# DUDE (DUALITY DESCRIPTOR): A ROBUST DESCRIPTOR FOR DISPARATE IMAGES USING LINE SEGMENT DUALITY

*Youngwook P. Kwon*[†] *Hyojin Kim*[⋆] *Goran Konjevod*[⋆] *Sara McMains*[†]

[†] UC Berkeley, Berkeley, CA 94720-1740
[⋆] Lawrence Livermore National Laboratory, Livermore, CA 94550

## ABSTRACT

We present a novel descriptor algorithm (DUDE) using line/point duality and a randomization strategy that provides simple but robust, consistent feature extraction and correspondence. Using duality enables us to effectively capture a distribution of line segments, and the proposed randomization strategy improves repeatability over existing techniques by generating more line features in common between two images. We demonstrate the effectiveness of our approach using a challenging set of disparate image pairs, and show that the DUDE descriptor performs comparably to state-of-the-art methods with significantly less computation expense.

***Index Terms*—** Image matching, multi-modal disparate images, feature detection, feature descriptor.

## 1. INTRODUCTION

Automatic feature matching is now relatively robust for images taken with the same sensors under the same conditions, such as in image panorama stitching. Mikolajczyk and Schmid's evaluation of local descriptors [1] found SIFT [2] had the overall best performance.

However, a more demanding situation arises from dramatic appearance changes such as occurs with images from different modalities (e.g., photo vs. painting), sensors, age, lighting, etc. Some examples are shown in Fig. 1, and representative performance shown in Fig. 2. The nature of such disparate images makes the feature detection and description process difficult since appearance changes may even include some features being completely absent in one of the images, or missing or inconsistent image textures and gradients (including gradient reversals). For example, pixel gradient-based local feature descriptors such as SIFT work poorly with disparate images where pixel gradients are frequently inconsistent, such as for multimodal medical imaging (e.g., CT vs. MRI).

This work was performed at UC Berkeley under partial funding from awards B615778 and B611926 and under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-PROC-692138.

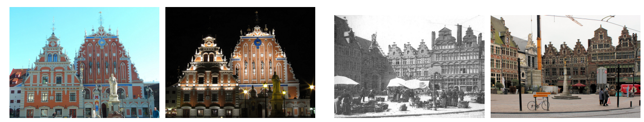

**Fig. 1**. Examples from [3, 4] of disparate image pairs.


**Fig. 2**. Top 40 highest confidence matches of SIFT (left pair) and our algorithm, DUDE (right pair). Green and red lines represent correct and incorrect matches, respectively.

Recently, more successful disparate image matching is reported by [3, 4]. The method by Hauagge and Snavely [3] detects symmetric structures in images and encodes them into descriptors. They also present a challenging dataset, mostly architectural scenes that include symmetric shapes, exhibiting dramatic variations in lighting, time period, modality, etc. Using the same dataset, Bansal and Daniilidis evaluate a new method [4] that analyzes the eigen-spectrum of the joint image graph constructed from all pixels in the images, and achieve impressive experimental results. However, [3] may not be suitable for images that do not include symmetric objects, and [4] is expensive both in time and memory for eigen-decomposition of a huge matrix.

We consider two questions: (1) what information tends to be consistently preserved across disparate images despite significant appearance change (e.g., reversal of brightness and darkness, partial absence of edges, etc.), and (2) how can we capture the information and create descriptors effectively and efficiently?

Inspired by these questions, we detect (hundreds or thousands of) line segments from a given image and exploit them as input to our descriptor. Lines compromise nicely between low-level and high-level information. Pixel-level gradients are not preserved well in our target images. High-level contours (such as humans rely on) are often well preserved, but extracting them is a challenging and parameter-sensitive task. Mid-level line segments can capture meaningful information about contours and are easy to handle mathematically.

Nevertheless, line detection is still parameter-sensitive, and does not guarantee unique endpoints. Occlusion and appearance change can cause line segments to become disconnected. Existing line-based techniques are not suitable for such cases: [5] requires a known epipolar geometry, [6] relies on intensities using a SIFT-like strategy, and [7, 8] rely on pairwise geometric relationships between line segments whose endpoints typically vary considerably under multi-modality.

To overcome the issue of inconsistent line segment detection, we propose a novel descriptor system named *DUDE* (DUality DEscriptor) that uses a 3D cylindrical histogram based on a transformation of line segments to a dual space (or parameter space) of points. By exploiting line-point duality, DUDE is designed so as to be less affected by line segment disconnection, and at the same time, captures geometric relationships very efficiently. This can be viewed as a line segment version of Shape Contexts [9].

