Multiscale line segment detector for robust and accurate SfM — **Source link** ↗

Yohann Salaün, Renaud Marlet, Pascal Monasse

Institutions: École Normale Supérieure

Related papers:

- LSD: a Line Segment Detector

- Robust segmentation of relevant regions in low depth of field images

- Flexible depth segmentation method using phase-shifted wrapped phase sequences

- Bilayer Segmentation of Live Video

- Object-oriented hybrid segmentation using stereo images

Multiscale line segment detector for robust and accurate SfM

Yohann Salaün, Renaud Marlet, Pascal Monasse

Multiscale line segment detector for robust and accurate SfM

Yohann Salaün*†, Renaud Marlet* and Pascal Monasse*
* LIGM, UMR 8049, École des Ponts, UPE, Champs-sur-Marne, France
† CentraleSupélec, Châtenay-Malabry, France
Email: yohann.salaun@enpc.fr, renaud.marlet@enpc.fr, pascal.monasse@enpc.fr

Abstract—We propose a multiscale extension of a well-known line segment detector, LSD. We show that its multiscale nature makes it much less prone to over-segmentation, more robust to low contrast and less sensitive to noise, while keeping the parameterless advantage of LSD and still being fast. Moreover, we show that in scenes with little or no feature points, but where it is however possible to perform structure from motion from matched line segments, the accuracy is significantly improved. This provides an objective and automatic quantitative assessment of our detector that goes much beyond the usual qualitative visual inspection found in the literature.

I. INTRODUCTION

In many common settings, Structure from Motion (SfM) techniques based on detected salient points are reliable enough to estimate the relative pose of cameras and reconstruct a 3D point geometry. However these techniques may fail or yield inaccurate results on low-textured scenes, such as indoor environments with large uniform wall surfaces, which have not, or not enough salient points. Another failure case is with buildings with large glass areas, where reflection on windows invalidates point matching. Mostly planar scenes may also lead to degenerate settings for point-based camera calibration.

To address these issues, a number of authors advocate the use of lines rather than points in such situations. Indeed, lines can generally be detected even in presence of uniform areas, at edges. Lines at window edges are also robust to reflection variations when the viewpoint changes, and calibration with lines naturally deals well with scenes that may be planar. As a matter of fact, a few line-based SfM methods [2], [9], [13] have recently been developed, that outperform classical point-based methods on this kind of scenes, while not relying on segment endpoints, which are too uncertain with all known segment detectors.

The argument that lines are useful for calibration and 3D reconstruction is heavily hammered home by authors that propose new lines detectors and/or new line matchers [7], [11], [12]. However, their validation is often "visual", i.e., qualitative based on displayed detections or matches. When quantitative data is provided, it only evaluates quality indirectly and unreliably, possibly with an arbitrary bias due to a human assessment. Typical such measures include the number of detected lines, supposed to be low as an indicator for little over-segmentation [7], possibly using synthetic data (sketched lines) where the number of lines is known. Another measure is a human count of the number of matches [11],
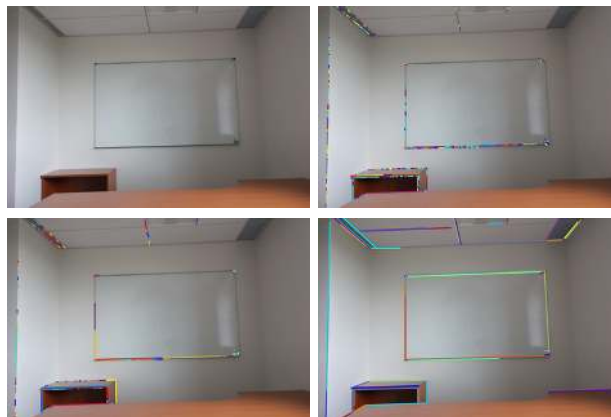

Fig. 1: Original 15 Mpx picture (top left); lines detected with LSD [6] (top right), EDLines [1] (bottom left), our method (MLSD, bottom right).

[12], which can be done only at a small scale, with an arbitrary notion of perceived lines and possibly a penalized score for over-segmentation [7]. To evaluate robustness, the same assessments can be performed with noise [5], [6] or other image transformations [11].

