matches are significantly separated in the description space by only accepting matches whose distance between the two closest correspondences is above the double of the distance to the best match.

2) Line Segment Features: In our previous work [5] we detect line segments with the Line Segment Detector (LSD) [3] and also employ the Line Band Descriptor (LBD) [2] for the stereo and frame-to-frame matching. Although this method provides a high precision and repeatability, it still presents very high computational requirements, simple detection and matching requires more than 30ms with 752×480, thus its use limits their application in real-time. In order to reduce the computational burden of the Stereo VO system, in this work we have also employed the Fast Line Detector (FLD) [23], which is based on connecting collinear Canny edges [24]. This detector works faster than LSD at expense of a poorer performance in detecting meaningful lines, i.e. lines with strong local support along all the lines, since, unlike LSD, detection is only based on image edges.

3) Motion Estimation: After obtaining a set of point and line correspondences, we then recover the camera motion with iterative Gauss-Newton minimization of the projection errors in each case (in the case of line segments we employ the distance from the projected endpoint, to the line in the next frame). To mitigate the undesirable effect of outliers and noisy measurements, we perform a two steps minimization for which we weight the observations with a Pseudo-Huber loss function, and then we remove the outliers and refine the solution.

V. Experimental Validation

We evaluate the performance and robustness of our proposal in several public datasets for two different line segment detectors, LSD [3] and FLD [23], when employing two different matching strategies: our proposal, and traditional appearance based tracking with LBD [2]. All the experiments have run on an Intel Core i7-3770 CPU @ 3.40 GHz and 8GB RAM without GPU parallelization. In our experiments we have employed a fixed number of detected lines set to 100, and 600 ORB [1] features for the case of points and line based VO.

A. Tracking Performance

First, we compare the line segment tracking performance of our proposal against traditional feature matching approaches. For that, we took several sequences (at different speeds) and classified each match as an inlier if the correspondent line segment projection error is less than one pixel when employing the groundtruth transformation. In order to compare the algorithms under dynamic illumination changes, we have employed two specific datasets: one extracted from [4] (*hdr*) taken with an RGB-D sensor under HDR situations, and another one from our previous work [25] (*dnn*) containing a number of difficult dynamic illumination conditions. In addition, we also have employed the Tsukuba Stereo Dataset [26] , a synthetic dataset rendered under 4 different illuminations, i.e. *fluorescent*, *lamps*, *flashlight*, and *daylight*. For a more challenging set of experiments, we have also employed to use all combinations (taking *fluorescent* as reference) of the rendered sequences, by setting the left one to the reference and the right one to all different possibilities. It is worth noticing that illumination changes from the considered datasets are produced punctually, and after that, the scene illumination usually keeps constant until the next change. This benefits to descriptor-based techniques when evaluating the tracking performance during the whole sequence, for

TABLE I

TRACKING PERFORMANCE OF OUR PROPOSAL AND TRADITIONAL LINE SEGMENT FEATURE MATCHING (NUMBER OF MATCHES - INLIERS).

Dataset	Resolution	LSD + LBD		LSD + L1		FLD + LBD		FLD + L1	
hdr/bear	640×480	72	80 %	39	94 %	65	73 %	45	86 %
hdr/desk	640×480	70	84 %	40	93 %	60	83 %	41	88 %
hdr/floor1	640×480	27	77 %	22	85 %	29	77 %	33	85 %
hdr/floor2	640×480	71	80 %	42	93 %	61	80 %	41	87 %
hdr/sofa	640×480	58	81 %	41	90 %	49	81 %	45	84 %
hdr/whiteboard	640×480	41	76 %	35	95 %	42	75 %	35	90 %
hdr/flicker1	640×480	42	84 %	45	98 %	44	80 %	40	95 %
hdr/flicker2	640×480	86	89 %	45	98 %	74	85 %	45	97 %
dnn/1-light	752×480	16	94 %	15	100 %	19	90 %	18	100 %
dnn/2-lights	752×480	29	94 %	18	100 %	38	92 %	24	97 %
dnn/3-lights	752×480	49	90 %	35	100 %	57	91 %	35	97 %
dnn/change-light	752×480	19	85 %	14	100 %	24	95 %	18	100 %
dnn/hdr1	752×480	55	90 %	35	100 %	51	88 %	35	95 %
dnn/hdr2	752×480	50	90 %	32	97 %	52	93 %	39	95 %
dnn/overexp	752×480	84	94 %	55	98 %	75	92 %	47	96 %
dnn/overexp-change-light	752×480	82	92 %	50	98 %	72	90 %	43	96 %
dnn/low-texture	752×480	53	88 %	35	100 %	52	90 %	35	97 %
dnn/low-texture-rot	752×480	43	90 %	30	100 %	40	90 %	30	95 %
tsukuba	640×480	43	86 %	34	84 %	36	75 %	21	86 %
tsukuba/fluor(L)-daylight(R)	640×480	5	40 %	15	80 %	5	40 %	12	80 %
tsukuba/fluor(L)-flashlight(R)	640×480	2	33 %	5	60 %	2	50 %	4	50 %
tsukuba/fluor(L)-lamps(R)	640×480	1	0 %	5	60 %	1	0 %	3	67 %

which we also recommend to watch the attached video for visual evaluation under such circumstances.

