RIDE: Reversal Invariant Descriptor Enhancement

Lingxi Xie[1], Jingdong Wang[2], Weiyao Lin[3], Bo Zhang[1] and Qi Tian[4]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China

[2]Microsoft Research, Beijing, China [3]Shanghai Jiao Tong University, Shanghai, China

[4]Department of Computer Science, University of Texas at San Antonio, Texas, USA

ABSTRACT

In many fine-grained object recognition datasets, image *orientation* (left/right) might vary from sample to sample. Since handcrafted descriptors such as SIFT are not reversal invariant, the stability of image representation based on them is consequently limited. A popular solution is to augment the datasets by adding a left-right reversed copy for each original image. This strategy improves recognition accuracy to some extent, but also brings the price of almost doubled time and memory consumptions.

In this paper, we present **RIDE** (**Reversal Invariant Descriptor Enhancement**) for fine-grained object recognition. RIDE is a generalized algorithm which cancels out the impact of image reversal by estimating the *orientation* of local descriptors, and guarantees to produce the identical representation for an image and its left-right reversed copy. Experimental results reveal the consistent accuracy gain of RIDE with various types of descriptors.

CONTRIBUTION

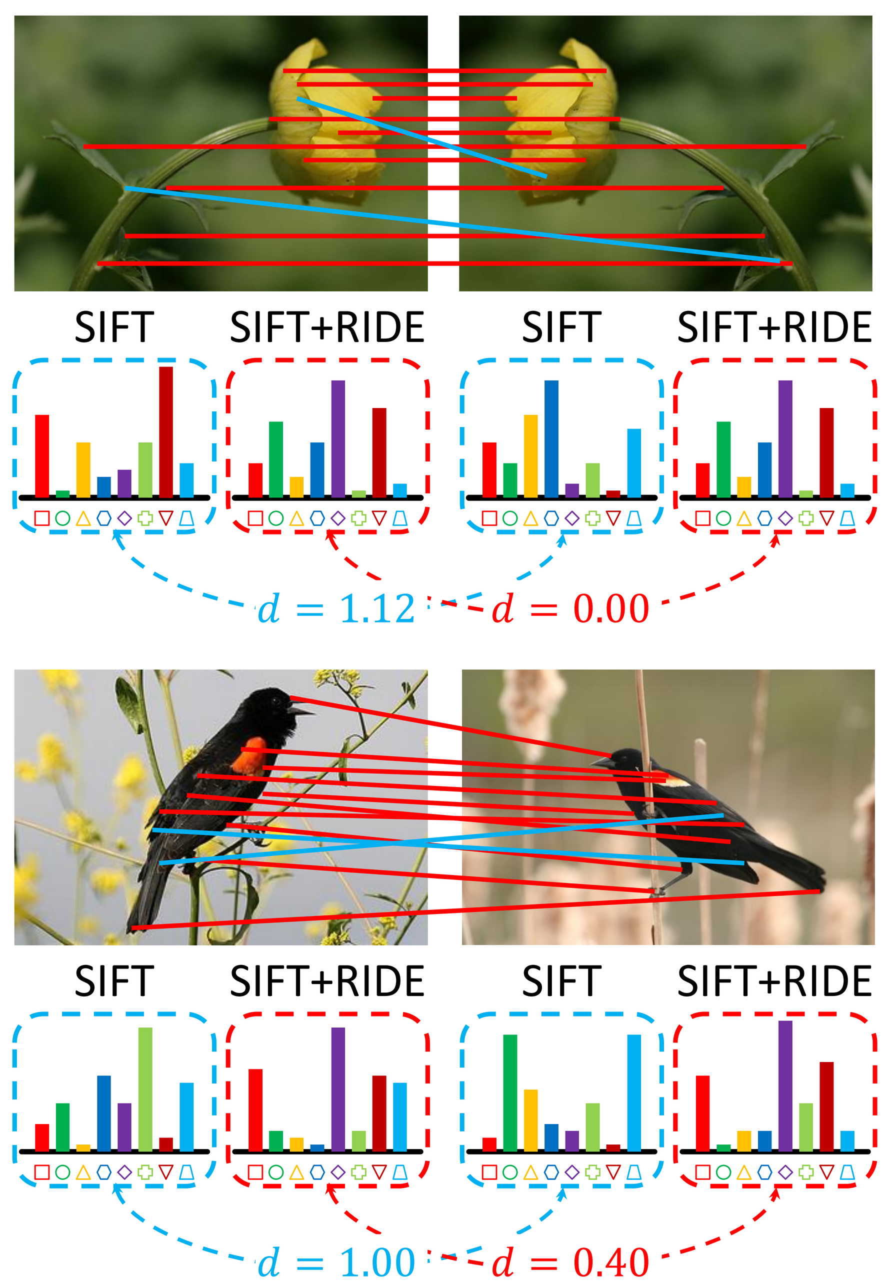
In this paper, we propose a novel algorithm named RIDE (Reversal Invariant Descriptor Enhancement) to cancel out the transformation on local image descriptors when the input image is reversed. Based on which, we obtain a reversal invariant image representation, which produces significant improvement in image classification tasks.