Although it is presented as one of their main motivation, these papers do not evaluate the quality of their line detector or matcher w.r.t. image registration and 3D reconstruction. Yet our experiments show that, indeed, line-based SfM is sensitive to the quality of detections and matches. Detectors and matchers that put forward SfM as an application should thus be crafted w.r.t. an actual *quantitative* SfM evaluation, not just a *qualitative* assessment, or indirect/unreliable measurements.

In this paper, we propose a new, enhanced line segment detector that is based on a multiscale exploration (MLSD). It detects some lines that fail to be detected by other detectors while reducing over-segmentation. Besides, it detects lines more accurately. We show, using not only qualitative but also quantitative results, that our detector does improve both the robustness and the accuracy of line-based SfM, i.e., calibration and reconstruction. For this, we compare our method to state-of-the-art line detectors, namely LSD [6] and EDLines [1].
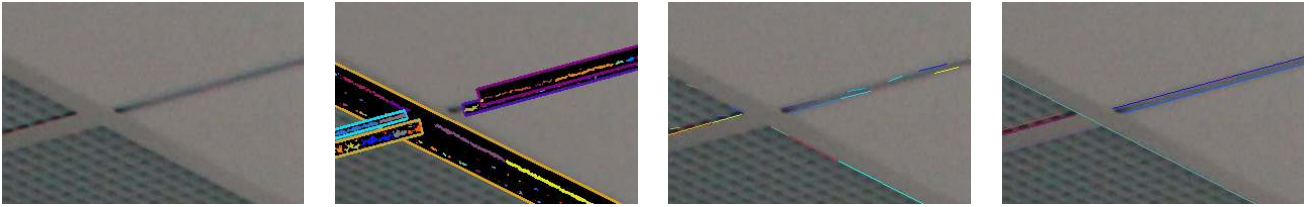
Fig. 2: From left to right: crop of the original high-resolution picture (suspended ceiling), multiscale clusters, detections with LSD, and final detections with MLSD. Colored rectangles filled with black represent coarse segments upsampled to the finest scale. Colored clusters inside are connected components of pixels with compatible gradient directions.

II. RELATED WORK

LSD [6] and EDLines [1] are currently state-of-the-art methods for line detection and are thus used in the mentioned SfM approaches [2], [9], [13]. They are based on an *a contrario* theory that automatically defines if a line is meaningful or not with a score called Number of False Alarms (NFA).

In LSD [6], segments are detected as connected components of pixels with similar gradient direction. The similarity criterion not being transitive, the authors use a greedy method to compute these components. However, because of noise or lack of contrast, some components may be separated by a few pixels and the detector then yields several shorter segments. This causes an over-segmentation phenomenon, especially with high resolution images (see Fig. 1).

Actually, both methods [1], [6] tend to compute lots of short segments ($\leq 5\%$ of the image size), which turns out to be redundant and noisy information for the matching and calibration methods that follow. The accuracy of these methods thus does not improve with high quality pictures.

Other authors [7] propose a multiscale method to detect lines without over-segmenting them. However, they use many different thresholds to merge their segments, based on parameters such as direction similarity, distance between endpoints and lines, line proximity, etc. If these parameters make sense, they can easily fail when used on different pictures, e.g., in the presence of slight curves, grid patterns, close parallel lines.

The absence of parameters in *a contrario* methods is the main reason for the quality of their results compared to others, such as Hough-based methods. They do not depend on parameter tuning and are consistently as efficient for different kinds of scenes. In this paper, we present a multiscale method to detect and arrange line segments at different scales. However, we stick to the *a contrario* framework to automatically define which segments should be merged. Although one LSD article [5] presents some experiments showing the influence of noise and the importance of detecting at lower resolutions, these were just illustrative and did not specify a way to choose the right resolution for detection and how to avoid detection redundancy across different scales.

In Sect. III, we propose a new multiscale segment detector based on LSD. In Sect. IV, we present qualitative and quantitative results. Finally, we draw some conclusions in Sect. V.

III. MULTISCALE SEGMENT DETECTOR

A. Number of False Alarms for segments and multi-segments

We use the same background model as generally used in *a contrario* methods: as in LSD, this model assumes that gradient directions are independent and uniformly distributed within the image. Significant deviations from this "null hypothesis" are deemed relevant and reveal the presence of a line segment, hence the *a contrario* principle of the approach.

