Neurocomputing 120 (2013) 380–390

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Performance evaluation of feature detection and matching in stereo visual odometry

Yunliang Jiang [c], Yunxi Xu [a,b,c], Yong Liu [a,b,*]

[a] Institute of Cyber-Systems and Control, Zhejiang Univeristy, 310027 Hangzhou, China
[b] State Key Laboratory of Industrial Control Technology, Zhejiang Univeristy, 310027 Hanzghou, China
[c] School of Information and Engineering, Huzhou Teachers College, 31300 Huzhou, China

ARTICLE INFO

*Article history:*
Received 26 December 2011
Received in revised form
18 April 2012
Accepted 5 June 2012
Available online 26 March 2013

*Keywords:*
Interest point detectors
Local descriptors
Matching
Stereo visual odometry
Evaluation

ABSTRACT

In this paper, we try to evaluate which detector and descriptor may be the most appropriate solution in stereo visual odometry and whether there is any bias on calculation methods in visual odometry applications. We summarize the state of art feature detectors and descriptors in visual odometry field and divide them based on their implemented details. We present three new evaluation criterions (*Detection Chain Repeatability*, *Average Detection Chain Re-projection Error* and *Matching Chain Precision*) of feature detectors and descriptors. We also design experiments to evaluate the performance of different detectors and descriptors from the robustness, precision and cost of computation.



## 1. Introduction

Feature detection and matching from images are hotspots in computer vision and robotics, and have been successfully implemented in many fields such as object recognition, 3D reconstruction, image retrieval, and camera localization etc. Recently, researchers have paid much attention on developing proper detectors and descriptors in visual odometry applications.

Visual odometry [1] is an important technique in robotics and assistance–driver systems, it can estimate the motion from the videos obtained by a single camera or multiple cameras. The motion between two consecutive frames is estimated from corresponding feature points in these two frames. The first step is to detect features in both frames with an interest point detector. Ideally, the detected features are the projections of the same 3D world points. In the second step, a feature descriptor is used to represent the detected features in a distinctive way. Correspondences are then obtained by using a matching strategy that compares feature descriptors based on a similarity measurement. During that process, the biggest challenge is data association, which means associate those feature points projected from a common world point in successive frames correctly. It is important that the features should be stable under the changes of lighting and viewpoint; it also should be distinctive and fast to compute so that features in successive frames can be tracked correctly by matching with descriptors.

Given the challenge on data association, the remaining question is which detector and descriptor is the most appropriate to estimate the accurate motion in visual odometry. Because there are varied calculation methods with respect to detectors and descriptors, whether the detectors and descriptors are in depended in visual odometry and whether there is any bias on calculation methods in visual odometry applications are all worth researching. As the monocular vision system is subject to scale uncertainty, in this work, we focus on detectors and descriptors implemented in stereo visual odometry.

The evaluation of the detectors and descriptors are performed in the standard framework of stereo visual odometry with different real world data sets. We have selected a number of detectors and descriptors which have previously shown a good performance in visual odometry, some newly detectors and descriptors are also selected to compare with the popular approaches using the same evaluation framework and data sets. The key problem in stereo visual odometry may quite different from the applications such as image mosaic, object recognition and image retrieval etc., thus the previous evaluation criterions [2–5] designed for image based applications may not be proper. And how to design the criterions suited for visual odometry is still an open challenge. In our evaluation, we design several new criterions, which are all concentrated on the challenge of data association on visual odometry.

The following sections are organized as follows: Section 2, we present a state of the art on detectors and descriptors

* Corresponding author at: Institute of Cyber-Systems and Control, Zhejiang Univeristy, 310027 Hangzhou, China. Tel.: +86 13805719977.
*E-mail addresses:* cckaffe@yahoo.com.cn, yongliu@iipc.zju.edu.cn (Y. Liu).

0925-2312/$ - see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.06.055

implemented in visual odometry applications and the related performance evaluation works. Section 3 presents the detailed stereo visual odometry framework in our evaluation. Sections 4 and 5 introduce the implementation details of detectors and descriptors evaluated in our experiments respectively. Section 6 presents our evaluation criterions and the experimental results. Finally, we conclude in Section 7.

