

KAZE FeaturePoint with Modified-SIFT Descriptor

Meixi Chen¹, Yule Yuan², Yong Zhao³

Abstract. In this paper, we present a novel descriptor, called Modified-SIFT (M-SIFT), which is more suitable for KAZE feature point than other descriptors such as SURF and SIFT. Their combining even outperforms SIFT matching algorithm and standard KAZE algorithm. We present extensive experimental image matching results on the Mikolajczyk and Schmid dataset which show clear advantages of our descriptor against others.

Keywords: feature descriptor; local feature; image match

1 Introduction

Given two or more images of a scene, the ability to match corresponding points between these images is an important component of many computer vision tasks such as image registration, object tracking, and object recognition.

On the basis of previous studies, Lowe proposed Scale Invariant Feature Transform (SIFT) [1] algorithm based on Gaussian scale space, and it had proven to be the most popular multiscale feature detection and description algorithm. However, it has some important drawbacks. Gaussian blurring does not respect the natural boundaries of objects and it smoothes details and noise in the same extent at all scale levels, the price to pay for this is a reduction in localization accuracy.

We should find a way to make blurring locally adaptive to the image data so

¹Meixi Chen

Key Lab. of Integrated Microsystem Sci. & Eng. Applications, Peking University Shenzhen Graduate School, China
e-mail: chenmeixi@sz.pku.edu.cn

²Yule Yuan

Key Lab. of Integrated Microsystem Sci. & Eng. Applications, Peking University Shenzhen Graduate School, China
e-mail: yuanyl@pkusz.edu.cn

³Yong Zhao (✉)

Key Lab. of Integrated Microsystem Sci. & Eng. Applications, Peking University Shenzhen Graduate School, China
e-mail: yongzhao@pkusz.edu.cn

that noise will be blurred, but details or edges will remain unaffected. To achieve this, Alcantarilla et al proposed KAZE features [2], a novel multiscale 2D feature detection and description algorithm in nonlinear scale spaces. They exploit nonlinear diffusion filtering in the context of multiscale feature detection and description using efficient schemes.

In this paper we go deeper in study of the descriptor of KAZE feature, and propose to find the most suitable descriptor. Our descriptor is named Modified SIFT (M-SIFT), based on SIFT descriptor.

The remainder of this paper is organized as follows: Section 2 gives an introduction to nonlinear diffusion filtering. Then, we describe KAZE feature detection in Section 3. Modified SIFT descriptor is explained in Section 4. Experimental results are provided in Section 5 and conclusions are presented in Section 6.

2 Nonlinear Diffusion Filtering

2.1 Peron -Malik Diffusion Equation

Nonlinear diffusion approaches describe the evolution of the luminance of an image through increasing scale levels as the divergence of a certain flow function that controls the diffusion process. These approaches are normally described by nonlinear partial differential equations (PDEs):

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \quad (2.1)$$

where div and ∇L are respectively the divergence and gradient operators. By setting the appropriate conduction function $c(x, y, t)$, we can make the diffusion adaptive to the local image structure. The time t is the scale parameter, and the larger its value, the image representations are more simple. Perona and Malik proposed conduction function [4] is defined as:

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|), \quad g = \frac{1}{1 + \frac{|\nabla L_\sigma|^2}{k^2}} \quad (2.2)$$

where ∇L_σ is the gradient of a Gaussian smoothed version of the original image L . The parameter k is the contrast factor that controls the level of diffusion, it determines which edges have to be enhanced and which have to be canceled. In KAZE algorithm, it takes an empirical value for k as the 70% percentile of the gradient histogram of a smoothed version of the original image.

2.2 AOS algorithm

For nonlinear partial differential equations, there are no analytical solutions. Generally we use numerical analysis method to get the iterative solution. One possible discretization of the diffusion equation is the so-called linear-implicit or semi-implicit scheme [5], it can be expressed as:

$$\frac{L^{i+1} - L^i}{\tau} = \sum_{l=1}^m A_l (L^i) L^{i+1} \quad (2.3)$$

where A_l is a matrix that encodes the image conductivities for each dimension. Fortunately, this linear system can be quickly solved by means of the Thomas algorithm and obtain the solution L^{i+1} as:

$$L^{i+1} = \left(I - \tau \sum_{l=1}^m A_l (L^i) \right)^{-1} L^i \quad (2.4)$$

3 KAZE Feature Detection

There are two main steps for KAZE feature detection. Firstly we build the nonlinear scale space using AOS techniques and PDEs. Then, we detect feature points of interest which are the local maxima of the scale-normalized determinant of the Hessian response through the nonlinear scale space.