We consider that a segment s is defined by a rectangle, and can thus have a width greater than 1 pixel. We note $|s|$ the area of s, p the probability that a random pixel q has the same direction $\theta(q)$ (orthogonal to its gradient) than s up to an angular tolerance of πp, and k_s the number of pixels inside the rectangle that have the same direction than s up to πp.

In LSD, the NFA of a segment s is defined for a given angular tolerance πp as

$$\text{NFA}(s,p) = \gamma N_L \mathcal{B}(|s|, k_s, p) \tag{1}$$

where N_L is the number of possible segments in the image, γ is a normalizing value (number of different tested values for p), and $\mathcal{B}(|s|, k_s, p)$ is the tail of the binomial distribution:

$$\mathcal{B}(|s|, k_s, p) = \sum_{j=k_s}^{|s|} \binom{|s|}{j} p^j (1-p)^{|s|-j}. \tag{2}$$

A segment is considered meaningful if and only if its NFA is below a given threshold (whose "natural" value is 1). In LSD [6], the authors show that the threshold value can be changed without significant differences for the detections.

Besides, Grompone von Gioi et al. [4] define an NFA for lines which is a 1D version of (1), as they only look for clusters of pixels that are perfectly aligned. They also define an $\text{NFA}_\mathcal{M}$ for multiple lines, to compare the meaningfulness of segmented lines to their merged version. We generalize this formula to a set of n segments $S = \{s_1, ..., s_n\}$ inside a sub-area of the picture of size $N \times M$:

$$\text{NFA}_\mathcal{M}(S,p) = \gamma N_L \binom{(NM)^{\frac{5}{2}}}{n} \prod_{i=1}^{n} (|s_i| + 1) \mathcal{B}(|s_i|, k_{s_i}, p) \tag{3}$$

where $(NM)^{\frac{5}{2}}$ approximates the number of possible segments in an area of size $N \times M$, as in LSD [6].

Given n segments $S = \{s_1, ..., s_n\}$, let $Seg(\cup_{i=1}^n s_i)$ be the best segment computed from the union of the clusters s_i,

defined as the smallest rectangle that contains the rectangles associated to all segments s_i. The corresponding fusion score of the set of segments is defined as:

$$\mathcal{F}(s_1, ..., s_n) = \log \left(\frac{\text{NFA}_\mathcal{M}(s_1, ..., s_n, p)}{\text{NFA}_\mathcal{M}(Seg(\cup_{i=1}^n s_i), p)} \right). \quad (4)$$

If the fusion score is positive, it means that the $\text{NFA}_\mathcal{M}$ is lower for the merged segment than for the segments individually, and thus that they should be merged. This defines a criterion for segment merging that does not rely on any parameter.

B. Multiscale processing

Input: Image I
Output: Set of segments $\mathcal{S}$
for $k = 0$ **to** K **do**
    Compute downscaled image I^k
    $\mathcal{S}^k \leftarrow \emptyset$
    if $k \neq 0$ **then**
        Upscale $\mathcal{S}^{k-1}$
        forall the $s_i \in \mathcal{S}^{k-1}$ **do**
            Compute the connected components of $\mathcal{P}_i^k$ (5)
            Merge w.r.t. fusion score (4) with greedy strategy
            Add in $\mathcal{S}^k$ the meaningful segments if any, else s_i
    Add in $\mathcal{S}^k$ the segments detected with LSD in I^k
    forall the $s_i \in \mathcal{S}^k$ *in increasing NFA order* **do**
        Find aligned neighbors
        Merge w.r.t. fusion score (4)
return $\mathcal{S}^K$

Algorithm 1: Multiscale Line Segment Detector

As observed in [5], the size of the picture matters a lot for segment detection. A smaller version of the same picture often yields fewer but proportionally longer segment detections, while the original picture may yield a lot of fragmented segments. We thus use a multiscale approach to find long segments at coarse scales, refining their location at a finer scale. At each scale, as described below, we consider new segment candidates at the same locations than segments at the previous scale, possibly merge them using the multi-segment criterion (4), and keep the resulting segments that pass the NFA condition (1). In the following, we denote with an upper index the scale of the image, from 0 (coarsest) to K (finest): $I = I^K$ is the original image and I^k is its downscaled version by a factor 2^{K-k} in both width and height. Following LSD [6], we also convolve with a Gaussian kernel to smooth each image I^k before any further detection.