To acquire repeatable and consistent line segments (and therefore feature points) across disparate images, we adopt an idea from an image segmentation technique in [10]. Instead of grouping to form segmentations (regions) based on predefined similarity in a bottom-up hierarchical manner, we similarly merge line segments from an initial line segment set (that may include broken line segments).

We evaluate our method with the challenging multi-modal dataset of [3, 4]. Although many of its images contain the symmetry required for Hauagge and Snavely [3]'s approach, DUDE outperforms [3] using a more general approach, and in combination with our feature detector, we achieve similar performance to the state-of-art [4] with significantly more efficient computation.

## 2. THE DUDE DESCRIPTOR

The idea underlying our descriptor is illustrated in Fig. 3, for two sets of line segments for the (manually chosen) corresponding regions centered at the red points. Despite similar appearance, the number and endpoints of corresponding line segments vary considerably. We propose to capture the *distribution*, relative to a feature location $p$, of line segment set $S$ despite disparate detection.

The DUDE descriptor takes advantage of line-point duality by transforming lines into points in dual space. We denote a line segment as $[r, \theta, f_1, f_2]$, instead of $[x_1, y_1, x_2, y_2]$, where $r$ and $\theta$ are defined by the infinite line[1] containing the segment, and the two $f$ values, calculated from $\sin(\psi)$, represent how far each endpoint is from the orthogonal projection

[1] $x \cos\theta + y \sin\theta - r = 0$.

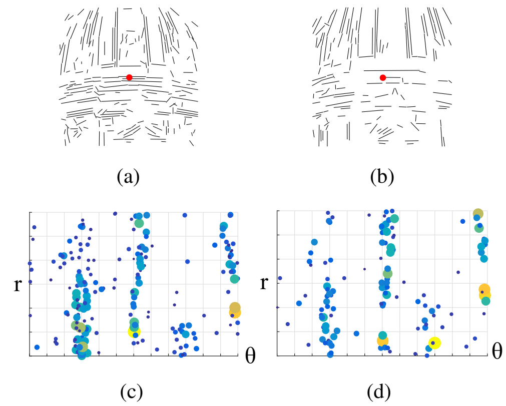


**Fig. 3**. (a, b) Line segment sets extracted from a pair of images of the same scene and (c, d) their $r$-$\theta$ dual space representations for the origins indicated by the red dots in (a, b)

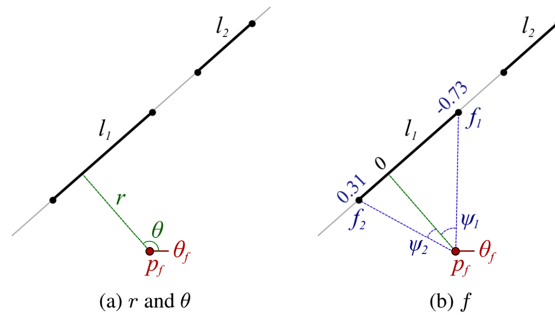


**Fig. 4**. Dual representation $r, \theta, f_1, f_2$ of a line segment

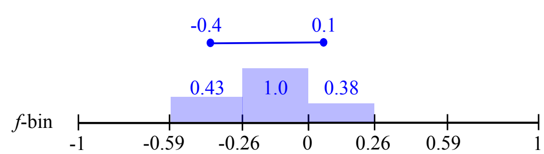


**Fig. 5**. An example of $f$-binning.

of the coordinate system origin onto the line (Fig. 4) [2].

Typically in image matching, when a set of features, $\mathbf{F}$, is given, a feature descriptor $\mathcal{D}_i$ is assigned to each $\mathcal{F}_i \in \mathbf{F}$. In many techniques such as SIFT, $\mathcal{F}_i$ is defined as $[x_i, y_i, s_i, \theta_i]$, representing the location, characteristic scale, and orientation, respectively. We also assume features have this form.

Our descriptor design is as follows. The set of line segment $\mathbf{S}$ is extracted from the image[3]. For a given feature $\mathcal{F}_i$, we first identify the set of line segments $\mathbf{S}_i \subset \mathbf{S}$ that are within a circle whose center is $(x_i, y_i)$ with radius $qs_i$, where $q$ is a parameter for local range of interest. We calculate $[r, \theta, f_1, f_2]$ values of all line segments $s \in \mathbf{S}_i$, relative to the $\mathcal{F}_i$-defined coordinate system with origin at $(x_i, y_i)$ and orientation $\theta_i$. Then for each $\mathbf{S}_i$ we create a 3D cylindrical coordinate $(r, \theta, f)$, $r \in [0, qs_i], \theta \in [0, 2\pi), f \in [-1, 1]$), and accumulate them in a histogram for $\mathcal{F}_i$. We divide the

[2] Naturally, $f$ has a value in $[-1, +1]$; we set $f_1 < f_2$.