## 2. Related works

To the present days, different combinations of detectors and descriptors have been used for stereo visual odometry. For example, Nister [1] used Harris corner detector [6] to find interest points in images and described them with a patch of image pixels centered at the detected points in monocular and stereo visual odometry. Gabor features and Haar features [7–9] are early important features in the field of computer vision and are widely used in object recognition fields. More recently, FAST [10] criterion for interest point detection has become increasingly popular in state-of-the-art methods with hard real-time constraints, and then AGAST [11] extended FAST with a better performance. Mei [12] used FAST in stereo visual SLAM. In recently, the BRISK [13], which combines FAST, AGAST, scale filter and can be regarded as a multiple-scale FAST corner detector, has been introduced. Besides corners, another most intuitive local feature is blob. Lowe [14,15] presented the scale-invariant feature transform (SIFT) detector and a descriptor scheme, which was initially applied to object recognition applications. SIFT uses scale-space approaches to achieve the desired scale invariance. First successful case of using SIFT features in a stereo odometry framework has been described by Se et al. [16]. SIFT is very popular and has been also widely applied in other computer vision fields such as image matching [17,18], image retrieval. SIFT detector uses difference of Gaussians (DoG) and the descriptor is based on gradient histograms with a 128-vector. They are relatively slow to compute. This can be a drawback for real-time applications such as visual odometry. PCA-SIFT [19] reduced the dimension of descriptor from 128 to 36 in compromising, however its lower distinctiveness and time-increasing in descriptor formation which almost annihilates the acceleration of matching speed. HOG is gradient histogram feature, similar to SIFT, and used widely in human detection [20–22]. Bay et al. [23] presented the SURF detector and descriptor, which was proved to be much faster than SIFT. Pang et al. [24] presented the FAIR-SURF which is an affine invariant version of SURF descriptor. Agrawal al et. [25] applied a center-symmetric local binary pattern as an alternative to SIFT's orientation histograms approach and used CensurE feature in large scale visual odometry for rough terrain. The most recent BRIEF [26] is designed for super-fast description which is a brightness comparison descriptor using simple binary tests between pixels in a smoothed image patch. As BRIEF is not scale invariance, it is sensitive to the large in-plane rotations and scale changes. To solve the disadvantages of BRIEF, Stefan et al. [13] presented the BRISK descriptor.

In the context of object recognition and image retrieval, many authors have presented their works evaluating interest point detectors and descriptors. A performance comparison among interest point detectors under scale invariance, viewpoint, lighting and manual noise conditions is carried out in [3]. Mikolajczyk et al. [4,27] have evaluated the affine based interest point detectors. Moreels and Perona [5] presented a performance evaluation addressed on interest point detectors of 3D objects. Performaces of detectors is evaluated by the repeatability rate, i.e., the percentage of points simultaneously present in two images. A higher the repeatability rate between two images will represent that more points can potentially be matched and the results of matching and recognition will be better. Carneiro and Jepson [28] evaluated the performance of feature descriptors using ROC (Receiver Operating Characteristics). Mikolajczyk and Schmid [2] compare the performance of descriptors computed for local interest regions. Their evaluation uses a criterion recall with respect to 1-precision and is carried out for different image transformations. Recall is the number of correctly matched regions with respect to the number of corresponding regions between two images of the same scene. The number of false matches relative to the total number of matches is represented by 1-precision. Local descriptors have also been evaluated in the context of texture classification [29,30].

Currently, the researches on performance evaluation of detectors and descriptors are almost focused in the context of the object recognition, image retrieval and texture classifying, and there are less work (or even ignored) focusing on the performance evaluation of real-time video applications. For visual odometry as a real-time video system, accuracy of feature localization and computation cost are crucial. Different from matching image applications with large viewpoint changes such as panorama stitching, object recognition and image retrieval, visual odometry is a video sequence matching between the successive frames. The challenge problem of detectors and descriptors in visual odometry is not focusing on the single criterion of change of viewpoint, lighting, or repeatability rate of image pairs; however, it may concern the assembly of varied factors, which will make influences on the estimation of motions, in the sequence of video. Thus the criterions and image dataset used in previous performance evaluation approaches may be not suitable for stereo visual odometry.

## 3. Stereo visual odometry

Before we evaluate the different detectors and descriptors implemented in stereo visual odometry, a general workflow for stereo visual odometry should be introduced. In stereo visual odometry, the vehicle or robot equip a stereo camera system whose intrinsic parameters and rigid relative pose are known, the goal of stereo visual odometry is to precisely estimate the global 6D orientation and position of the system at each stereo frame. The system estimates incrementally; in each new frame, it does the following processes:

(1) Performing epipolar rectification for each stereo image.
(2) Extracting features from each new frame in the left image.
(3) Obtaining corresponding feature points in the right image by standard stereo method, and then calculating the 3D location relative to the present frame.
(4) Extracted features in the left images are matched with features in previous left image.
(5) From these uncertain matches, we exclude outliers and obtain a consensus estimation of motion using RANSAC and PnP motion estimation method on three points [28].
(6) Refinement via incremental sparse bundle adjustment is used to polish the solution.