At the coarsest scale, we only detect segments classically with LSD. At subsequent scales, we first use information from the previous scale to find new segments. Given a coarse segment s_i^{k-1} of direction $\theta(s_i^{k-1})$ detected with some angular tolerance πp_i^{k-1}, we define $\mathcal{A}_i^k$ as the rectangular area of s_i^{k-1}

upscaled in I^k, and $\mathcal{P}_i^k$ as the subset of pixels in $\mathcal{A}_i^k$ that have the same direction as s_i^{k-1} up to πp_i^{k-1}:

$$\mathcal{P}_i^k = \left\{ q \in \mathcal{A}_i^k \text{ s.t. } |\theta(q) - \theta(s_i^{k-1})|_{(\text{mod } \pi)} < \pi p_i^{k-1} \right\}. \quad (5)$$

We then compute the set $\mathcal{C}_i^k$ of all the connected components in $\mathcal{P}_i^k$, using 8-neighborhood connectivity. These potential new segments are illustrated in Fig. 2. Different cases can occur, e.g., the components can belong to the same line, or to parallel and close lines that were merged at a coarser scale.

As a set $\mathcal{C}_i^k$ can contain many tentative segments, we cannot test all the possible fusion combinations to find the best ones. We thus resort to a greedy strategy to iterate over potential segments to merge. Beginning with the component c in $\mathcal{C}_i^k$ with lowest NFA (i.e., the most meaningful), we compute all the other components that are sufficiently aligned with it:

$$\mathcal{I}(c) = \{c' \in \mathcal{C}_i^k \backslash \{c\} \text{ s.t. } l(c) \cap c' \neq \emptyset\} \quad (6)$$

where $l(c)$ is the line passing through the center of c, with angle $\theta(c)$. We compute the fusion score of $\mathcal{I}(c)$ and, if positive, we replace the subsegments in $\mathcal{I}(c)$ by their merged version. This algorithm is iterated until all tentative segments c in $\mathcal{C}_i^k$ have been tested for fusion.

Finally, we compute the usual NFA on all these segments to only keep the meaningful ones. For a given coarse segment s_i^{k-1}, it may happen, e.g., with noise and/or when contrast is low (see Fig. 3), that no derived segment at the current finer scale is kept because none meets the NFA condition. In this case, we keep the original coarse segment, with scale information, but no longer try to find finer segments at the same locations at finer scales.

Line segments are then detected with LSD at current scale.

C. Post-detection merging

The multiscale processing helps finding low-contrast segments and merging some of the over-segmentations. However, as the segments used for fusion are detected at coarse scales, there still remain segmented parts. We thus add a refinement step that tries to merge segments after each multiscale detection, similar to the inter-scale merging above (Sect. III-B).

For each detected segment, in increasing NFA order, we find the neighboring segments with similar alignments. For this, we extend the segment to its corresponding line and find the intersected segments with similar directions as in (6). They are merged if the corresponding fusion score (4) is positive.

The whole algorithm is summarized in Algo. 1.

IV. EXPERIMENTS

In this section, we compare our detector (MLSD) to LSD [6] and EDLines [1]. For this, we use the following datasets:

- **Strecha et al.'s**: a standard dataset used for SfM, containing several outdoor scenes (e.g., castle courtyard), with a picture resolution of 6 Mpx [10]. The camera parameters are known and a LiDAR ground truth is available for a couple of scenes.
- **Hall**: a dataset of a hall building from [3] with a picture resolution of 6 Mpx.

Fig. 3: From left to right: detection upon a low contrast area for LSD [6], EDLines [1] and our method. Notice that our method does not suffer from over-segmentation and that upper and lateral borders of the white board are detected.

