KAZE Features

Pablo Fernández Alcantarilla[1], Adrien Bartoli[1], and Andrew J. Davison[2]

[1] ISIT-UMR 6284 CNRS, Université d'Auvergne, Clermont Ferrand, France
{pablo.alcantarilla,adrien.bartoli}@gmail.com
[2] Department of Computing, Imperial College London, UK
ajd@doc.ic.ac.uk

Abstract. In this paper, we introduce KAZE features, a novel multiscale 2D feature detection and description algorithm in nonlinear scale spaces. Previous approaches detect and describe features at different scale levels by building or approximating the Gaussian scale space of an image. However, Gaussian blurring does not respect the natural boundaries of objects and smoothes to the same degree both details and noise, reducing localization accuracy and distinctiveness. In contrast, we detect and describe 2D features in a nonlinear scale space by means of nonlinear diffusion filtering. In this way, we can make blurring locally adaptive to the image data, reducing noise but retaining object boundaries, obtaining superior localization accuracy and distinctiviness. The nonlinear scale space is built using efficient Additive Operator Splitting (AOS) techniques and variable conductance diffusion. We present an extensive evaluation on benchmark datasets and a practical matching application on deformable surfaces. Even though our features are somewhat more expensive to compute than SURF due to the construction of the nonlinear scale space, but comparable to SIFT, our results reveal a step forward in performance both in detection and description against previous state-of-the-art methods.

1 Introduction

Multiscale image processing is a very important tool in computer vision applications. We can abstract an image by automatically detecting features of interest at different scale levels. For each of the detected features an invariant local description of the image can be obtained. These multiscale feature algorithms are a key component in modern computer vision frameworks, such as scene understanding [1], visual categorization [2] and large scale 3D Structure from Motion (SfM) [3].

The main idea of multiscale methods is quite simple: Create the scale space of an image by filtering the original image with an appropriate function over increasing time or scale. In the case of the Gaussian scale space, this is done by convolving the original image with a Gaussian kernel of increasing standard deviation. For larger kernel values we obtain simpler image representations. With a multiscale image representation, we can detect and describe image features at different scale levels or resolutions. Several authors [4,5] have shown that under some general assumptions, the Gaussian kernel and its set of partial derivatives are possible smoothing kernels for scale space analysis. However, it is important to note here that the Gaussian scale space is just one instance of linear diffusion, since other linear scale spaces are also possible [6].

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part VI, LNCS 7577, pp. 214–227, 2012.

The Gaussian kernel is probably the simplest option (but not the only one) to build a scale space representation of an image. However, it has some important drawbacks. In Gaussian scale space, the advantages of selecting coarser scales are the reduction of noise and the emphasis of more prominent structure. The price to pay for this is a reduction in localization accuracy. The reason for this is the fact that Gaussian blurring does not respect the natural boundaries of objects and smoothes to the same degree both details and noise at all scale levels. This loss in localization increases as long as we detect features at coarser scale levels, where the amount of Gaussian blurring is higher.

It seems more appropriate to make blurring locally adaptive to the image data so that noise will be blurred, but details or edges will remain unaffected. To achieve this, different nonlinear scale space approaches have been proposed to improve on the Gaussian scale space approach [7,8]. In general, nonlinear diffusion approaches perform much better than linear ones [9,10] and impressive results have been obtained in different applications such as image segmentation [11] or denoising [12]. However, to the best of our knowledge, this paper is the first one that exploits nonlinear diffusion filtering in the context of multiscale feature detection and description using efficient schemes. By means of nonlinear diffusion, we can increase repeatability and distinctiviness when detecting and describing an image region at different scale levels through a nonlinear scale space.

Probably one of the reasons why nonlinear diffusion filtering has not been used more often in practical computer vision components such as feature detection and description is the poor efficiency of most of the approaches. These approaches normally consist of the discretization of a function by means of the *forward Euler scheme*. The Euler scheme requires very small step sizes for convergence, and hence many iterations to reach a desired scale level and high computational cost. Fortunately, Weickert *et al.* introduced efficient schemes for nonlinear diffusion filtering in [9]. The backbone of these schemes is the use of Additive Operator Splitting (AOS) techniques. By means of AOS schemes we can obtain stable nonlinear scale spaces for any step size in a very efficient way. One of the key issues in AOS schemes is solving a tridiagonal system of linear equations, which can be efficiently done by means of the *Thomas algorithm*, a special variant of the Gaussian elimination algorithm.

In this paper we propose to perform automatic feature detection and description in nonlinear scale spaces. We describe how to build nonlinear scale spaces using efficient AOS techniques and variable conductance diffusion, and how to obtain features that exhibit high repeatability and distinctiveness under different image transfromations. We evaluate in detail our novel features within standard evaluation frameworks [13,14] and a practical image matching application using deformable surfaces.