[3] We use the Line Segment Detector (LSD) [11] for initial extraction.

histogram's $r$ and $\theta$ axes uniformly into $n_r$ and $n_\theta$ bins, respectively, and divide the range $f$ into $n_f$ bins using a log scale. Because $f_1$ and $f_2$ denote the endpoints of the range of each line segment, segments are binned as a *range*, contributing to bins by the coverage percentage. Fig. 5 shows an example of the range histogram when $f_1 = -0.4, f_2 = 0.1$ and $n_f = 6$. In this manner, each $[r, \theta, f_1, f_2]$ of $s \in \mathbf{S}_i$ is accumulated in the 3D histogram. By concatenating the histogram bins, we have a $(n_r \times n_\theta \times n_f)$-dimension descriptor $\mathcal{D}_i$. [4]

We address two underlying difficulties that any line-based approach must overcome. First, there is the case that one long line segment in an image is detected as multiple short segments in the counterpart image. We lessen this "disconnected detection" problem by the nature of our descriptor design. Because collinear line segments share the same $r$ and $\theta$, and their $f$ ranges are accumulated, the disconnection does not cause much difference in descriptors. This is seen in the two line segment sets plotted in $r$-$\theta$ space in Fig. 3 (c, d). Instead of the $f$ dimension, we visualize longer line segments with larger and lighter colored discs; collinear line segments will occupy the same location. One can see that despite the disconnection they have similar patterns in $r$-$\theta$ space. Secondly, slight changes of endpoints can cause changes in $r$ and $\theta$ values. We solve this problem by intentional perturbation of endpoints. We duplicate each segment $d$ times, while randomly perturbing the endpoints of the additional segments within $\pm 3$ pixels, both in $x$ and $y$. This can be regarded as blurring the histograms and making them less sensitive to unstable endpoint detection.

## 3. FEATURE DETECTION

Since many feature detectors return features in the form of $[x_i, y_i, s_i, \theta_i]$, different detectors and descriptors can be combined. DUDE descriptors also can take existing feature detectors (e.g., SIFT) as input. However, in this section, we discuss how to generate a set of more repeatable features across disparate images, which are more suitable for our descriptors.

Because DUDE descriptors use line segments, the concept is basically to derive a feature per line segment: $(x_i, y_i)$ at its midpoint, $s_i$ as half of its length, and $\theta_i$ its orientation. However, regardless of line detector algorithms, most initial line segments are inconsistent across multi-modal images, due to dramatic appearance changes, as shown in Fig. 6 (middle column).

To increase the consistency of feature extraction, we propose to generate multiple *groupings* of line segments from the initial line segments by randomly merging them, inspired by an image segmentation technique in Kim et al. [10]. For image segmentation, they conducted randomized bottom-up merging from superpixels in a hierarchical structure, and

[4] $q = 10, n_r = 5, n_\theta = 5, n_f = 10$ for our experiments.

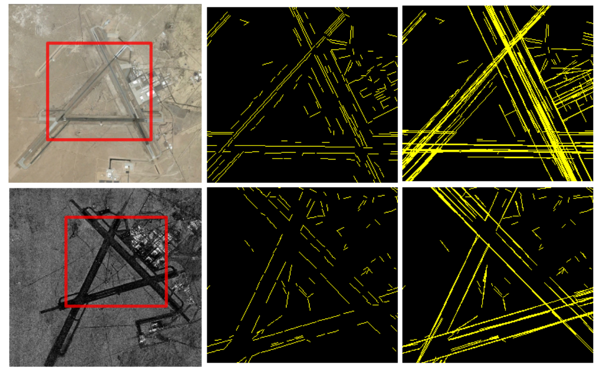

**Fig. 6**. An example of two disparate images and their line segments: An EO (Electro-Optics) and a SAR (Synthetic Aperture Radar) image from [12] (left column), initial line segments from LSD [11] (middle), and the proposed randomly merged line segments (right).

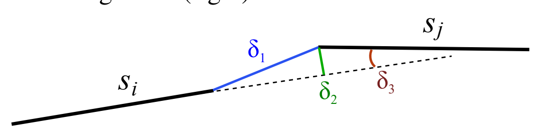


**Fig. 7**. Merging criterion for the randomized merging process; shortest distance $\delta_1$ (in pixels), perpendicular distance $\delta_2$ (in pixels), and angle $\delta_3$ (in degrees) between line segment $i$ and $j$.

identified meaningful regions. We use a similar randomized merging scheme (see [10] for details), but take line segments as initial input and use our own merging criterion.