Method Dataset		LSD [6]	EDLines [1]	MLSD (ours)
Castle [10]	# segments	5996	4568	4112
	average length	26.3	31.9	51.5
	total length	158k	146k	**212k**
	run. time	3.9	**0.7**	9.1
Office	# segments	5708	2849	1683
	average length	44.4	88.9	190.0
	total length	253k	253k	**320k**
	run. time	6.0	**1.4**	14.5
Hall	# segments	4291	2996	1787
	average length	26.8	39.1	86.2
	total length	115k	117k	**154k**
	run. time	2.1	**0.7**	4.3

TABLE I: Output of different line detectors: number of detected segments, average length (in pixels), total length (in pixels), computation time (in seconds).

- **Office**: our own dataset of an office room with a picture resolution of 18 Mpx. It contains a calibration ground truth obtained by taking auxiliary pictures at the same camera positions but with a video-projected pattern rich in feature points, allowing to perform accurate structure from motion based on point matches [8].
- **Building**: our own dataset of urban pictures (18 Mpx).

A. Visual results

We detected segments using each method on all datasets. Fig 4 illustrates some of the detections. Comparing the first two rows to the last two, we can observe that our method is especially efficient for high resolution pictures (Office and Building), where it merges many segments that belong to the same line. It also better detects low-contrast lines where noise and shadows make it difficult for classical detection methods.

Although visual inspection shows the quality of our results, it cannot be measured objectively this way. We also give in Table I a quantitative comparison of these methods, reporting the number of segments, their average length and the sum of

Method Dataset	LSD [6]	EDLines [1]	MLSD (ours)
Strecha et al.'s [10]	0.47	1.32	**0.29**
Office	2.77	2.02	**0.72**

TABLE II: Average rotation error (in °) of line-based external calibration [2] for different detectors.

their length. We observe that our method tends to detect fewer but significantly longer segments than other methods, hence being less prone to over-segmentation. The visual inspection of Fig. 4 confirms that our method tends to give a better structural information compared to the others. The multiscale processing and fusion process have an acceptable computation overhead: they double or so the time of LSD, which is a fast algorithm with an online, real-time demonstrator [6].

B. Calibration results

To compare quantitatively the different line segment detectors, we compare their impact on the accuracy of line-based external calibration. For this, we use an automatic line matcher, LBD+S&G [12]. Matched segments are then provided as input to a line-based SfM algorithm [2] that computes the relative rotations between pairs of images. It relies on vanishing points to identify projections of 3D parallel lines and assumes a Manhattan world, with orthogonal principal vanishing points. It uses a minimal solver using two parallel lines and one line in an orthogonal direction, which is embedded in a RANSAC procedure to ensure robustness w.r.t. outliers.

The average rotation error between pairs of images for the datasets with calibration ground truth are reported in Table II. The segments detected by MLSD yield a significantly better accuracy than the segments detected by LSD or EDLines.