Our features are named KAZE, in tribute to Iijima [15], the father of scale space analysis. KAZE is a Japanese word that means *wind*. In nature wind is defined as the flow of air on a large scale and normally this flow is ruled by nonlinear processes. In this way, we make the analogy with nonlinear diffusion processes in the image domain. The rest of the paper is organized as follows: In Section 2 we describe the related work. Then, we briefly introduce the basis of nonlinear diffusion filtering in Section 3.

The KAZE features algorithm is explained in detail in Section 4. Finally, exhaustive experimental results and conclusions are presented in Section 5 and 6 respectively.

2 Related Work

Feature detection and description is a very active field of research in computer vision. Obtaining features that exhibit high repeatability and distinctiveness against different image transformations (e.g. viewpoint, blurring, noise, etc.) is of extreme importance in many different applications. The most popular multiscale feature detection and description algorithms are the Scale Invariant Feature Transform (SIFT) [16] and the Speeded Up Robust Features (SURF) [17].

SIFT features were a milestone in feature detection and image matching and are still widely used in many different fields such as mobile robotics and object recognition. In SIFT, feature locations are obtained as the maxima and minima of the result of a Difference of Gaussians (DoG) operator applied through a Gaussian scale space. For building the scale space, a pyramid of Gaussian blurred versions of the original image is computed. The scale space is composed of different sublevels and octaves. For the set of detected features, a descriptor is built based on the main gradient orientation over a local area of interest of the detected keypoint. Then, a rectangular grid of normally 4×4 subregions is defined (according to the main orientation) and a histogram of the gradient orientations weighted by its magnitude is built, yielding a descriptor vector of 128 elements.

Inspired by SIFT, Bay *et al.* proposed the SURF detector and descriptor. SURF features exhibit better results with respect to repeatability, distinctiveness and robustness, but at the same time can be computed much faster thanks to the use of the integral image [18], meaning that Gaussian derivatives at different scale levels can be approximated by simple box filters without computing the whole Gaussian scale space. Similar to SIFT, a rectangular grid of 4×4 subregions is defined (according to the main orientation) and a sum of Haar wavelet responses (weighted by a Gaussian centered at the interest keypoint) is computed per region. The final descriptor dimension is normally 64 or 128 in its extended counterpart. In [19], Agrawal and Konolige introduced some improvements over SURF by using center-surround detectors (CenSurE) and the Modified-SURF (M-SURF) descriptor. The M-SURF is a variant of the original SURF descriptor, but handles better descriptor boundary effects and uses a more robust and intelligent two-stage Gaussian weighting scheme.

Both of these approaches and the many related algorithms which have followed rely on the use of the Gaussian scale space and sets of Gaussian derivatives as smoothing kernels for scale space analysis. However, to repeat, Gaussian scale space does not respect the natural boundaries of objects and smoothes to the same degree both details and noise at all scale levels. In this paper we will show that by means of nonlinear diffusion filtering it is possible to obtain multiscale features that exhibit much higher repeatability and distinctiveness rates than previous algorithms that are based on the Gaussian scale space. At the cost of a moderate increase in computational cost compared to SURF or CenSurE, our results reveal a big step forward in performance in both feature detection and description.

3 Nonlinear Diffusion Filtering

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), due to the nonlinear nature of the involved differential equations that diffuse the luminance of the image through the nonlinear scale space. Equation 1 shows the classic nonlinear diffusion formulation:

$$\frac{\partial L}{\partial t} = div\left(c\left(x,y,t\right)\cdot\nabla L\right) , \tag{1}$$

where *div* and ∇ are respectively the divergence and gradient operators. Thanks to the introduction of a *conductivity* function (c) in the diffusion equation, it is possible to make the diffusion adaptive to the local image structure. The function c depends on the local image differential structure, and this function can be either a scalar or a tensor. The time t is the scale parameter, and larger values lead to simpler image representations. In this paper, we will focus on the case of variable conductance diffusion, where the image gradient magnitude controls the diffusion at each scale level.

