

# Feature Pairs Connected by Lines for Object Recognition

Muhammad Awais and Krystian Mikolajczyk

Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK

*m.rana, k.mikolajczyk@surrey.ac.uk*

## Abstract

*In this paper we exploit image edges and segmentation maps to build features for object category recognition. We build a parametric line based image approximation to identify the dominant edge structures. Line ends are used as features described by histograms of gradient orientations. We then form descriptors based on connected line ends to incorporate weak topological constraints which improve their discriminative power. Using point pairs connected by an edge assures higher repeatability than a random pair of points or edges. The results are compared with state-of-the-art, and show significant improvement on challenging recognition benchmark Pascal VOC 2007. Kernel based fusion is performed to emphasize the complementary nature of our descriptors with respect to the state-of-the-art features.*

## 1. Introduction

State-of-the-art object recognition approaches [8, 10] are based on local descriptors, codebook representation and kernel based classifiers. Despite very active research in the recognition community there is no different approach applicable to a general recognition problem that would lead to competitive results. The main improvements are therefore being done in the components of this approach. This paper focuses on the features extracted from the image. Instead of designing a new descriptor it proposes an approach to capture more complex shape structures by combining existing state-of-the-art descriptors [6]. We focus on dominant edge structures to represent the shape of many object categories. Edges have been successfully used in the context of object recognition, in particular [7] has exploited edges to detect objects. Connected edge pairs were used in [3] to improve the discriminative properties of shape descriptors. However, the repeatability of such built structures is low due to ambiguities in selecting edges for pairs. Large number of edge combinations was used to overcome this problem which resulted in

over-complete and redundant representations. Pairs of interest points were also used in [9] to combine appearance and topology of local image structures in the context of wide baseline matching. This representation is not robust to occlusion or background clutter as random pairs of points were formed. The main problem is to identify the features that can form a pair likely to repeat across examples of the same object.

We propose to use segmentation maps that provide labeled regions including boundaries and vertices. To reduce the number of vertices that are often due to noise we fit lines into region boundaries and keep only dominant structures. We then use the dominant lines to find the features in the form of line centers and end points. The edge patterns at these points are then described by a histogram of gradient orientations similar to SIFT [5]. The main contribution is the novel representation based on pairs of such features that capture more structural information than interest points, yet repeatable across different instances. Pairs of points that are connected within the same segment are likely to belong to the same object thus more robust to occlusion and viewpoint change. They are also more discriminative than the individual points. The proposed representation is extensively evaluated on challenging Pascal 2007 dataset [2]. The results are presented and compared with the best performing features within a state-of-the-art recognition system. We obtain significant improvement over existing point based representations and demonstrate that our features are complementary to them.

The remainder of this paper is organized as follows. Section 2 presents the approach to obtain approximate line based images using segmentation maps as well as features based on connected line ends. Section 3 discusses the classification approach and section 4 presents the experimental results.

## 2. Feature Extraction Approach

In this section we discuss our feature extraction approach based on dominant line segments. We first

present the line segment extraction method and then the features based on this representation. The features include line segment descriptors and connected line ends that exploit the topological relations between region vertices.

## 2.1. Dominant Line Segment Extraction

Dominant line segment extraction starts with an edge detection algorithm. The edge map can be computed by a standard edge detector or from the boundaries of an image segmentation. In contrast to edge detectors such as Canny, the segmentation methods also identify which edges and vertices belong to the same regions. Moreover, the segmentation methods filters out small noise edges that typically remain after Canny detector. The segmentation map used in our system is computed using standard Watershed [4] approach for its efficiency. The map is illustrated in Figure 1. To identify the dominant edges we fit lines into segment boundaries. We have used RANSAC to estimate 5 parameters for each line segment: mid-point  $(x_{cen}, y_{cen})$ , length  $(l)$ , perpendicular distance  $(\rho)$  of the line from the origin of the image, and orientation  $(\theta)$  of this perpendicular line from the origin. To capture dominant structure and avoid noise due to small segments we only keep the lines above a significant length. Parametric line representations makes it possible to control the degree of approximation and gives smoothing effect over the segmentation boundaries. Moreover, insignificant lines can be filtered out in the parameter space rather than image space. Given this parametric representation the line based image approximation can be reconstructed for a given degree of accuracy and regions are simplified to polygons. Figure 1 illustrates the process: (a) is the original image, (b) is the corresponding watershed segmented image with segment boundaries, (c) is the reconstructed line based image approximation. As one can notice the reconstructed image is sufficiently similar to the original one to allow for recognition of objects based on their shapes.