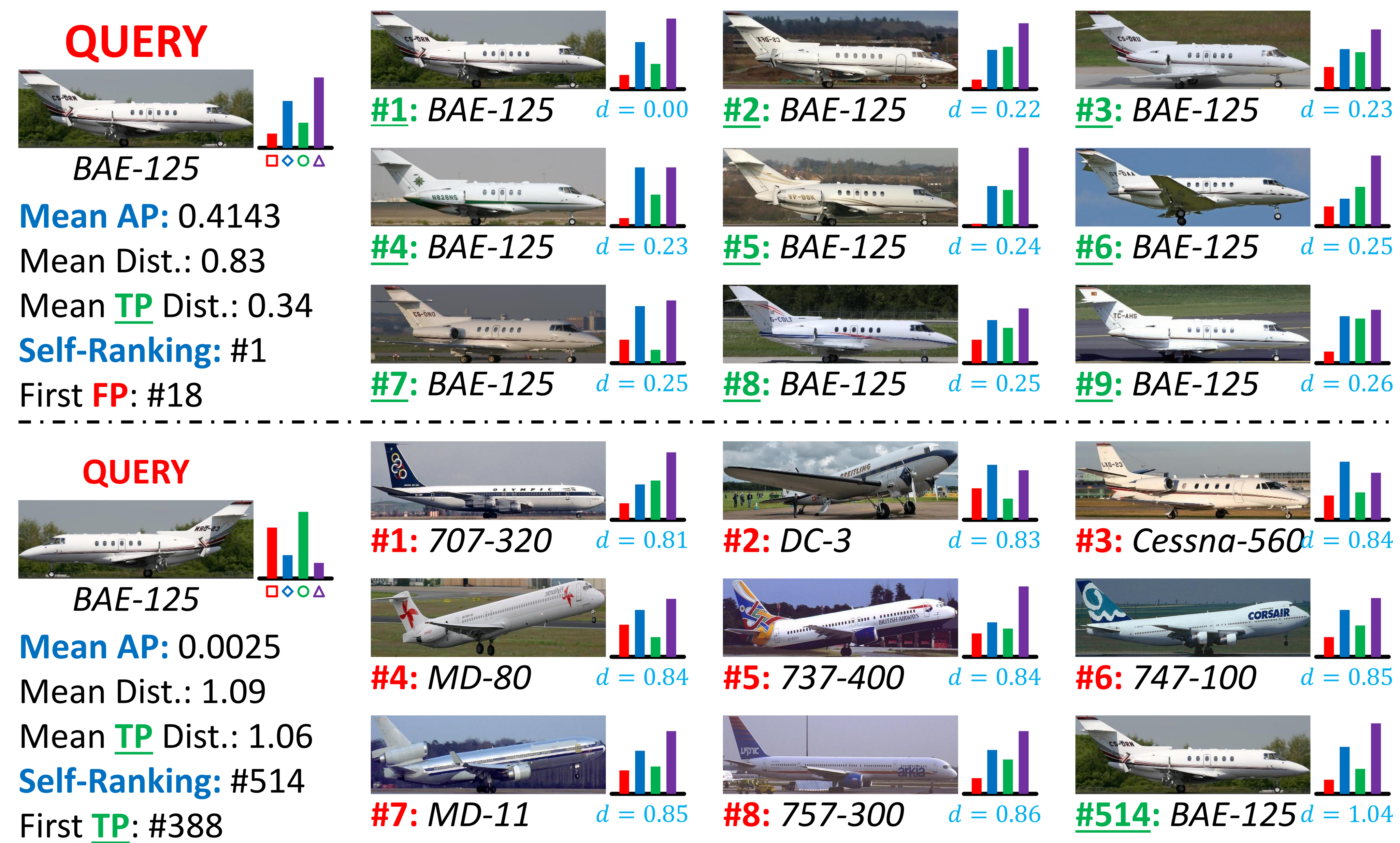
1. With several verification experiments, we demonstrate that reversal invariance is very important for robust image representation, but it is not considered in most popular local descriptors such as SIFT.
2. We propose RIDE, a simple algorithm to add reversal invariance to local descriptors. The algorithm is based on the observation of the structure of SIFT descriptors. We compute the *orientation* of each SIFT descriptor based on the oriented gradients, and transfer the orientation to other local descriptors.
3. To the best of our knowledge, RIDE is the first efficient algorithm to extract reversal invariant image representation for image classification.