3.1 Perona and Malik Diffusion Equation

Nonlinear diffusion filtering was introduced in the computer vision literature in [7]. Perona and Malik proposed to make the function c dependent on the gradient magnitude in order to reduce the diffusion at the location of edges, encouraging smoothing within a region instead of smoothing across boundaries. In this way, the function c is defined as:

$$c\left(x,y,t\right) = g\left(\left|\nabla L_{\sigma}\left(x,y,t\right)\right|\right) , \tag{2}$$

where the luminance function ∇L_{σ} is the gradient of a Gaussian smoothed version of the original image L. Perona and Malik described two different formulations for the conductivity function g:

$$g_1 = \exp\left(-\frac{|\nabla L_{\sigma}|^2}{k^2}\right) , \; g_2 = \frac{1}{1+\frac{|\nabla L_{\sigma}|^2}{k^2}} , \tag{3}$$

where the parameter k is the contrast factor that controls the level of diffusion. The function g_1 promotes high-contrast edges, whereas g_2 promotes wide regions over smaller ones. Weickert [11] proposed a slightly different diffusion function for rapidly decreasing diffusivities, where smoothing on both sides of an edge is much stronger than smoothing across it. That selective smoothing prefers intraregional smoothing to interregional blurring. This function, which we denote here as g_3, is defined as follows:

$$g_3 = \begin{cases} 1 & , |\nabla L_{\sigma}|^2 = 0 \\ 1 - \exp\left(-\frac{3.315}{(|\nabla L_{\sigma}|/k)^8}\right) & , |\nabla L_{\sigma}|^2 > 0 \end{cases} . \tag{4}$$

The contrast parameter k can be either fixed by hand or automatically by means of some estimation of the image gradient. The contrast factor determines which edges

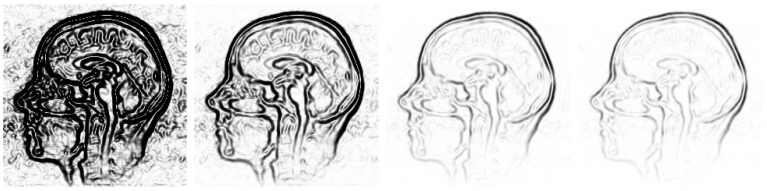


Fig. 1. The conductivity coefficient g_1 in the Perona and Malik equation as a function of the parameter k. Notice that for increasing values of k only higher gradients are considered. We consider grey scale images of range 0-255.

have to be enhanced and which have to be canceled. In this paper we take an empirical value for k as the 70% percentile of the gradient histogram of a smoothed version of the original image. This empirical procedure gives in general good results in our experiments. However, it is possible that for some images a more detailed analysis of the contrast parameter can give better results. Figure 1 depicts the conductivity coefficient g_1 in the Perona and Malik equation for different values of the parameter k. In general, for higher k values only larger gradients are taken into account.

3.2 AOS Schemes

There are no analytical solutions for the PDEs involved in nonlinear diffusion filtering. Therefore, one needs to use numerical methods to approximate the differential equations. One possible discretization of the diffusion equation is the so-called *linear-implicit* or *semi-implicit* scheme. In a vector-matrix notation and using a similar notation to [9], the discretization of Equation 1 can be expressed as:

$$\frac{L^{i+1} - L^i}{\tau} = \sum_{l=1}^{m} A_l \left(L^i\right) L^{i+1} , \tag{5}$$

where A_l is a matrix that encodes the image conductivities for each dimension. In the semi-implicit scheme, for computing the solution L^{i+1}, one needs to solve a linear system of equations. The solution L^{i+1} can be obtained as:

$$L^{i+1} = \left(I - \tau \sum_{l=1}^{m} A_l \left(L^i\right) \right)^{-1} L^i . \tag{6}$$

The semi-implicit scheme is absolutely stable for any step size. In addition, it creates a discrete nonlinear diffusion scale-space for arbitrarily large time steps. In the semi-implicit scheme, it is necessary to solve a linear system of equations, where the system matrix is tridiagonal and diagonally dominant. Such systems can be solved very efficiently by means of the *Thomas algorithm*, which is a variation of the well-known Gaussian elimination algorithm for tridiagonal systems.

4 KAZE Features

In this section, we describe our novel method for feature detection and description in nonlinear scale spaces. Given an input image, we build the nonlinear scale space up to a maximum evolution time using AOS techniques and variable conductance diffusion. Then, we detect 2D features of interest that exhibit a maxima of the scale-normalized determinant of the Hessian response through the nonlinear scale space. Finally, we compute the main orientation of the keypoint and obtain a scale and rotation invariant descriptor considering first order image derivatives. Now, we will describe each of the main steps in our formulation.