### 3.1. Motion estimation between frames and outlier rejection

The relative poses (the rotation matrix $R$ and the translation $t$) between the consecutive frames, as shown in Fig. 1, can be calculated using Absolute Orientation (AO) methods or Perspective-n-Point (PnP) methods [31]. In this paper, we exclude outliers and obtain a consensus estimation of motion using RANSAC and three points PnP motion estimation method [32], which has been proved the most accurate PnP motion estimation

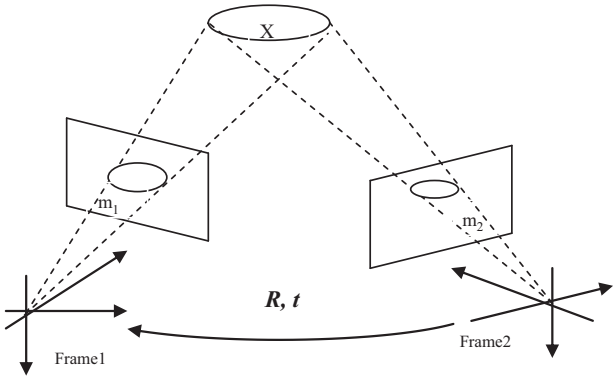


**Fig. 1.** Rotation and translation between the consecutive frames.

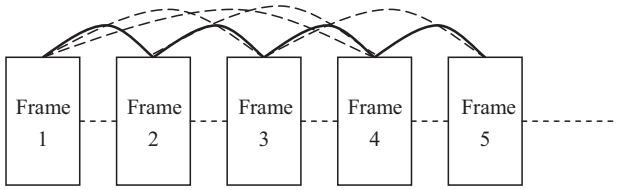


**Fig. 2.** Matching chains and sequence motion refinement.

method [31]. The RANSAC samples three points repeatedly to estimate the motion model and exclude outliers according to that model.

### 3.2. Sequence motion refinement

During the motion estimation from consecutive frame pairs with matched features, matched features are linked in multiple frames through consecutive pairs of images. Fig. 2 shows the process. The linked feature points are corresponded to the same point in real world, thus the sparse bundle adjustment [33] can process the information from several successive stereo images. The matching chains along the images can be seen as constraints that have to be regarded in the estimation. Then the sparse bundle adjustment can achieve more accurate results compared with two-frame estimation.

Each camera projects $X_j$ to $x_{ij}=P_iX_j$, $x_{ij}$ is the image homograph coordinate of the $j$th world point in the $i$th image. In order to obtain best projection matrices $P_i$ and world coordinates $X_j$, we minimize the summed squared re-projection error

$$\min_{P_i,X_j} = \sum_{i,j} d(P_iX_j,x_{ij})$$

The position and orientation of each camera are contained in the projection matrix$P_i$. To refine the pose estimates and achieve real-time performance, we use a local (or incremental) sparse bundle adjustment approach [31,34–37].

## 4. Interest point detectors

In this section, we will give a brief introduction on the state-of-the-art detectors, which have been implemented in visual odometry system, e.g. Harris in [1,31,38], SIFT in [16,39], SURF in [40], FAST in [6,11], CenSurE in [8]. And there are also the newest detectors presented recently, which may be used in visual odometry, e.g. AGAST and BRISK.

### 4.1. Harris corner detector

The Harris corner detection algorithm, proposed by Harris and Stephens [6], is probably one of the most popular corner extraction methods, and is also firstly used in stereo visual odometry. The Harris detector is based on the second moment matrix which is often used for feature detection and for describing local image structures. This matrix describes the gradient distribution in a local neighborhood of a point

$$M = \sum_u \sum_v W(p,q) \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix}$$

$W$ denotes the Gaussian weight filter. Harris then defines the corner response to be

$$C = \mathrm{Det(M)} - k\cdot \mathrm{Tr(M)}^2$$

with Det(M) the determinant and Tr(M) the trace of the matrix M. A typical value for $k$ is 0.04.

When used as an interest point detector, local maxima of the cornerness function are extracted using non-maximum suppression.

### 4.2. Difference of Gaussians (DoG, SIFT-detector)

Lowe proposed a Scale Invariant Feature Transform (SIFT) detector/descriptor scheme [14,15]. The descriptor will be introduced in Section 5.3. SIFT Interest points that are invariant to scale and orientation are extracted with a Difference of Gaussian (DOG) detector. The scale-space of an image $L(x,y,\sigma)$ is produced from the convolution of the input image with a variable-scale Gaussian. The differences of Gaussian is calculated as follows:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y\sigma)$$

Sample points are compared with its eight neighbors in current image and nine neighbors in the adjacent scales and selected as a potential interest point if it is larger or smaller than all of these neighbors. The interest point locations are then refined to sub-pixel accuracy using the quadratic Taylor expansion of the DoG scale-space function.