Table I shows the tracking accuracy and the number of features tracked, for all the sequences from each considered dataset. First, we observe a slightly inferior performance of FLD [23] in comparison against LSD [3], due to its lower repeatability in contrast with its superior computational performance (see Table III). In general, we observe that our matching method decreases the number of features, due to the very requiring assumptions of our matching technique, however, it provides a higher ratio of inliers thanks to the extra stage explained in Section III.

As for the Tsukuba dataset, we observe that the number of features successfully tracked dramatically decreases as the response of the detectors is not capable of producing a compatible set of lines from the same images. However, we observe that our method technique is capable of recovering more matches, specially in the less challenging case (*fluorescent* and *daylight*), that can be employed along different sensing to extract more information from the environment in such difficult situations.

B. Robustness Evaluation in Stereo Visual Odometry

In this set of experiments, we test the performance of the compared algorithms in the EuRoC [27] dataset. In order to simulate changes in exposure time or illumination within the EuRoC dataset [27] (we will refer to simulated sequences with an asterisk) we change the gain and bias of the image with two uniform distribution, i.e. $\alpha = \mathcal{U}(0.5, 2.5)$ and $\beta = \mathcal{U}(0, 20)$ pixels every 30 seconds. For that comparison, we not only focus in the accuracy of the estimated trajectories, but also in the robustness of the algorithms under different environment conditions (we mark a dash those experiments where the algorithm lose the track). We compare the accuracy of trajectories obtained with our previous stereo VO system, PLVO [5], against our proposal tracking strategy, PLVO-L1, when employing LSD or FLD features.

TABLE II

RELATIVE RMSE ERRORS IN THE EUROC MAV DATASET [27].

Sequence	LVO (FLD)	LVO-L1 (FLD)	LVO (LSD)	LVO-L1 (LSD)
MH-01-easy	0.0641	0.0788	0.0669	0.0716
MH-02-easy	0.0826	0.0923	0.0740	0.0881
MH-03-med	0.0886	0.1011	0.0898	0.1004
MH-04-diff	0.1500	0.1536	0.1429	0.1518
MH-05-diff	0.1350	0.1529	0.1391	0.1561
V1-01-easy	0.0890	0.0969	0.0876	0.0954
V1-02-med	0.0662	0.0847	0.0606	0.0947
V1-03-diff	0.2261	0.1518	0.0765	0.1103
V2-01-easy	0.1980	0.1868	0.1662	0.1898
V2-02-med	0.1634	0.2294	0.1982	0.2562
V2-03-diff	0.2329	0.2342	0.2354	0.2275
MH-01-easy*	0.0787	0.0897	0.0741	0.0728
MH-02-easy*	0.0873	0.1015	0.8237	0.0981
MH-03-med*	0.0982	0.1578	0.0916	0.1141
MH-04-diff*	0.1540	0.1780	0.1354	0.1621
MH-05-diff*	-	0.1603	-	0.1863
V1-01-easy*	0.0880	0.1011	0.0997	0.1041
V1-02-med*	0.0858	0.0953	0.0713	0.1096
V1-03-diff*	-	0.2087	-	0.1598
V2-01-easy*	-	0.2396	-	0.2080
V2-02-med*	-	0.2472	-	0.2563
V2-03-diff*	-	-	-	0.2631

Table II contains the results by computing the relative RMSE in translation for the estimated trajectories. As we can observe, in the raw dataset our approach performs slightly worse than standard appearance-based tracking techniques, mainly due to the lower number of correspondences provided by our algorithm, as mentioned in previous Section. In contrast, we can observe a considerable decrease in accuracy of our approaches, however, they are capable of estimating the motion in all sequences with an lower accuracy, mainly due to the less number of matches, due to restrictive constraints. For this reason we believe our matching technique a

TABLE III

COMPARISON OF THE COMPUTATIONAL PERFORMANCE OF THE DIFFERENT CONSIDERED ALGORITHMS.

	Monocular Tracking	Stereo Tracking
LSD + LBD	39.342 ms	51.347 ms
LSD + Our	25.897 ms	35.828 ms
FLD + LBD	18.147 ms	33.266 ms
FLD + Our	7.654 ms	23.445 ms

suitable option to address the line segment tracking problem under severe appearance changes, in combination with prior information from different sensors and/or algorithms.