Experimental results on several popular image classification benchmarks verify that RIDE effectively improves the image feature quality. Compared to data augmentation, we produce slightly better recognition results with only half computational costs. The success of our algorithm also lays the foundation of combining reversal invariance to some new algorithms such as training a reversal-invariant Convolutional Neural Network (CNN).

THE PROPOSED FRAMEWORK

Motivation from Image Matching (left) & Retrieval (right)





SIFT Reversal

Mapping each ***original*** index to a ***reversed*** index

Original SIFT

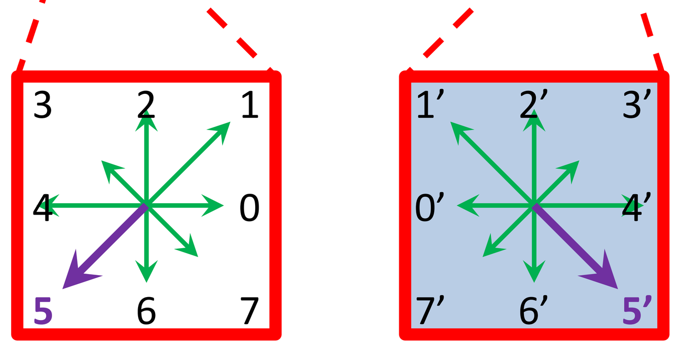
0	1	2	3
4	5	6	7
8	9	10	11
12	13	$\boxed{14}$	15