**Line Based Features:** We first define a baseline feature that uses lines fitted to the segment boundaries. These lines are important as they carry the information on object shapes and represent the transition area between two regions. To capture the appearance of the line and its neighbourhood we define a circular region of interest centred on the middle of the line with its diameter determined by the length of the line. The circular region corresponding to line  $l_1$  in illustrated in Figure 1 (d). The edge patterns in the area of dominant lines are described by histogram of gradient orientation similar to SIFT [5].

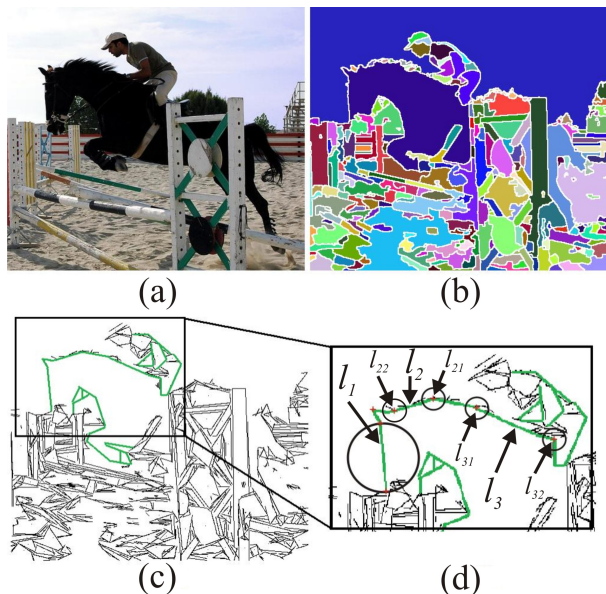


Figure 1. Line based representation.

**Connected Line Ends:** Vertices of the simplified segmentation regions can be used as features that encode local discontinuities of region boundaries. In a similar way to features centred on lines we define a circular regions centred at line end points and encode their neighbourhood by a histogram of gradient orientation as shown by line ends  $l_{21}, \dots, l_{32}$  in Figure 1. However, individual vertices are not sufficiently discriminative in large datasets. To capture more complex edge structures we combine features into pairs. Pairs have been used in past to improve the discriminative representations but the main difficulty is to select features that are likely to co-occur together. Vertices connected by a line are more likely to belong to the same object thus more often co-occur in different examples of the same object category. For example, man made objects are often highly symmetric and can be easily represented by straight lines. Two vertices connected by a line can represent a discriminative pattern yet are repeatable enough to generalise to unobserved object instances. The angle between x-axis and the connecting line defines the order of the line ends that form a pair which is repeatable. For acute angles the line end point closer to the y-axis is taken as the first point as shown by point  $l_{31}$  in Figure 1(d). For obtuse angles the line end point farther from the y-axis is taken as the first point as shown by point  $l_{21}$  in Figure 1(d). Although rotation invariance was not considered in this work it can be obtained by overlying two circular regions and building one histogram of their gradients. In our implementation, pairs of descriptors that are connected by a dominant line are concatenated and

form a new descriptor.

To summarize, three types of descriptors are built based on the parametric line approximation: line based (LB), line ends (LE), and connected line ends (ConLE) features. We evaluate these features in the context of object category recognition within a state-of-the-art system.

### 3. Object Recognition System

In this section we discuss our object recognition approach based on Spectral Regression Kernel Discriminant Analysis (SR-KDA) introduced in [8]. SR-KDA complexity scales linearly with respect to the number of classes while its performance is consistent with state-of-the-art methods [8]. For each object category  $\alpha \in Y$  a separate binary classifier  $h_\alpha : X \rightarrow \{-\alpha, \alpha\}$  is learned, where  $Y = \{1, 2, \dots, N\}$  is finite number of object classes. Thus, the object recognition problem is modelled as two class problem by dividing the original data set into  $N$  sets.