3.1 Nonlinear Scale Space Building

We discretize the scale space in logarithmic steps arranged in a series of O octaves and S sub-levels as done in SIFT. However, we always work with the original image resolution, without performing any downsampling at each new octave. The octave index o and the sub-level index s are mapped to their corresponding scale σ through the following formula:

$$\sigma_i(o, s) = \sigma_0 2^{o+s/S}, o \in [0 \dots O-1], s \in [0 \dots S-1], i \in [0 \dots N], \quad (3.1)$$

where σ_0 is the base scale level and M is the total number of filtered images. For obtaining a set of evolution times from which we build the nonlinear scale space, we use the following mapping $\sigma_i \rightarrow t_i$:

$$t_i = \frac{1}{2} \sigma_i^2, i = \{0 \dots N\}, \quad (3.2)$$

Given an input image, we firstly compute the image gradient histogram and obtain the contrast parameter k . Then, given the contrast parameter and the set of evolution times t_i , we build the nonlinear scale space in an iterative way using the AOS schemes (which are absolutely stable for any step size) as:

$$L^{i+1} = \left(I - (t_{i+1} - t_i) \sum_{l=1}^m A_l (L^i) \right)^{-1} L^i. \quad (3.3)$$

3.2 Feature Detection

We compute the response of scale-normalized determinant of the hessian at multiple scale levels for detecting points of interest [8]. Hessian matrix is calculated as follows:

$$L_{Hessian} = \sigma^2 (L_{xx} L_{yy} - L_{xy}^2), \quad (3.4)$$

where (L_{xx}, L_{yy}) are the second order horizontal and vertical derivatives respectively, and L_{xy} is the second order cross derivative. In order to detect the local maxima of the nonlinear scale space L_i , each sample point is compared to its eight neighbors in the current filtered image and nine neighbors in the scale above and below.

4. Modified SIFT descriptor

4.1 Finding the Dominant Orientation.

In order to achieve invariance to image rotation, we identify a dominant orientation for the interest points. Similar to SURF [3], we first calculate first order derivatives L_x and L_y in a circular neighbourhood of radius $6\sigma_i$ with a sampling step of size σ_i . These first order derivatives are weighted with a Gaussian centered at the interest point. Then, the derivative responses are represented as points in vector space and the dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of size $\pi/3$. The longest such vector over all windows defines the dominant orientation.

4.2 Building descriptor

The descriptor for each point is created based on a patch of pixels in its local neighborhood. Since we get the location, scale, orientation of each keypoint, the descriptor is robust to the rotation of image. The descriptor is computed as a set of histograms on 16×16 pixel neighborhoods. Every histogram contains 4×4 subregions and every subregion contains 8 bins for directions, this results in an 8-dimensional vector. Finally, the descriptor of each keypoint contains a $128(4 \times 4 \times 8)$ element of feature vector. In order to reduce the effects of illumination change, the vector is normalized to the unit length.

5 Experiment results and conclusions

These experiments are carried out on an Intel(R) Pentium(R) CPU G630 @ 2.70 GHz computer with 2G RAM and the operation system is windows XP. All experiments run on the VS2010 and OpenCV 2.4.3.

In this section, we present extensive experimental results obtained on the standard evaluation set of Mikolajczyk et al [6, 7]. This standard dataset includes eight image sets (each sequence generally contains 6 images) with different geometric and photometric transformations such as image blur, lighting, viewpoint, scale changes, zoom, rotation and JPEG compression.

We compare our method against SIFT, standard KAZE, KAZE feature point with SIFT descriptor and KAZE feature point with SURF descriptor.

5.1 Time consumption

We compute the average detection time of three different features on eight image sets of the Mikolajczyk and Schmid dataset. Table 1 shows the detection time of each image in same image sets. We can find the KAZE feature point average detection time is close even less than the SURF feature point for whole image. This is also an important reason that we use KAZE feature point.