C. 3D reconstruction results

Strecha et al.'s dataset [10] also contains a LiDAR point cloud used as a ground truth for reconstruction. We project all the matching segments computed by the different detectors
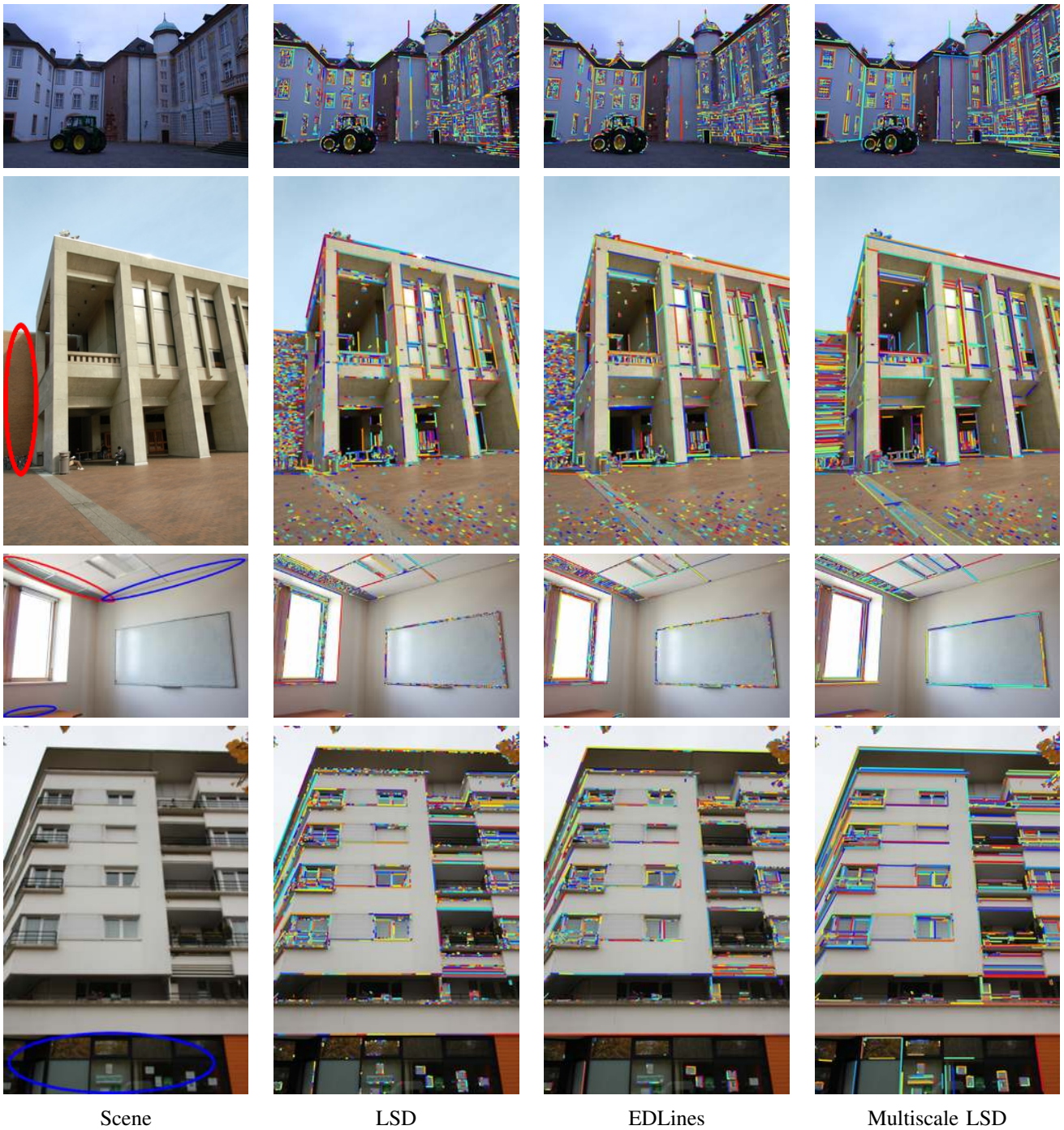
Fig. 4: Visual results for different line segment detectors and datasets. Top to bottom: Castle (from Strecha et al.'s), Office, Hall, Building. Left to right: original image and detections with LSD [6], EDLines [1] and our multiscale LSD (MLSD). Note that our method also merges segments following a grid pattern (see red ellipse areas) and better detects segments in areas with high noise or low contrast (see blue ellipses).
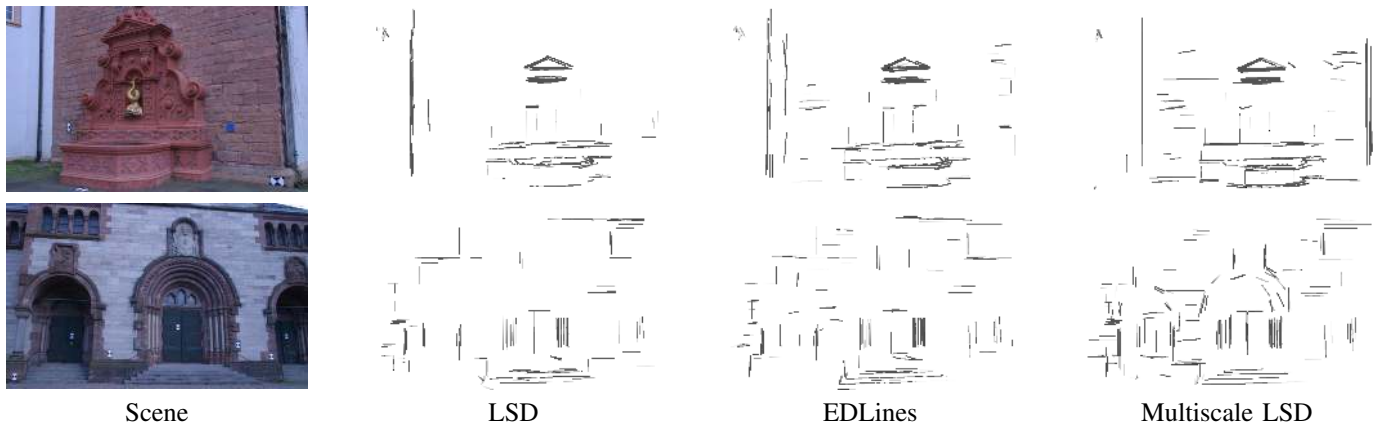
Fig. 5: Fronto-parallel 3D reconstructions of matched segments (only inliers according to the ground truth) on scenes Fountain-P11 (top) and Herz-Jesu-P8 (bottom) from Strecha et al.'s. The tested methods are LSD [6], EDLines [1] and the proposed multiscale LSD. Notice that our method has more relevant segments than the other two.