We compute the descriptor as described in section 2 and used bag-of-words model [8] to represent the image. There are three types of feature points combined with SIFT descriptor (i) line segments, (ii) line ends and (iii) pairs of line ends. For visual codebooks, each descriptor is clustered using k-means to form a codebook of 4000 clusters. The image is divided into 4 spatial grids [8]: entire image, horizontal bars (1x3), vertical bars (3x1) and image quarters (2x2). Each spatial grid is then represented by histograms of codebook occurrences. For each of 4 spatial location grids, a separate kernel matrix is computed which are then combined by simple averaging. The kernel function to compute the entry  $(i, j)$  of the kernel matrix is based on the  $\chi^2$  distance between features  $F_i$  and  $F_j$ :

$$k(F_i, F_j) = e^{-\frac{1}{A} dist_{\chi^2}(F_i, F_j)}$$

where,  $A$  is a scalar for normalizing the distance, and is set to average  $\chi^2$  distance between all features [8].

### 4. Experimental Setup

This section presents the experimental results on challenging data set Pascal VOC 2007 [2] which consists of 20 object classes with 9963 image examples (2501 training, 2510 validation, and 4952 testing images). The data includes both indoor and outdoor scenes, truncated and occluded objects at various scales and different lighting conditions. The edge maps form segmentation of these images include complex structure, e.g., curves, arcs, which are reliable capture by the proposed approach. We present the results using average precision (AP) [2], which is proportional to the

area under precision recall curve. Mean average precision (MAP) is computed by averaging scores for all 20 classes.

### 4.1. Results and Discussion

The classification results of approaches discussed earlier are presented in Table 1 and their comparison is made with other methods from [10]. Dense and DenseOPP are densely sampled SIFT and Opponent colour SIFT, respectively. These two are our reference methods which gave top scores in [10] as well as in other methods evaluated in Pascal Challenge [2]. We experimented with other interest point detectors from [6] but their results were lower than Dense SIFT due to much lower number of detected features. Note that our approach is not using colour information even though colour brings 1.8% overall improvement according to the scores for these two methods in Table 1. Moreover, our features carry at least 2.5 times less data than Dense SIFT. Significant improvements by all our edge based representations upon Dense and DenseOPP can be noticed. For example, LB (line based) gives overall 4.5% and 2.7% improvement over Dense and DenseOPP, respectively. LE (line ends encoded by SIFT) leads to larger improvements of 6.6% over Dense SIFT. This advantage is further increased by 1% with ConLE (connected line ends) based on pairs of descriptors. In most cases average precision (AP) for individual categories in LB, LE and ConLE shows improvements upon Dense and even upon DenseOPP. For shape based classes such as car, sofa, horse and train the improvement is between 5 to 10%. For other rigid classes such as bicycle, bus, diningtable, motorbike improvement is even higher and reaches nearly 20% for diningtable. These results suggest that our representation carry more salient information about object shape than densely sampled features.

To investigate the complementarity of our features to the state-of-the-art representation we carry out their kernel level fusion. Kernels produced by the proposed features are averaged with Dense one and a classifier is trained. There is an improvement of 10% or more in Dense method by combining it with any of our features. It shows that our approach is indeed complementary to dense sampling and leads to significant increase of performance.

To produce state-of-the-art results on this dataset we combined our features with 17 other descriptors by averaging it with 17 kernels from [1, 10]. Fourteen of these kernels are based on dense sampling as well as on Harris-Laplace detector encoded with SIFT, HVS-SIFT, Opponent-SIFT, RGB-SIFT, and other colour variants of SIFT [10], and three of them are linear discrimi-