### 4.3. Fast Hessian (SURF-detector)

The Fast-Hessian detector is used in the SURF detector-descriptor scheme proposed by Bay et al. [23]. The SURF descriptor will be introduced in Sect. 5.4. The Hessian matrix at scale $\sigma$ is defined as follows:

$$H(x,y,\sigma) = \begin{bmatrix} I_{xx}(x,y,\sigma) & I_{xy}(x,y,\sigma) \\ I_{xy}(x,y,\sigma) & I_{yy}(x,y,\sigma) \end{bmatrix}$$

with $I_{xx}$ etc. second order Gaussian smoothed image derivatives. Bay et al. used simple box filters, as shown in Fig. 3, to approximate convolution with the Gaussian second order

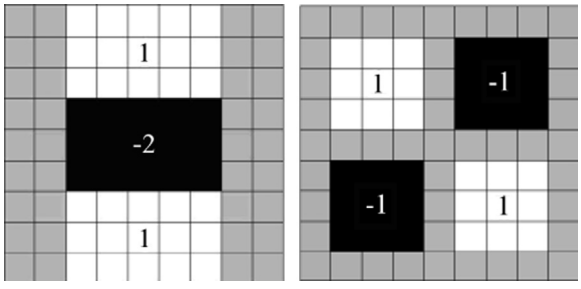


**Fig. 3.** Box filters as used by Fast Hessian as approximations to second order derivatives of Gaussians.

derivatives. So box filters can be computed in constant time using the integral image. The approximate determinant of Hessian matrix then is

$$\mathrm{Det}[H(x,y,\sigma)]\approx D_{xx}D_{yy}-(0.912D_{xy})^2$$

Here $D_{xx}$, $D_{xy}$ and $D_{yy}$ are the results of the image convoluted with box filters. Finally, $3x3\times 3$-neighborhood non-maximum suppression and sub-pixel refinement are then applied.

### 4.4. Center-surround extrema (censure)

CenSurE (Center Surround Extremas) [25] approximates the Laplacian with the centered surround filters which are designed in the forms of octagons, hexagons or boxes. We use the star-version which is also referred to as the Star Keypoint Detector in OpenCV. The filters are shown in Fig. 4. Agrawal et al. propose the use of slanted integral images for octagons, hexagons and star. All the filters can be computed rapidly with integral images. The center-surround filter is computed at all locations and scales and then local extrema in a neighborhood can be found.

### 4.5. Features from accelerated segment test (FAST)

The FAST (Features from Accelerated Segment Test) feature detector, proposed by Rosten and Drummond [10], is very fast to compute. So FAST criterion for interest point detection has become increasingly popular in state-of-the-art methods with hard real-time constraints. A feature is detected at pixel p if the intensities of at least 9 contiguous pixel of a surrounding circle of 16 pixels are all below or above the intensity of p by a threshold t, as shown in Fig. 5. The algorithm was further accelerated by training a decision tree to test as few pixels as possible for classifying a candidate pixel as corner or non-corner. Since the segment test does not compute a corner response function, non-maximal suppression cannot be applied directly to the resulting features. Consequently, a score function, $V$ must be computed for each detected corner, and non-maximal suppression applied to this to remove corners which have an adjacent corner with higher $V$. $V$ is given by:

$$V=\max\Big(\sum_{x\in S_{bright}}\left|I_{p\to x}-I_p\right|-t,\ \sum_{x\in S_{dark}}\left|I_p-I_{p\to x}\right|-t\Big)$$

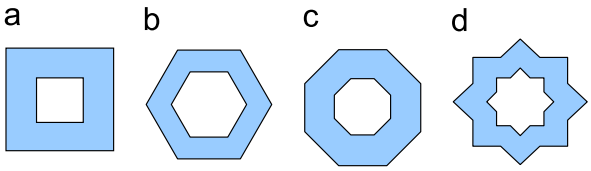


**Fig. 4.** CenSurE's Bi-Level filters.

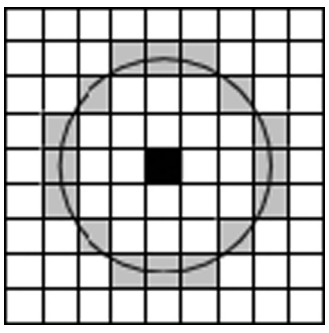

**Fig. 5.** FAST feature detection: the highlighted squares are the pixels used in the feature extraction.

### 4.6. AGAST (Adaptive and generic accelerated segment test)

AGAST was proposed by Mair et al. [11] is a highly efficient corner detector based on the same corner test as FAST. AGAST approach only the way the decision trees for the accelerated segment test are built and used has been significantly improved. AGAST uses also the same non-maximum suppression as FAST.

Unlike FAST, AGAST detector does not have to be trained for a specific scene, but it dynamically adapts to an arbitrary scene which makes the accelerated segment test generic.

FAST computes the decision tree by learning the distribution of the corner configuration from a training set of a specific environment, the result can be quite suboptimal and some corner configurations may be missing in the training set which leads to false positive and false negative responses of the corner detector. In AGAST, the optimal trees are found by exploiting the full binary configuration space. The tree is optimal for a certain probability of similar pixels in the accelerated segment test mask.