(ms)	bark	bikes	boat	graf	leuven	trees	ubc	wall
SIFT	382	388	361	466	203	989	388	612
SURF	989	936	1295	1226	842	1848	1204	2002
KAZE	649	1166	1197	1015	901	1750	965	1326

Table 1 .average detection time of each image

5.2 Matching performance

Initially we intend to adopt the method of combining kaze feature point with SIFT descriptor, but we find it does not have rotational invariance. The Fig1 shows the match result of the method of combining KAZE feature point with SIFT descriptor, it can't find any match pairs between the left image with the right. The Fig2 shows the match result of our method, it finds enough match pairs. In order to evaluate its overall performance, we have done more detailed experiments in the public datasets.



Fig1. KAZE+SIFT (Brute Force match)

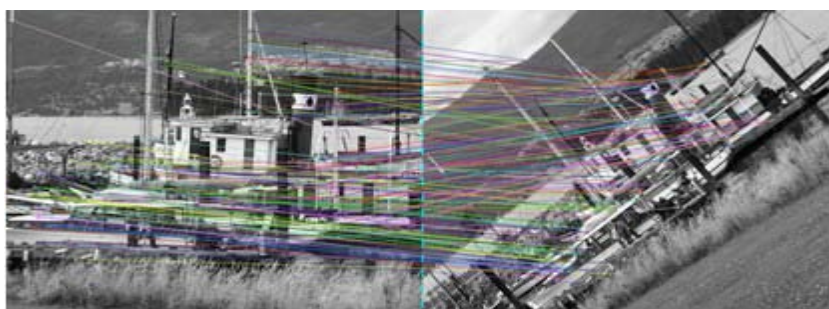
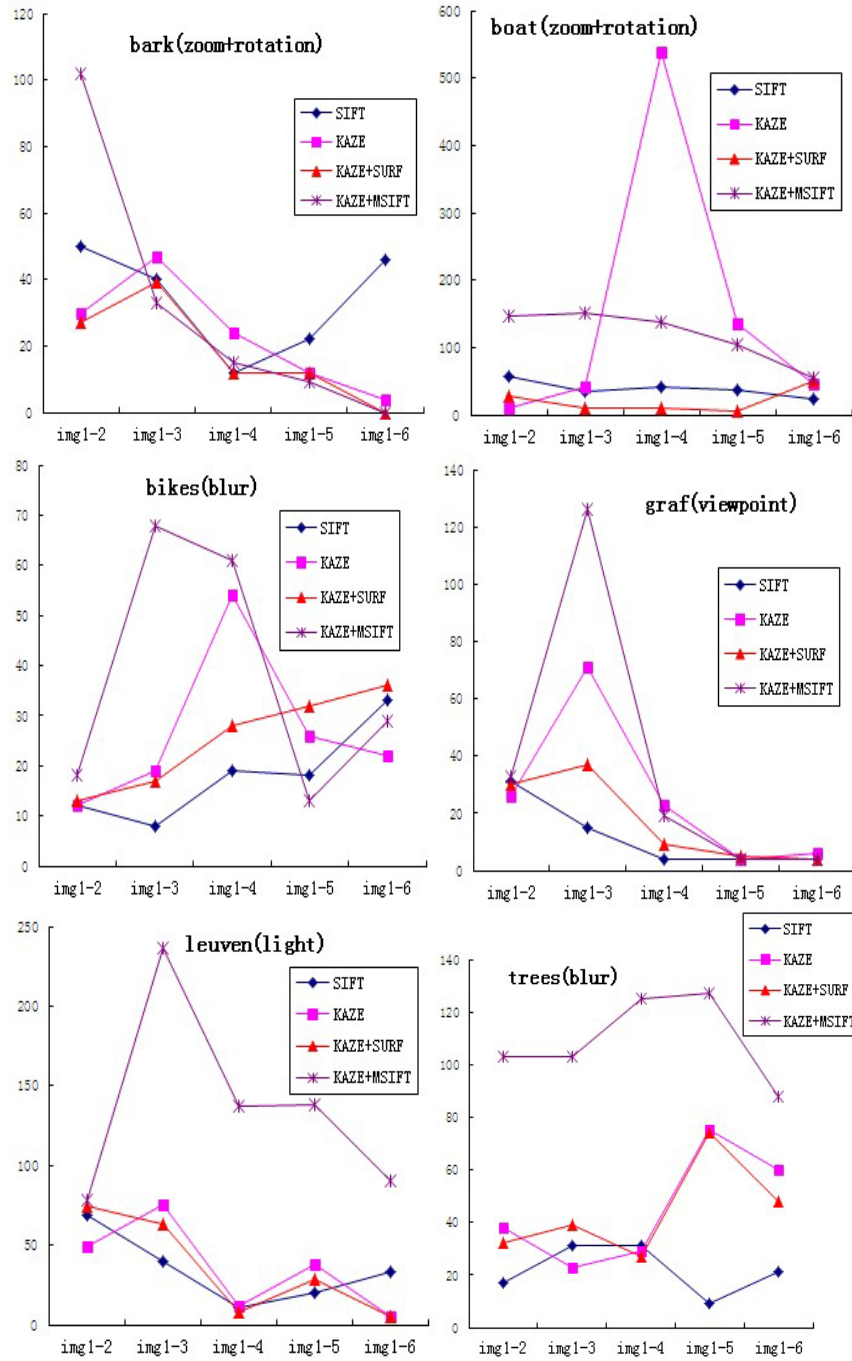