	Distance (m)	1	0.1	0.01	0.001
Dataset					
Herz-Jesu-P8 [10]	LSD [6]	246	214	174	118
	EDLines [1]	309	244	172	97
	MLSD	**357**	**287**	**230**	**137**
Fountain-P11 [10]	LSD [6]	268	230	180	124
	EDLines [1]	319	267	177	97
	MLSD	**372**	**297**	**214**	**140**

TABLE III: Number of correct matches up to a given distance to the ground-truth mesh. Our MLSD method outperforms the others: although it detects fewer segments (see Table I), more inliers are present.

and use the distance to the point cloud to define inlier and outlier matches (see Table III).

In Fig. 5, we also display some reconstructions, keeping only inliers as the presence of several outliers impedes the understanding of the scene. No bundle adjustment was performed, which explains why some 3D lines are not perfect. Our method yields a denser reconstruction.

V. CONCLUSION

We presented MLSD, a multiscale extension to the popular Line Segment Detector (LSD). MLSD is less prone to over-segmentation and is more robust to noise and low contrast. Being based on the *a contrario* theory, it retains the parameter-less advantage of LSD, at a moderate additional computation cost. Not only are the visual results better but, more importantly, the improvement can be measured quantitatively by the accuracy of line-based structure from motion. We argue that such an objective quantitative measure is very important for assessing the quality of line segment detectors. This is in stark contrast with usual experiments presented in the literature, where little or no quantitative measures are provided, and when they are, they are not as objective as residual SfM errors.

REFERENCES

[1] C. Akinlar and C. Topal. EDLines: A real-time line segment detector with a false detection control. *Pattern Recogn. Lett.*, 32(13):1633–1642, Oct. 2011.
[2] A. Elqursh and A. Elgammal. Line-based relative pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3049–3056, June 2011.
[3] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.
[4] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. On straight line segment detection. *Journal of Mathematical Imaging and Vision*, 32(3):313–347, Nov. 2008.
[5] R. Grompone 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*, 32(4):722–732, April 2010.
[6] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: a line segment detector. *Image Process. On Line (IPOL 2012)*, 2:35–55, 2012. http://dx.doi.org/10.5201/ipol.2012.gjmr-lsd.
[7] J. López, R. Santos, X. R. Fernández-Vidal, and X. M. Pardo. Two-view line matching algorithm based on context and appearance in low-textured images. *Pattern Recognition*, 48(7):2164–2184, 2015.
[8] P. Moulon, P. Monasse, and R. Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3248–3255, 2013.
[9] Y. Salaün, R. Marlet, and P. Monasse. Robust and accurate line-and/or point-based pose estimation without Manhattan assumptions. In *European Conference on Computer Vision (ECCV)*, 2016.
[10] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pages 1–8, June 2008.
[11] Z. Wang, F. Wu, and Z. Hu. MSLD: A robust descriptor for line matching. *Pattern Recognition*, 42(5):941–953, 2009.
[12] L. Zhang and R. Koch. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *J. Vis. Comun. Image Represent.*, 24(7):794–805, Oct. 2013.
[13] L. Zhang and R. Koch. Structure and motion from line correspondences: Representation, projection, initialization and sparse bundle adjustment. *Journal of Visual Communication and Image Representation*, 25(5):904–915, 2014.