AGAST increase the performance of the accelerated segment test and is the currently most efficient corner detection algorithm to our knowledge.

### 4.7. BRISK (Binary robust invariant scalable keypoints)

The BRISK was proposed by Stefan. et al. [13], and able to achieve invariance to scale by searching for maxima not only in the image plane, but also in scale-space using image scale filters, AGAST detecting strategy, and the FAST scores as a measure for saliency. The BRISK estimates the true scale of each interest point in a continuous scale-space. In order to evaluate effect of multi-scales on visual odometry performance, we used especially BRISK (0) detector which the value of octaves parameter is 0 in experiment evaluation. Table 1 shows summarizes of the detectors.

## 5. Local descriptors

We will also introduce the state-of-art descriptors implemented in stereo visual odometry systems, e.g. SAD in [12,41], NCC in [1,31,38], SURF in [40] and SIFT in [16,39]. Although the descriptor of BRISK is not implemented in visual odometry yet, we will employ it as a potential optimal descriptor in our evaluation due to its impressive performance in both the accuracy and cost of calculation.

### 5.1. SAD

In Sum-of-absolute-differences (SAD) approach, the corresponding pixels from two image patches $I_1$and $I_2$ with the same size M by N are subtracted pair wise and the absolute difference of their grey values is summed up

$$SAD=\sum_{m=1}^{M}\sum_{n=1}^{N}abs(I_1(m,n)-I_2(m,n))$$

### 5.2. NCC

NCC (normalized cross-correlation) is more complicated than SAD, but is supposed to offer more robustness to lighting and contrast changes. For two image patches $I_1$and $I_2$ with the same

**Table 1**
Summarizes of the detectors.

| | Harris | FAST | AGAST | BRISK(0) | BRISK | SIFT | SURF | CenSurE |
|---|---|---|---|---|---|---|---|---|
| Scale invariance | No | No | No | No | Yes | Yes | Yes | Yes |
| Feature type | Corner | Corner | Corner | Corner | Corner | Blob | Blob | Blob |

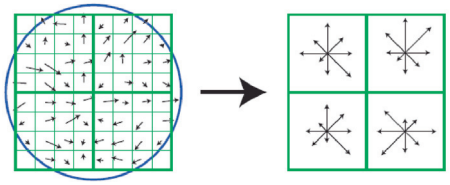


**Fig. 6.** A $2 \times 2$ descriptor array computed from an $8 \times 8$ set of samples.

size M byN, NCC can be expressed as

$$NCC = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} I_1(m,n)I_2(m,n)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N} I_1^2(m,n)\sum_{m=1}^{M}\sum_{n=1}^{N} I_2^2(m,n)}}.$$

### 5.3. SIFT (Scale invariant feature transform)

SIFT descriptor [14,15] is a 3D histogram of gradient location and orientation. The magnitudes are weighted by a Gaussian window with$\sigma$equal to one half the width of the descriptor window. These samples are then accumulated into orientation histograms (with eight bins) summarizing the contents over $4 \times 4$ sub-regions. The feature vector contains the values of all orientation histograms entries. With a descriptor window size of $16 \times 16$ samples leading to 16 sub-regions the resulting feature vector has $16 \times 8 = 128$ elements. Fig. 6 shows the computation of the interest point descriptor.

### 5.4. Speeded up robust features (SURF)

The SURF descriptor [23] encodes the distribution of pixel intensities in the neighborhood of the detected feature at the corresponding scale. To extract the SURF-Descriptor, the first step is to construct a square window of size $20\sigma$ ($\sigma$ is scale) around the interest point oriented along the dominant direction. The window is divided into $4 \times 4$ regular sub-regions. Then for each sub-region the values of $\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|$ are computed, where $d_x$ and $d_y$ refer to the Haar wavelet responses in horizontal and vertical directions in relation to the dominant orientation. This leads to an overall vector of length $4x4 \times 4 = 64$. If the rotational invariance is not required, the upright version of the descriptor, called U-SURF, can be used.

### 5.5. BRIEF (Binary robust independent elementary features)

Calonder et al. proposed BRIEF descriptor [26], which used binary strings as feature point descriptor. It is highly discriminative even when only using relatively few bits and can be computed using simple intensity difference tests. Furthermore, the descriptor similarity can be evaluated using the Hamming distance, which is very efficient to compute, instead of the L2 norm distance as is usually done. So, BRIEF is very fast both to build and match.