4.1 Computation of the Nonlinear Scale Space

We take a similar approach as done in SIFT, discretizing the scale space in logarithmic steps arranged in a series of O octaves and S sub-levels. Note that we always work with the original image resolution, without performing any downsampling at each new octave as done in SIFT. The set of octaves and sub-levels are identified by a discrete octave index o and a sub-level one s. The octave and the sub-level indexes are mapped to their corresponding scale σ through the following formula:

$$\sigma_i(o,s) = \sigma_0 2^{o+s/S},\ o \in [0 \dots O-1],\ s \in [0 \dots S-1],\ i \in [0 \dots N], \qquad (7)$$

where σ_0 is the base scale level and N is the total number of filtered images. Now, we need to convert the set of discrete scale levels in pixel units σ_i to time units. The reason of this conversion is because nonlinear diffusion filtering is defined in time terms. In the case of the Gaussian scale space, the convolution of an image with a Gaussian of standard deviation σ (in pixels) is equivalent to filtering the image for some time $t = \sigma^2/2$. We apply this conversion in order to obtain a set of evolution times and transform the scale space $\sigma_i(o,s)$ to time units by means of the following mapping $\sigma_i \to t_i$:

$$t_i = \frac{1}{2}\sigma_i^2,\ i = \{0 \dots N\}, \qquad (8)$$

It is important to mention here that we use the mapping $\sigma_i \to t_i$ only for obtaining a set of evolution times from which we build the nonlinear scale space. In general, in the nonlinear scale space at each filtered image t_i the resulting image does not correspond with the convolution of the original image with a Gaussian of standard deviation σ_i. However, our framework is also compatible with the Gaussian scale space in the sense that we can obtain the equations for the Gaussian scale space by setting the diffusion function g to be equal to 1 (i.e. a constant function). In addition, as long as we evolve through the nonlinear scale space the conductivity function tends to be constant for most of the image pixels except for the strong image edges that correspond to the objects boundaries.

Given an input image, we firstly convolve the image with a Gaussian kernel of standard deviation σ_0 to reduce noise and possible image artefacts. From that base image we compute the image gradient histogram and obtain the contrast parameter k in an automatic procedure as described in Section 3.1. Then, given the contrast parameter and

Fig. 2. Comparison between the Gaussian and nonlinear diffusion scale space for several evolution times t_i. First Row: Gaussian scale space (linear diffusion). The scale space is formed by convolving the original image with a Gaussian kernel of increasing standard deviation. Second Row: Nonlinear diffusion scale space with conductivity function g_3.

the set of evolution times t_i, it is straightforward to build the nonlinear scale space in an iterative way using the AOS schemes (which are absolutely stable for any step size) as:

$$L^{i+1} = \left(I - (t_{i+1} - t_i) \cdot \sum_{l=1}^{m} A_l \left(L^i \right) \right)^{-1} L^i \, . \tag{9}$$

Figure 2 depicts a comparison between the Gaussian scale space and the nonlinear one (using the g_3 conductivity function) for several evolution times given the same reference image. As it can be observed, Gaussian blurring smoothes for equal all the structures in the image, whereas in the nonlinear scale space strong image edges remain unaffected.

4.2 Feature Detection

For detecting points of interest, we compute the response of scale-normalized determinant of the Hessian at multiple scale levels. For multiscale feature detection, the set of differential operators needs to be normalized with respect to scale, since in general the amplitude of spatial derivatives decrease with scale [5]:

$$L_{Hessian} \; = \; \sigma^2 \left(L_{xx} L_{yy} - L_{xy}^2 \right) \, , \tag{10}$$

where (L_{xx}, L_{yy}) are the second order horizontal and vertical derivatives respectively, and L_{xy} is the second order cross derivative. Given the set of filtered images from the nonlinear scale space L^i, we analyze the detector response at different scale levels σ_i. We search for maxima in scale and spatial location. The search for extrema is performed in all the filtered images except $i = 0$ and $i = N$. Each extrema is searched over a rectangular window of size $\sigma_i \times \sigma_i$ on the current i, upper $i + 1$ and lower $i - 1$ filtered images. For speeding-up the search for extrema, we firstly check the responses over a window of size 3×3 pixels, in order to discard quickly non-maxima responses. Finally, the position of the keypoint is estimated with sub-pixel accuracy using the method proposed in [20].

The set of first and second order derivatives are approximated by means of 3×3 Scharr filters of different derivative step sizes σ_i. Second order derivatives are approximated by using consecutive Scharr filters in the desired coordinates of the derivatives. These filters approximate rotation invariance significantly better than other popular filters such as Sobel filters or standard central differences differentiation [21]. Notice here that although we need to compute multiscale derivatives for every pixel, we save computational efforts in the description step, since we re-use the same set of derivatives that are computed in the detection step.