Reversed SIFT

3′	2′	1′	0′
7′	6′	5′	4′
11′	10′	9′	8′
15′	$\boxed{14'}$	13′	12′

Original Index

$14 \times 8 + 5 = 117$



Reversed Index

$13 \times 8 + 7 = 111$

Gradient Histogram

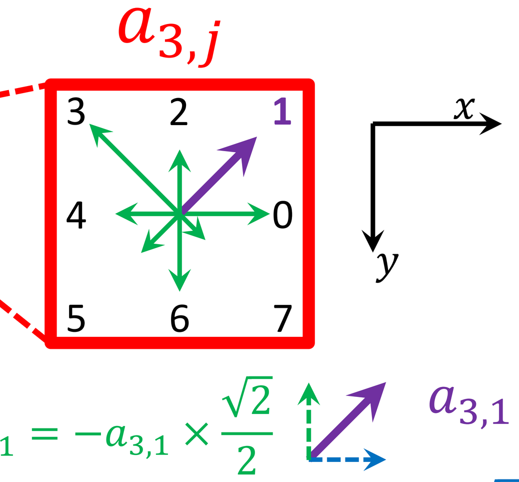
Computing Orientation

Oriented ***gradients*** of SIFT for computing orientation

$a_{i,j}$

0	1	2	$\boxed{3}$
4	5	6	7
8	9	10	11
12	13	14	15

$a_{3,j}$



$G_{3,1}^{Y} = -a_{3,1} \times \frac{\sqrt{2}}{2}$

$G_{3,1}^{X} = a_{3,1} \times \frac{\sqrt{2}}{2}$

$$G^X = \sum_{i=0}^{15} \sum_{j=0}^{7} G_{i,j}^{X} \qquad G^Y = \sum_{i=0}^{15} \sum_{j=0}^{7} G_{i,j}^{Y}$$

The ***combination*** of G^X and G^Y handles up to **8** cases (the sign of G^X, the sign of G^Y, G^Xcompared to G^Y)

Useful Notes

1. A simplified version of RIDE is **Max-SIFT**, which is extremely easy to implement and has ***no cost*** to improve classification.
2. RIDE can handle various types of reversal transformations, including left-right reversal, upside-down reversal and 90, 180 or 270-degree rotations.
3. RIDE can be applied to any types of local descriptors, by transforming the orientation computed on SIFT to them.
4. RIDE can be generalized to fast binary features, such as BRIEF or ORB, by computing orientation stats on them.

RESULTS

CUB-200-2011 Accuracy (%)

Results on ***small*** codebooks (GMM, $K = 32$)

	ORIG	RIDE	AUGM	RIDEx2
SIFT	25.77	32.14	31.60	**34.07**
LCS	36.18	38.50	38.97	**40.16**
SIFT+LCS	38.11	44.73	43.98	**46.38**
RGB-SIFT	31.36	39.16	38.79	**41.73**
OPP-SIFT	35.40	42.18	41.72	**44.30**

Results on ***large*** codebooks (GMM, $K = 256$)

	ORIG	AUGM	RIDEx2
SIFT+LCS, large vocab.	47.61	50.16	**50.81**
SIFT+LCS, part det. [6]	56.6	59.1	**60.7**
SIFT+LCS, part det. [12]	65.3	67.0	**67.4**

Please refer to our paper for more results

Conclusions

In this paper, we propose RIDE (Reversal Invariant Descriptor Enhancement) which brings reversal invariance to local descriptors. RIDE cancels out the impact of image/object reversal by estimating the orientation of each descriptor, and then forcing all the descriptors to have the same orientation. Experiments reveal that RIDE significantly improves the accuracy of fine-grained object recognition and scene classification with very few computational costs. Compared with dataset augmentation, RIDE produces better results with lower time/memory consumptions.

REFERENCES

Key references are numbered as they appear in the paper.

[7] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman. The Devil is in the Details: An Evaluation of Recent Feature Encoding Methods. BMVC, 2011.

[19] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. CVPR, 2006.

[21] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 2004.

[29] F. Perronnin, J. Sanchez, and T. Mensink, Improving the Fisher Kernel for Large-scale Image Classification. ECCV, 2010.

[39] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-Constrained Linear Coding for Image Classification. CVPR, 2010.

[46] L. Xie, Q. Tian, J. Wang, and B. Zhang. Max-SIFT: Flipping Invariant Descriptors for Web Logo Search. ICIP, 2014.

ACKNOWLEDGEMENTS

This work was done when Lingxi Xie was an intern at MSR, Beijing. This work is supported by the 973 Program of China 2013CB329403 and 2012CB316301, NSFC 61332007, 61273023, 61429201, 61471235, Tsinghua ISRP 20121088071, ARO grants W911NF-15-1-0290 and W911NF-12-1-0057, and Faculty Research Awards, NEC Lab of America.