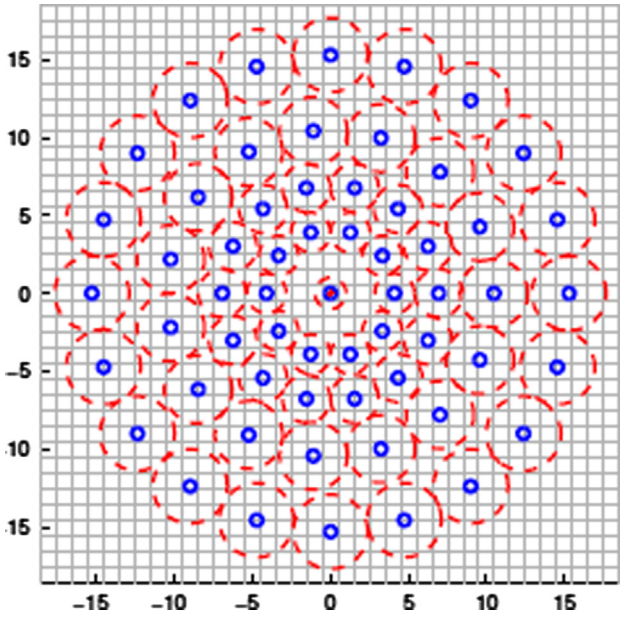


**Fig. 7.** BRISK sampling pattern with $N=60$ points.

Consider a smoothed image patch, p. A binary test $\tau$ is defined by

$$\tau(p;x,y) = \begin{cases} 1 & \text{if } p(x) < p(y) \\ 0 & \text{otherwise} \end{cases}$$

$p(x)$ is the intensity of p at a point $x$. Choosing a set of $n_d(x, y)$ location pairs uniquely defines a set of binary tests. BRIEF descriptor is a $n_d$dimensional bitstring

$$f_{n_d}(p) = \sum_{1 \le i \le n_d} 2^{i-1}\tau(p;x_i,y_i)$$

There are multiple distributions of tests considered in [26], here we use one of the best performers, a Gaussian distribution around the center of the patch.

### 5.6. BRISK descriptor

The BRISK descriptor [13] is also composed as a binary string by concatenating the results of simple brightness comparison tests like BRIEF. The key concept of the BRISK descriptor makes use of a pattern used for sampling the neighborhood of the interest point. BRISK carefully select the brightness comparisons with the focus on maximizing descriptiveness. The pattern, illustrated in Fig. 7, defines $N$ locations equally spaced on circles concentric with the interest point.

For the formation of the rotation- and scale-normalized descriptor, BRISK applies the sampling pattern rotated by $\theta$ around the interest point k. $\theta$ is the orientation of interest point. The

**Table 2**
Summarizes of descriptors.

| | SAD | NCC | SIFT | SURF | U-SURF | BRIEF | BRISK | U-BRISK |
|---|---|---|---|---|---|---|---|---|
| Rotation invariance | No | No | Yes | Yes | No | No | Yes | No |
| Descriptor type | Image patch pixels | Image patch pixels | Gradient histogram | Gradient histogram | Gradient histogram | Brightness comparison | Brightness comparison | Brightness comparison |

bit-vector descriptor is assembled by performing intensity comparisons of point pairs. If the rotational invariance is not required, the upright version of the descriptor, called U-BRISK, can be used.

Table 2 shows summarizes of the descriptors.

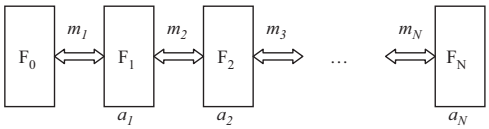


**Fig. 8.** Chain for the detected feature points and matched points with N frames

## 6. Evaluation and experiments

### 6.1. Evaluation criterion

In visual odometry system, the stereo cameras may only detect few corresponds or even some wrong corresponds and will lead to failure in motion estimation due to the quick movement of robots or vehicles. In general, those feature points, which can satisfy the homography projection constrained by the correct estimation of translation and rotation, is called inliers. An intuition evaluation criterion for the performance of stereo visual odometry is the *percentage of inliers in all the detected feature points and matched points*, which can be used to evaluate the performances of detectors and descriptors, especially when there are notable changes on the scales, viewpoints and image blurring or distortion caused by the fast moving vehicles.

As the detector and descriptor are different phrases in stereo visual odometry, they should be considered independently, we then present their criterions on percentage of inliers as follow:

We define percentage of detected inliers as

$$\mathrm{PD} = \frac{\text{the numbers of detected inliers}}{\text{the numbers of detected feature points}}$$

In experimental, it is easy to fail to estimate the motion when the *percentage of detected inliers* is lower than 5%, so we regard it as *detection failure* once the percentage of detected inliers is lower than 5%. Then the *rate of detection failure* will indicates the robustness of the feature detection methods directly in stereo visual odometry.

We define percentage of matched inliers as

$$\mathrm{PM} = \frac{\text{the numbers of matched inliers}}{\text{the numbers of matched points}}$$

According to experimental analysis, the *feature matching failure* is defined as the *percentage of matched inliers* is lower than 20%, which may often fail to estimate motions. Then the *rate of matching failure* will indicates the robustness of the feature matching methods directly in stereo visual odometry.