4.3 Feature Description

Finding the Dominant Orientation. For obtaining rotation invariant descriptors, it is necessary to estimate the dominant orientation in a local neighbourhood centered at the keypoint location. Similar to SURF, we find the dominant orientation in a circular area of radius $6\sigma_i$ with a sampling step of size σ_i. For each of the samples in the circular area, first order derivatives L_x and L_y 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 found by summing the responses within a sliding circle segment covering an angle of $\pi/3$. From the longest vector the dominant orientation is obtained.

Building the Descriptor. We use the M-SURF descriptor adapted to our nonlinear scale space framework. For a detected feature at scale σ_i, first order derivatives L_x and L_y of size σ_i are computed over a $24\sigma_i \times 24\sigma_i$ rectangular grid. This grid is divided into 4×4 subregions of size $9\sigma_i \times 9\sigma_i$ with an overlap of $2\sigma_i$. The derivative responses in each subregion are weighted with a Gaussian ($\sigma_1 = 2.5\sigma_i$) centered on the subregion center and summed into a descriptor vector $d_v = (\sum L_x, \sum L_y, \sum |L_x|, \sum |L_y|)$. Then, each subregion vector is weighted using a Gaussian ($\sigma_2 = 1.5\sigma_i$) defined over a mask of 4×4 and centered on the interest keypoint. When considering the dominant orientation of the keypoint, each of the samples in the rectangular grid is rotated according to the dominant orientation. In addition, the derivatives are also computed according to the dominant orientation. Finally, the descriptor vector of length 64 is normalized into a unit vector to achieve invariance to contrast.

5 Experimental Results and Discussion

In this section, we present extensive experimental results obtained on the standard evaluation set of Mikolajczyk *et al.* [13,14] and on a practical image matching application on deformable surfaces. The standard dataset includes several 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. In addition, the ground truth homographies are also available for every image transformation with respect to the first image of every sequence.

We also evaluate the performance of feature detectors and descriptors under image noise transformations. We created a new dataset named *Iguazu*. This dataset consists of

Fig. 3. Iguazu dataset images with increasing random Gaussian noise values per image

6 images, where the image transformation is the progressive addition of random Gaussian noise. For each pixel of the transformed images, we add random Gaussian noise with increasing variance considering grey scale value images. The noise variances for each of the images are the following: Image $2 \pm \mathcal{N}(0, 2.55)$, Image $3 \pm \mathcal{N}(0, 12.75)$, Image $4 \pm \mathcal{N}(0, 15.00)$, Image $5 \pm \mathcal{N}(0, 51.00)$ and Image $6 \pm \mathcal{N}(0, 102)$, considering that the grey value of each pixel in the image ranges from 0 to 255. Figure 3 depicts the *Iguazu* dataset.

We compare KAZE features against SURF, SIFT and CenSurE features. For SURF we use the original closed-source library[1] and for SIFT we use Vedaldi's implementation[2] [22]. Regarding CenSurE features we use the OpenCV based implementation, which is called STAR detector [3]. After detecting features with the STAR detector, we compute a M-SURF descriptor plus orientation as described in [19]. Therefore, we will denote in this section the STAR method as an approximation of CenSurE feature detector plus the computation of a M-SURF descriptor. We use for all the methods the same number of scales $O = 4$, and sublevels $S = 3$ for the SIFT and KAZE cases. The feature detection thresholds of the different methods are set to proper values to detect approximately the same number of features per image.

5.1 KAZE Detector Repeatability

The detector repeatability score between two images as defined in [13], measures the ratio between the corresponding keypoints and the minimum number of keypoints visible in both images. The overlap error is defined as the ratio of the intersection and union of the regions $\epsilon_s = 1 - (\mathbf{A} \bigcap \mathbf{H}^t \mathbf{B} \mathbf{H}) \,/\, (\mathbf{A} \bigcup \mathbf{H}^t \mathbf{B} \mathbf{H})$, where $\mathbf{A}$ and $\mathbf{B}$ are the two regions and $\mathbf{H}$ is the corresponding homography between the images. When the overlap error between two regions is smaller than 50%, a correspondence is considered.

Figure 4 depicts the repeatability scores for some selected sequences from the standard dataset. We show repeatability scores for SURF, SIFT, STAR and KAZE considering the different conductivities (g_1, g_2, g_3) explained in Section 3.1. As it can be observed, the repeatibility score of KAZE features clearly outperforms their competitors by a large margin for all the analyzed sequences. Regarding the Iguazu dataset (Gaussian noise), the repeatability score of the KAZE features is for some images 20% higher than SURF and STAR and 40% higher than SIFT. The reason for this is because nonlinear diffusion filtering smoothes the noise but at the same time keeps the boundaries of the objects, whereas Gaussian blurring smoothes in the same degree details and noise. Comparing the results of the different conductivities, g_2 exhibits a slightly