**Table 1. Mean Average Precision of PASCAL VOC 2007.**

Concepts	Dense	DenseOPP	LB	LE	ConLE	LB+	LE+	ConLE+	17 Kernels [10, 1]	ConLE+ 17 kernels
	[10]	[10]				Dense	Dense	Dense		
Aeroplane	66.0	70.9	64.9	70.2	68.3	76.6	78.1	77.4	80.5	81.1
Bicycle	48.7	50.4	53.1	<b>58.4</b>	<b>60.7</b>	61.7	63.4	64.8	67.7	69.0
Bird	39.8	43.7	35.5	41.3	42.8	46.0	47.1	47.8	60.0	60.7
Boat	54.7	60.4	58.7	59.4	60.9	67.3	66.2	66.9	72.1	72.2
Bottle	16.7	17.6	21.3	19.1	20.0	21.8	20.3	20.1	27.2	28.0
Bus	44.3	43.3	<b>52.3</b>	<b>59.1</b>	<b>57.7</b>	63.9	65.4	64.4	67.7	68.8
Car	70.4	69.4	74.2	75.7	77.6	78.4	78.6	79.4	80.4	81.1
Cat	43.4	38.4	<b>53.7</b>	<b>53.6</b>	<b>53.5</b>	57.4	57.0	56.7	58.0	59.4
Chair	41.6	40.2	45.8	46.4	46.3	49.3	48.9	48.4	51.9	52.6
Cow	28.8	29.2	32.1	31.8	33.0	37.5	36.8	36.4	46.1	45.9
Diningtable	23.5	31.2	45.9	52.7	58.7	54.0	57.1	59.4	57.6	63.0
Dog	37.6	36.6	37.6	39.1	38.8	40.3	40.6	39.1	46.8	47.1
Horse	68.6	73.9	73.8	76.3	77.0	78.2	78.3	78.8	82.3	82.7
Motorbike	48.2	50.7	56.1	<b>60.5</b>	<b>62.9</b>	64.0	64.6	65.9	67.5	68.7
Person	79.9	81.8	82.4	82.9	83.6	85.0	84.8	84.9	87.6	88.0
Pottedplant	12.2	15.7	20.3	23.3	21.9	24.5	25.7	23.9	38.2	39.2
Sheep	26.5	38.2	30.2	26.3	26.0	29.2	26.6	25.7	48.5	46.8
Sofa	35.7	31.1	35.3	37.5	40.5	43.1	44.6	45.2	48.6	49.4
Train	66.6	69.0	72.5	73.8	75.7	78.4	78.3	78.8	84.8	85.1
Tvmonitor	43.3	39.6	40.2	41.0	42.1	48.5	48.7	49.1	53.3	54.6
MAP	44.8	46.6	<b>49.3</b>	<b>51.4</b>	<b>52.4</b>	55.3	55.6	55.7	61.3	62.2

nant projections [1] of descriptors based on dense sampling and SIFT descriptor. The results for 17 kernels are given in Table 1. Combined kernels score is significantly better than any individual one, and the improvement is nearly 10% compared to our best features. We added our best performer ConLE to the sum of 17 kernels (ConLE +17 kernels in Table 1). This further improves state-of-the-art score by 1%. It is a significant gain given small differences between results for this dataset produced by various recognition systems reported in the literature. Note that a poor descriptor can significantly decrease the performance even when combined with 17 other kernels.

## 5. Conclusions

We have proposed a novel approach to extract local image features based on line segments fitted into dominant edges in images. We have also proposed a method to combine local descriptors into repeatable pairs that capture more complex image shapes. We have extensively evaluated the features within a state-of-the-art recognition system on a challenging Pascal07 benchmark data. The proposed methods lead to significant improvements over state-of-the-art features. The results indicate that using the end points and junctions of significant edge structures enables filtering out less salient points which are frequently detected by interest point detectors or obtained by dense sampling. Pairs of line ends allow encoding more complex structures and result in higher performance. We have also demonstrated that our method is complementary to other features which

together produce state-of-the-art results on this dataset. **Acknowledgements.** This research was supported by UK EPSRC EP/F0034 20/1 and the BBC R&D grants.

## References

- [1] H. Cai, K. Mikolajczyk, and J. Matas. Learning Linear Discriminant Projections for Dimensionality Reduction of Image Descriptors. In *BMVC*, 2008.
- [2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results.
- [3] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *IEEE T. PAMI*, 30(1):36–51, 2008.
- [4] P. Koniusz and K. Mikolajczyk. Segmentation based interest points and evaluation of unsupervised image segmentation methods. In *BMVC*, 2009.
- [5] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [6] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE T. PAMI*, 27(10):1615–1630, 2005.
- [7] J. Shotton, A. Blake, and R. Cipolla. Contour-based learning for object detection. In *ICCV*, volume 1, 2005.
- [8] M. A. Tahir, J. Kittler, K. Mikolajczyk, F. Yan, K. van de Sande, and T. Gevers. Visual category recognition using spectral regression and kernel discriminant analysis. In *International Workshop on Subspace, ICCV*, 2009.
- [9] D. Tell and S. Carlsson. Combining Appearance and Topology for Wide Baseline Matching. In *ECCV*, 2002.
- [10] K. van de Sande, T. Gevers, and C. Snoek. Evaluation of color descriptors for object and scene recognition. In *CVPR*, 2008.