In visual odometry, the corresponding feature points in every two neighbored frames are continuous associated to form a chain, shown in Fig. 8. $F_i$ is the *ith* frame of the image sequence, $a_i$ is the set of detected feature points in frame $i$, $m_i$ is the set of matched feature points between frame $i$ and frame $i+1$, $b_i$ is the set of corresponding detected inliers between frame $i$ and frame $i+1$, as the inliers contained in the set of detected feature points may not be correctly matched into the matched inliers set, we use $d_i$ to present the set of corresponding matched inliers between frame $i$ and frame $i+1$.

The feature corresponds along that chain may be the most important parameters in visual odometry, the performances of motion estimation in visual odometry are almost determined by the quality of that chain. So we present three new criterions based on the chain.

(1) *Detection Chain Repeatability (DCR)*

$$\mathrm{DCR} = \frac{\sum_{i=1}^{N} C(b_i)}{\sum_{i=1}^{N} C(a_i)} \cdot \frac{\sum_{i=1}^{N} C(b_i)}{N \cdot C(W(b_1 \cup b_2 \cup \ldots \cup b_N))}$$

Here, $N$ is the number of frames, $C(b_i)$ will return the number of points in set $b_i$. $W(Q)$ is a function that can return all the real world points corresponding to each point in $Q$, that is, $W(Q) = \{\cup p | \forall q \in Q, q = P(p)\}$, $P(m)$ is a projection function, which projects the point of real world, $m$, into the image coordinate with the homography matrix.

The *DCR* is a parameter less than 1, and a larger *DCR* will represent that there are more points are correctly detected in that chain, and after executing the sparse bundle adjustment, it will achieve more accurate estimation results in visual odometry.

(2) *Average Detection Chain Re-projection Error (ADCRE)*

$$ADCRE = \frac{\sum ||P_{R,t}(W(b_i)) - b_i||}{\sum_{i=1}^{N} C(b_i)}$$

The ADCRE calculates the average re-projection error for each detected feature points to evaluate the detection accuracy in visual odometry.

(3) *Matching Chain Precision (MCP)*

$$\mathrm{MCP} = \frac{\sum_{i=1}^{N} C(d_i)}{\sum_{i=1}^{N} C(m_i)} \cdot \frac{\sum_{i=1}^{N} C(d_i)}{N \cdot C(W(d_1 \cup d_2 \cup \ldots \cup d_N))}$$

Similarly, a MCP could be calculated to see how matched features are correct inliers.

During the evaluation of detectors in stereo visual odometry system, we adopt a two-phrase method to minimize the affect of descriptors:

(1) Estimation the motions with method introduced in Section 3. To achieve accurate motions, we iterate the RANSAC as much as possible and optimize the $R$, $t$ by the sparse bundle adjustment;

(2) With the accurate geometric constraints between two pairs of stereo images, we test each feature point-pairs whether they

can fulfill the geometric constraints, the inliers will satisfy the following equation:

$|m-P(RX+t)| < \tau, m = (u,v)^T$, $\tau$ is a non-negative threshold.

During the evaluation of descriptors, we implement a hybrid detector method to obtain the original feature points for matching, that is to use three high performance detectors, i.e. SIFT, BRISK0, CenSurE to detect the original feature points and extract 1/3 of the best detected features from each detector to construct our hybrid based detected feature points. We use the nearest-neighbor matching to search the matched descriptors, and the calculation for matched inliers is similar with the detected inliers, it also employs the two-phrase method, the only different is only those matched feature points are used to verify their geometric constraints.

The error of localization must be the most useful criterion in motion estimation, so we also evaluate the relationships of different detectors and descriptors with the localization error in stereo visual odometry. Localization error measures the error between estimated position in stereo visual odometry and true position. It is a comprehensive criterion to represent performances of detectors and descriptors, because accuracies of feature detection and descriptor matching, the number of inliers will all affect the localization error in stereo visual odometry.

The sample iterations in RANSAC may be another interested criterion to evaluate the performance of descriptors. The iteration number of RANSAC indicates how easily RANSAC is able to obtain accurate initial motion estimation and thus it is a direct measure of time spent for outlier removal.

The detector and descriptors are two essential steps in stereo visual odometry, to evaluate the detectors and descriptors fairly and ignore the bias, we use the descriptor with highest MCP when evaluating different detectors with respect to their localization errors. When concerning the localization performances of different descriptors, we choose the detector with highest DCR and lowest ADCRE to minimize the affect of detectors.