[1] Available from `http://www.vision.ee.ethz.ch/ surf/`
[2] Available from `http://www.vlfeat.org/`
[3] Available from `http://opencv.willowgarage.com/wiki/`

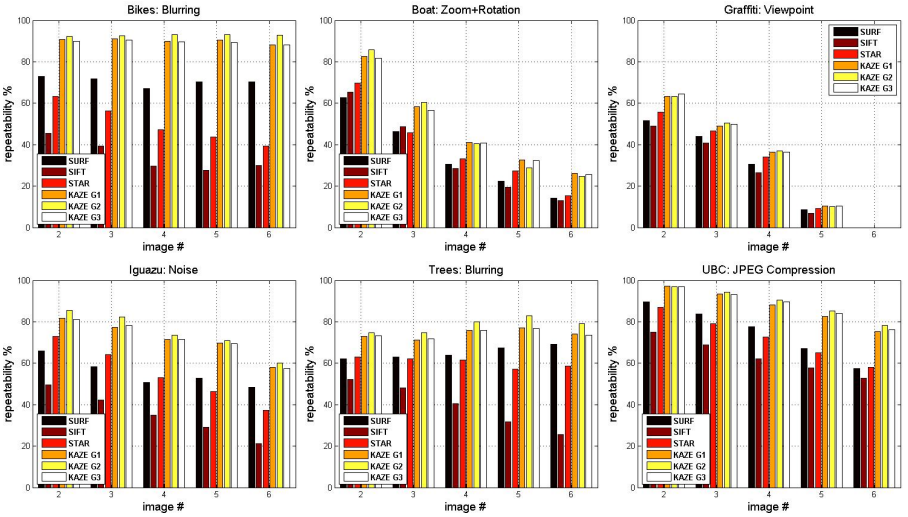


Fig. 4. Detector repeatability score for an overlap area error 50%. Best viewed in color.

higher repeatability. This can be explained by the fact that g_2 promotes wide area regions which are more suitable for blob-like features such as the ones detected by the determinant of the Hessian. In contrast g_1 and g_3 promote high-contrast edges which may be more suitable for corner detection.

5.2 Evaluation and Comparison of the Overall KAZE Features

We evaluate the joint performance of the detection, description and matching for each of the analyzed methods. Descriptors are evaluated by means of precision-recall graphs as proposed in [14]. This criterion is based on the number of correct matches and the number of false matches obtained for an image pair:

$$recall = \frac{\#correct\ matches}{\#correspondences}\ ,\ 1 - precision = \frac{\#false\ matches}{\#all\ matches}\ , \tag{11}$$

where the number of correct matches and correspondences is determined by the overlap error. For the Bikes, Iguazu, Trees and UBC sequences, we show results for the upright version of the descriptors (no dominant orientation) for all the methods. The upright version of the descriptors is faster to compute and usually exhibits higher performance (compared to its corresponding rotation invariant version) in applications where invariance to rotation is not necessary, such is the case of the mentioned sequences.

Figure 5 depicts precision-recall graphs considering the nearest neighbor matching strategy. As it can be seen, KAZE features obtain superior results thanks in part due to the much better detector repeatability in most of the sequences. For the Boat and Graffiti sequences SURF and SIFT obtain comparable results to KAZE features. However, the number of found correspondences by KAZE is approximately two times higher than the ones found by SURF and SIFT. Note that in all the analyzed image pairs, except the Boat and Graffiti ones, KAZE features exhibit recall rates sometimes 40% higher than SURF, SIFT and STAR for the same number of detected keypoints.