The line merging process is as follows. Given an initial line segment set, we build a graph where each node represents a line segment, and each edge connects two neighboring line segments with a corresponding weight. The edge weight $w_{i,j}$ (merging criterion) between line segment $i$ and $j$ consists of three terms: shortest distance $\delta_{1,i,j}$, perpendicular distance $\delta_{2,i,j}$, and angle $\delta_{3,i,j}$ between the two line segments (Fig. 7):

$$w_{i,j} = (1 - \delta_{1,i,j}/\alpha)(1 - \delta_{2,i,j}/\beta)(1 - \delta_{3,i,j}/\gamma) \qquad (1)$$

where $\alpha$, $\beta$ and $\gamma$ are predefined thresholds (80, 16, and 15, respectively). We collect edges whose weight is larger than a predefined threshold $\delta_w$ (0.5), and sort them in the descending order. We then incrementally merge line segments, as in [10], updating the graph and computing new edge weights. The merged line segments from the hierarchies are then used for extracting line features.

## 4. EXPERIMENTAL RESULTS

To evaluate the efficacy of our detector and descriptor, we follow the evaluation methodology of [3, 4], using the same challenging dataset of 46 image pairs from [3].

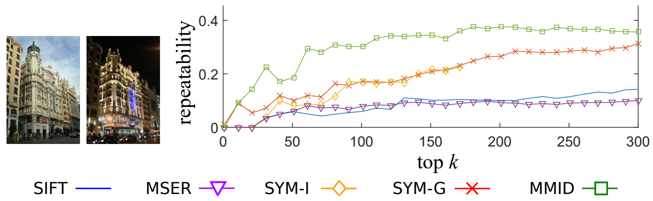


**Fig. 8**. Repeatability. Five detectors are compared: SIFT, MSER, SYM-I, SYM-G, and MMID (ours). For a given image pair (left), repeatability ($y$-axis) is computed when considering top-$k$ features ($x$-axis).

| | SIFT | SYM-I | SYM-G | JSPEC | MMID |
|---|---|---|---|---|---|
| k=100 | 0.144 | 0.135 | 0.173 | 0.287 | 0.217 |
| k=200 | 0.153 | 0.184 | 0.228 | 0.292 | **0.310** |

**Table 1**. Mean repeatability for the entire dataset

First, to test the efficacy of our detector, we calculate *repeatability*, a common measure indicating an ability for a feature detector to extract common feature points across image pairs. For feature sets $\mathbf{F}_1$ and $\mathbf{F}_2$, from an image pair, repeatability is the fraction of the number of repeatedly detected features over the total number of features. To determine if $\mathcal{F}_i \in \mathbf{F}_1$ and $\mathcal{F}_j \in \mathbf{F}_2$ are repeatedly detected, we use the overlap measure [13] (see also [3, 4] for details).

Note that simply producing a larger numbers of features can increase repeatability. Therefore, for evaluation, repeatability should be computed and compared only by considering the "top-$k$" detections. To select a subset of $k$ detections, we use scale descending ordering as in [4].

We compare the performance of five detectors: SIFT keypoints (DoG), SYM-I, SYM-G, JSPEC and our detection based on the merged line segments, MMID. Fig. 8 shows repeatability curves generated by varying $k$ values ($x$-axis), except for JSPEC, for which source code was not available. We also provide the performance table suggested by [3, 4] in Table 1, which indicates the average repeatability over the entire set of 46 image pairs when $k = 100$ and $k = 200$, respectively.

There are a few things to note. First, JSPEC requires significantly higher computation cost both in time and space. Second, the repeatability should be understood carefully. Repeatability is only one aspect of feature detection, and it can be easily biased (e.g., to the number of detections or density of detections) and Table 1 represents only two cross-sections ($k = 100$ or 200). A detection with higher repeatability does not always mean a better input for feature descriptors. Because the final goal is to find a better set of correct correspondences, the evaluation should also be considered in combination with descriptors, as below. Our detector does, however, show excellent repeatability.

For descriptor evaluation, we calculate the standard NNDR (Nearest Neighbor Distance Ratio) score [2] for each of the matches paired by descriptor similarity. By varying the threshold on the NNDR score, and identifying which matches are correct (with a known ground truth transformation), we obtain a precision-recall curve (see our project page [14]) from which mean average precision (mAP) is calculated.

| | Detectors | | | | | |
|---|---|---|---|---|---|---|
| **Descriptors** | GRID | SIFT | SYM-I | SYM-G | JSPEC | MMID |
| SIFT | 0.49 | 0.21 | 0.28 | 0.25 | **0.61** | 0.24 |
| SYMD | 0.41 | 0.22 | 0.20 | 0.25 | - | 0.26 |
| SIFT-SYMD | 0.58 | 0.28 | 0.35 | 0.36 | - | - |
| DUDE | **0.63** | **0.35** | **0.40** | - | - | **0.57** |

**Table 2**. Mean average precision (mAP) for different combinations of detector and descriptors from [2, 3, 4], and DUDE.