Fig2. KAZE+M-SIFT (Brute Force match)

We respectively match the first image with the other five images in eight image sets, and compute the matching number in four different algorithms. In order to improve matching accuracy, we use RANSAC algorithm [9] to optimize our results. As shown in Fig3, our method of combining KAZE feature point with M-SIFT descriptor is better than others. Although in some image sets advantage is not very obvious, such as bark sets, because KAZE feature is weaker than SIFT feature on the scale invariance. In other sets, our method shows better performance for viewpoint changes, image rotation, blur, jpeg compression and illumination changes than others.



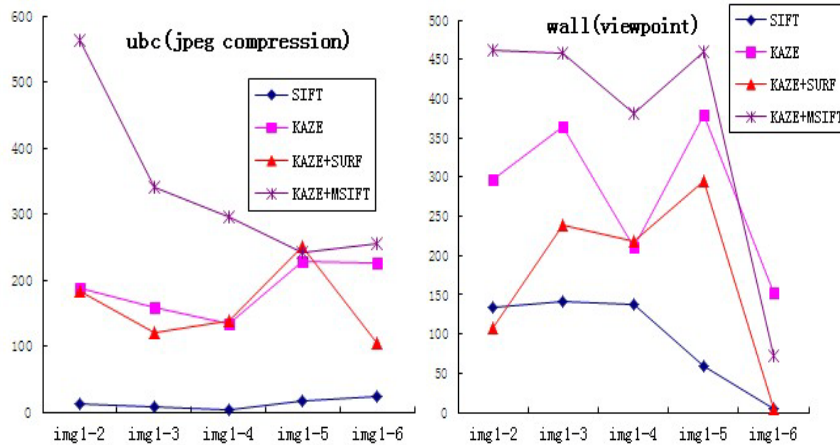


Fig3. matching numbers in eight image sets for four different algorithms (KAZE, SIFT, KAZE+SURF and KAZE+M-SIFT), img1-2 means the first image match with the second image

6 Conclusions

In this paper, we have presented M-SIFT descriptor, and our results show a better performance against SIFT and standard KAZE matching algorithms. In addition, our descriptor is more suitable for KAZE feature than SURF and SIFT descriptor.

7 References

1. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Intl. J. of Computer Vision* 60(2004) 91–110
2. Fernandez, P., A. Bartoli, and A. J. Davison.: KAZE Features. In *Proceedings of European Conference on Computer Vision*, 214–227. Firenze, Italy, October 7–13, 2012.
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. *Computer Vision and Image Understanding* 110(2008) 346–359
4. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.* 12(1990) 1651–1686
5. Weickert, J., ter Haar Romeny, B., Viergever, M.A.: Efficient and reliable schemes for non-linear diffusion filtering. *IEEE Trans. Image Processing* 7(1998)
6. Mikolajczyk, K., Tuytelaars, T., Schmid, C., et.al.: A comparison of affine region detectors. *Intl. J. of Computer Vision* 65(2005) 43–72
7. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Machine Intell.* 27(2005) 1615–1630
8. T. Lindeberg, Feature detection with automatic scale selection, *IJCV* 30 (2) (1998) 79–116

9. Fischler M, Bolles R.:Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24 (1981)