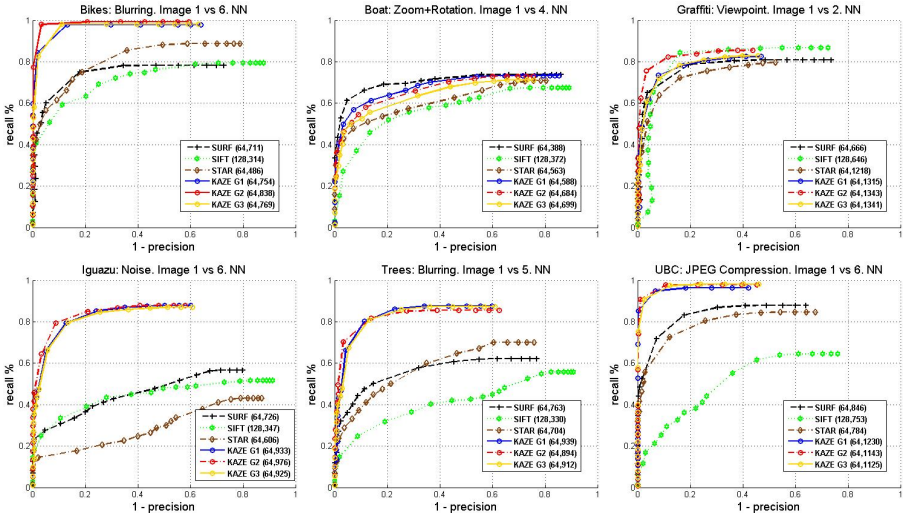
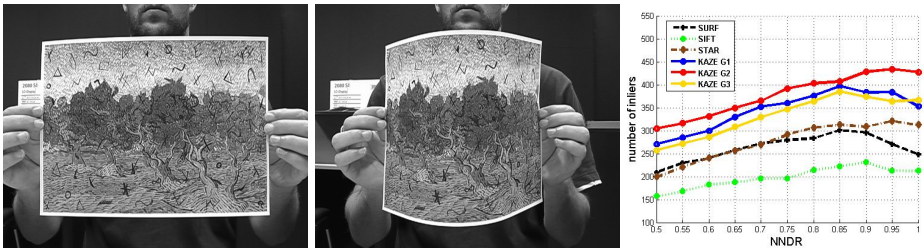


Fig. 5. Recall vs 1-precision graphs for nearest neighbor matching strategy. The graphs depict the overall results for detection, description and matching steps jointly for each of the methods. In parenthesis, next to the name of each of the methods we show the dimension of the descriptor and the number of found correspondences. Best viewed in color.

5.3 Image Matching for Deformable Surfaces

Complementary to the extensive evaluation on benchmark datasets, we also show results of image matching in deformable surfaces. In particular, we use the deformable surface detection method described in [23]. This method, based on local surface smoothness, is capable of discarding outliers from a set of putative matches between an image template and a deforming target image. In template-based deformable surface detection and reconstruction [24,25], is very important to have a high number of good correspondences between the template and the target image to capture more accurately the image deformation.

Figure 6(a,b) depicts two frames from the *paper* dataset [24] where we performed our image matching experiment. We detect features from the first image and then match these features to the extracted features on the second image. Firstly, a set of putative correspondences is obtained by using the nearest neighbor distance ratio (NNDR) strategy as proposed in [16]. This matching strategy takes into account the ratio of distance from the closest neighbor to the distance of the second closest. Then, we use the set of putative matches between the two images (that contains outliers) as the input for the outlier rejection method described in [23]. By varying the distance ratio, we can obtain a graph showing the number of inliers for different values of the distance ratio. Figure 6(c) depicts the number of inliers graphs obtained with SURF, SIFT, STAR and KAZE features for the analyzed experiment. According to the results, we can observe that KAZE features exhibit also good performance for image matching applications in deformable surfaces, yielding a higher number of inliers than their competitors.



(a) (b) (c)

Fig. 6. Image matching in deformable surfaces example. Two frames from the *paper* dataset [24]. (a) Frame 262 (b) Frame 315 (c) Number of inliers as a function of the nearest neighbor distance ratio. Best viewed in color.

5.4 Timing Evaluation

In this section we perform a timing evaluation for the most important operations in the process of computing KAZE features with conductivity function g_2 and a comparison with respect to SURF, SIFT and STAR. We take into account both the detection and the description of the features (computing a descriptor and dominant orientation or few of them in the case of SIFT). All timing results were obtained on a Core 2 Duo 2.4GHz laptop computer. Our KAZE code is implemented in C++ based on OpenCV data structures. The source code and the *Iguazu* dataset can be downloaded from `www.robesafe.com/personal/pablo.alcantarilla/kaze.html`.

Table 1. Computation times in seconds for the main steps of the KAZE features computation with conductivity function g_2 and comparison with respect to SURF, SIFT and STAR

KAZE	**UBC 1**	**Trees 6**
Nonlinear Scale Space	1.14	1.53
Feature Detection	0.68	0.93
Feature Description	0.38	0.20
Total Time	2.20	2.66
SURF	0.89	0.63
SIFT	2.66	2.77
STAR	0.25	0.32
Image Resolution	800×640	1000×700
Number of Keypoints	1463	765

In particular, Table 1 shows timing results in seconds for two images of different resolution from the standard dataset. As it can be observed, KAZE features are computationally more expensive than SURF or STAR, but comparable to SIFT. This is mainly due to the computation of the nonlinear scale space, which is the most consuming step in our method. However, at the cost of a slight increase in computational cost, our results reveal a big step forward in performance. In our implementation, we parallelized

the AOS schemes computation for each image dimension, since AOS schemes split the whole diffusion filtering in a sequence of 1D separable processes. Nevertheless, our method and implementation are subject to many improvements that can speed-up the computation of the KAZE features tremendously.