To evaluate the efficacy of detector and descriptor separately, we compare the results from different combinations. Table 2 shows the summarized results over the entire set of 46 image pairs. The column headings list the different feature detectors tested (GRID[5], SIFT [2], SYM-I, SYM-G [3], JSPEC [4], and MMID (ours)), and the row headings list the different descriptors (SIFT [2], SYMD, SIFT-SYMD [3], DUDE (ours)). Regardless of detector, DUDE outperforms other descriptors.

When we test the DUDE descriptors combined with our feature detector, we achieved comparable performance to the state-of-art result of JSPEC, but with much lower computation cost. To achieve its high performance, JSPEC requires eigen-value decomposition of a huge affinity matrix. Constructed by densely sampling every 5 pixels and assigning a descriptor to each sampled pixel, for the example image pair in Fig. 8, each $700 \times 500$ image will generate 14000 descriptors of dimension 256. Then one needs to solve an eigen-value decomposition of a $28000 \times 28000$ matrix. This seems infeasible using C++ OpenCV 2.4.11 (Windows 10, Intel i7-4790K, and 8GB RAM), and although MATLAB performs better, even the eigen-value decomposition of an $18000 \times 18000$ matrix takes 1.6 hours. In contrast, with DUDE, the entire feature matching procedure for this representative pair of images takes only 24 seconds (11 s. for detectors, 7 s. for descriptors, and 6 s. for matching and filtering).

## 5. CONCLUSION

In this paper we propose a novel detection and descriptor system for disparate image matching. The proposed DUDE descriptor can capture the relative distributions of unstably detected line segment sets in a consistent and efficient manner, and be general enough to be integrated with any choice of feature detection. DUDE outperforms existing descriptors, and in combination with our feature detector, we achieve similar performance to that of the state-of-art with significantly more efficient computation.

[5]GRID is a synthetic feature detector suggested in [3], that can be regarded as a perfect feature detector of repeatability 1, allowing pure evaluation of descriptor performance when detection is perfect.

## 6. REFERENCES


[1] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005. 1

[2] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, 2004. 1, 4

[3] D. C. Hauagge and N. Snavely, "Image matching using local symmetry features," in *Computer Vision and Pattern Recognition*, 2012, pp. 206–213. 1, 2, 3, 4

[4] M. Bansal and K. Daniilidis, "Joint spectral correspondence for disparate image matching," in *Computer Vision and Pattern Recognition*, 2013, pp. 2802–2809. 1, 2, 3, 4

[5] C. Schmid and A. Zisserman, "Automatic line matching across views," in *Computer Vision and Pattern Recognition*, 1997, pp. 666–671. 2

[6] Z. Wang, H. Liu, and F. Wu, "MSLD: A robust descriptor for line matching," *IEEE International Conference on Computer-Aided Design and Computer Graphics, CAD/Graphics*, vol. 42, pp. 128–133, 2009. 2

[7] L. Wang, U. Neumann, and S. You, "Wide-baseline image matching using Line Signatures," in *International Conference on Computer Vision*. 2009, pp. 1311–1318, IEEE. 2

[8] F. Tombari, A. Franchi, and L. Di, "BOLD features to detect texture-less objects," *International Conference on Computer Vision*, pp. 1265–1272, 2013. 2

[9] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002. 2

[10] H. Kim, J. Thiagarajan, and P.-T. Bremer, "Image segmentation using consensus from hierarchical segmentation ensembles," in *IEEE International Conference on Image Processing*, Oct 2014, pp. 3272–3276. 2, 3

[11] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: a Line Segment Detector," *Image Processing On Line*, vol. 2, pp. 35–55, 2012. 2, 3

[12] [Web page], "SAR images from Sandia National Laboratories," `https://web.archive.org/web/*/http://www.sandia.gov/radar/images/*`, Last accessed on Apr 21, 2015. 3

[13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, a. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005. 4

[14] [Web page], "DUDE (DUality DEscriptor): A robust descriptor for disparate images using line segment duality," `http://me.berkeley.edu/~mcmains/research/data/KwonMcMainsICIP16/`, Last accessed on May 16, 2016. 4