C. Computational Cost

Finally, we compare the computational performance of the different tracking algorithms in the considered datasets considering the time of processing one image (similarly to the VO framework). In the both cases we can observe the superior performance of our proposal, it runs between 1.5 and 2 times faster depending on the detector employed, thanks to the efficient implementation of the geometric-based tracking thus making it very suitable for robust real-time application, most likely in combination with other sensing, such as inertial measurement unit sensors (IMU).

VI. CONCLUSIONS

In this work, we have proposed a geometrical approach for the robust matching of line segments for challenging stereo streams, such as sequences including severe illumination changes or HDR environments. For that, we exploit the nature of the matching problem, i.e. every observation can only be corresponded with a single feature in the second image or not corresponded at all, and hence we state the problem as a sparse, convex, ℓ_1-minimization of the matching vector regularized by the geometric constraints. Thanks to this formulation we are able of robustly tracking line segments along sequences recorded under dynamic changes in illumination conditions or in HDR scenarios where usual appearance-based matching techniques fail. We validate the claimed features by first evaluating the matching performance in challenging video sequences, and then testing the system in a benchmarked point and line based VO algorithm showing promising results.

REFERENCES

[1] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2564–2571, IEEE, 2011.

[2] L. Zhang and R. Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency," *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 794–805, 2013.

[3] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: A fast line segment detector with a false detection control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722–732, 2010.

[4] S. Li, A. Handa, Y. Zhang, and A. Calway, "HDRFusion: HDR SLAM using a low-cost auto-exposure RGB-D sensor," in *3D Vision (3DV), 2016 Fourth International Conference on*, pp. 314–322, IEEE, 2016.

[5] R. Gomez-Ojeda and J. Gonzalez-Jimenez, "Robust stereo visual odometry through a probabilistic combination of points and line segments," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 2521–2526, IEEE, 2016.

[6] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pp. 225–234, IEEE, 2007.

[7] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, 2017.

[8] E. Eade and T. Drummond, "Edge landmarks in monocular SLAM," *Image and Vision Computing*, vol. 27, pp. 588–596, apr 2009.

[9] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect visual odometry for monocular and multicamera systems ," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2017.

[10] A. Bartoli and P. Sturm, "Structure-from-motion using lines: Representation, triangulation, and bundle adjustment," *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 416–441, 2005.

[11] L. Ma, C. Kerl, J. Stückler, and D. Cremers, "CPA-SLAM: Consistent plane-model alignment for direct RGB-D SLAM," in *ICRA 2016*, pp. 1285–1291, IEEE, 2016.

[12] T. Koletschka, L. Puig, and K. Daniilidis, "MEVO: Multi-environment stereo visual odometry," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 4981–4988, IEEE, 2014.

[13] G. Zhang, J. H. Lee, J. Lim, and I. H. Suh, "Building a 3-D Line-Based Map Using Stereo SLAM," *IEEE Transactions on Robotics*, vol. 31, no. 6, pp. 1364–1377, 2015.

[14] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[15] Z. Zhang, C. Forster, and D. Scaramuzza, "Active Exposure Control for Robust Visual Odometry in HDR Environments," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2017.

[16] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 15–22, IEEE, 2014.

[17] R. Gomez-Ojeda, Z. Zhang, J. Gonzalez-Jimenez, and D. Scaramuzza, "Learning-based Image Enhancement for Visual Odometry in Challenging HDR Environments," 2017.

[18] M. Elad, M. A. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.

[19] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with ℓ^1-graph for image analysis," *IEEE transactions on image processing*, vol. 19, no. 4, pp. 858–866, 2010.

[20] Y. Latif, G. Huang, J. Leonard, and J. Neira, "Sparse optimization for robust and efficient loop closing," *Robotics and Autonomous Systems*, vol. 93, pp. 13–26, 2017.

[21] M. S. Asif, *Primal Dual Pursuit: A homotopy based algorithm for the Dantzig selector*. PhD thesis, Georgia Institute of Technology, 2008.

[22] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ^1-norm minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.

[23] J. H. Lee, S. Lee, G. Zhang, J. Lim, W. K. Chung, and I. H. Suh, "Outdoor place recognition in urban environments using straight lines," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp. 5550–5557, IEEE, 2014.

[24] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[25] R. Gomez-Ojeda, Z. Zhang, J. Gonzalez-Jimenez, and D. Scaramuzza, "Learning-based image enhancement for visual odometry in challenging HDR environments,"

[26] M. Peris, S. Martull, A. Maki, Y. Ohkawa, and K. Fukui, "Towards a simulation driven stereo vision system," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 1038–1042, IEEE, 2012.

[27] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.