6 Conclusions and Future Work

In this paper, we have presented KAZE features, a novel method for multiscale 2D feature detection and description in nonlinear scale spaces. In contrast to previous approaches that rely on the Gaussian scale space, our method is based on nonlinear scale spaces using efficient AOS techniques and variable conductance diffusion. Despite of moderate increase in computational cost, our results reveal a step forward in performance both in detection and description against previous state-of-the-art methods such as SURF, SIFT or CenSurE.

In the next future we are interested in going deeper in nonlinear diffusion filtering and its applications for feature detection and description. In particular, we think that higher quality nonlinear diffusion filtering such as coherence-enhancing diffusion filtering [21] can improve our current approach substantially. In addition, we will work in the direction of speeding-up the method by simplifying the nonlinear diffusion process and by using GPGPU programming for real-time performance. Furthermore, we are also interested in using KAZE features for large-scale object recognition and deformable 3D reconstruction. Despite a tremendous amount of progress that has been made in the last few years in invariant feature matching, the final word has by no means been written yet, and we think nonlinear diffusion has many things to say.

Acknowledgments. Pablo F. Alcantarilla and Adrien Bartoli would like to acknowledge the support of ANR through project SYSEO. Andrew J. Davison was supported by ERC Starting Grant 210346.

References

1. Liu, C., Yuen, J., Torralba, A.: Dense scene alignment using SIFT flow for object recognition. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR (2009)
2. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR (2006)
3. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a Day. In: Intl. Conf. on Computer Vision, ICCV, Kyoto, Japan (2009)
4. Koenderink, J.: The structure of images. Biological Cybernetics 50, 363–370 (1984)
5. Lindeberg, T.: Feature detection with automatic scale selection. Intl. J. of Computer Vision 30, 77–116 (1998)
6. Duits, R., Florack, L., De Graaf, J., ter Haar Romeny, B.: On the axioms of scale space theory. Journal of Mathematical Imaging and Vision 20, 267–298 (2004)
7. Perona, P., Malik, J.: Scale-space and edge detection using annisotropic diffusion. IEEE Trans. Pattern Anal. Machine Intell. 12, 1651–1686 (1990)

8. Álvarez, L., Lions, P., Morel, J.: Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Numerical Analysis (SINUM) 29, 845–866 (1992)
9. Weickert, J., ter Haar Romeny, B., Viergever, M.A.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans. Image Processing 7 (1998)
10. ter Haar Romeny, B.M.: Front-End Vision and Multi-Scale Image Analysis. Multi-Scale Computer Vision Theory and Applications, written in Mathematica. Kluwer Academic Publishers (2003)
11. Weickert, J.: Efficient image segmentation using partial differential equations and morphology. Pattern Recognition 34, 1813–1824 (2001)
12. Qiu, Z., Yang, L., Lu, W.: A new feature-preserving nonlinear anisotropic diffusion method for image denoising. In: British Machine Vision Conf., BMVC, Dundee, UK (2011)
13. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. Intl. J. of Computer Vision 65, 43–72 (2005)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. 27, 1615–1630 (2005)
15. Weickert, J., Ishikawa, S., Imiya, A.: Linear scale-space has first been proposed in Japan. Journal of Mathematical Imaging and Vision 10 (1999)
16. Lowe, D.: Distinctive image features from scale-invariant keypoints. Intl. J. of Computer Vision 60, 91–110 (2004)
17. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. Computer Vision and Image Understanding 110, 346–359 (2008)
18. Viola, P., Jones, M.J.: Robust real-time face detection. Intl. J. of Computer Vision 57, 137–154 (2004)
19. Agrawal, M., Konolige, K., Blas, M.R.: CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 102–115. Springer, Heidelberg (2008)
20. Brown, M., Lowe, D.: Invariant features from interest point groups. In: British Machine Vision Conf., BMVC, Cardiff, UK (2002)
21. Weickert, J., Scharr, H.: A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. Journal of Visual Communication and Image Representation 13, 103–118 (2002)
22. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008), `http://www.vlfeat.org/`
23. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion reasoning. Intl. J. of Computer Vision 97, 54–70 (2012)
24. Salzmann, M., Hartley, R., Fua, P.: Convex optimization for deformable surface 3D tracking. In: Intl. Conf. on Computer Vision, ICCV, Rio de Janeiro, Brazil (2007)
25. Bartoli, A., Gérard, Y., Chadebecq, F., Collins, T.: On template-based reconstruction from a single view: Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island, USA (2012)