### 6.2. Experimental data sets

As the stereo visual odometry is mostly implemented in field robot systems, it is important to evaluate the detectors and descriptors on data sets gathered from real-world robot platforms. In our evaluation, three different data sets of stereo frames are adopted; these data sets are widely used in the benchmark test of stereo visual odometry algorithms:

(1) The Oxford New College dataset [42], kindly provided to the vision research community by the Oxford Mobile Robotics


**Fig. 9.** Large image deformations due to the motion of vehicle.

**Table 3**
Rate of detection failure in large image deformation conditions.

| | Scale changes | Viewpoint changes | Image blurring |
|---|---|---|---|
| Harris | 7% | 6% | 14% |
| FAST | 5% | 4% | 15% |
| CenSurE | 1% | 1% | 5% |
| AGAST | 4% | 3% | 11% |
| SIFT | 1% | 0 | 2% |
| SURF | 1% | 0 | 0 |
| BRISK | 1% | 2% | 3% |
| BRISK(0) | 3% | 2% | 9% |

**Table 4**
Rate of matching failure in large image deformation conditions.

| | Scale changes | Viewpoint changes | Image blurring |
|---|---|---|---|
| NCC | 65% | 55% | 89% |
| SAD | 63% | 52% | 91% |
| SURF | 0 | 0 | 1% |
| U-SURF | 0 | 0 | 0 |
| SIFT | 1% | 1% | 4% |
| BRISK | 3% | 2% | 3% |
| U-BRISK | 2% | 3% | 2% |
| BRIEF | 5% | 2% | 3% |

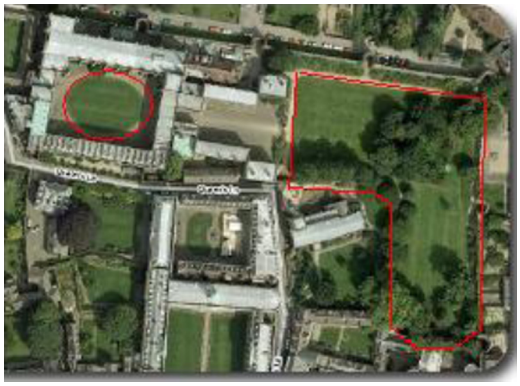

**Fig. 10.** Oxford New College dataset 1 and dataset 2: aerial view of location.

Group. This dataset comprises a total of 52,478 stereo frame pairs, collected in about 47 min over a total path length of about 2844 m. The accurate ground truth is not available.

(2) The Amsterdam Hague dataset [39], kindly provided to the computer vision research community by the Intelligent Autonomous Systems Group. This dataset comprises Hague1 and Hague2. All data sets are accompanied by DGPS. Hague1 dataset encompass a loop of approx. 600 m, and contain around 2500 stereo images. Hague2 dataset encompass a loop of approx. 800 m, and contain around 3000 stereo images.

(3) Karlsruhe dataset [43], kindly provided to Andreas Geiger professor by Institute of Measurement and Control Systems. The dataset contains high-quality stereo sequences recorded from a moving vehicle in Karlsruhe. The dataset encompass 16 sub-datasets. The ground truth odometry from an OXTS RT 3000 GPS/IMU system is provided.

### 6.3. Robustness evaluation

As mentioned in previous, there may be three challenges in feature detection and matching of visual odometry, i.e. significant changes on scales, viewpoints and blurring images. So we extract a couple of stereo frame-pairs from data sets introduced in Section 6.2, those pairs[1] are all considered to contain the either the significant changes on scales, changes on viewpoints or image blurring, as shown in Fig. 9. We then apply the detectors and descriptors in those selected pairs and calculate the *rate of detection failure* and *rate of matching failure* respectively as follows:

$$rate\ of\ \text{detection}\ failure = \frac{N_{PD<5\%}}{N_{total}},\quad rate\ of\ matching\ failure = \frac{N_{PM<20\%}}{N_{total}}$$

here $N_{PD<5\%}$ is the number of pairs whose *percentage of detection inliers (PD)* is less than 5%; $N_{PM<20\%}$ is the number of pairs whose *percentage of matching inliers (PM)* is less than 20%; $N_{total}$is the number of total pairs. The experimental results of detectors and descriptors are shown in Table 3 and Table 4 respectively.

Table 1 shows the detection failure rate of different feature detector. From the Table 1, we can see that Harris performs the worst. Performances of SIFT, SURF, CensurE, and BRISK feature are similar and SURF is slightly better. BRISK(0) is worse than BRISK. Image blurring may be the mainly negative factor with respect to the robustness of feature detectors.

From Table 4, we can see that NCC and SAD are worst with significant poor performance on robustness. Performances of SURF, U-SURF, SIFT, BRIEF, BRISK and U-BRISK are comparative and are very high. The reason may be that SAD and NCC employ the patch of pixels in computation, which will be quite sensitive to the changes of scale, viewpoint and image blurring and distortions and failed to match.

**Table 5**
Detection accuracy of different feature detectors in stereo visual odometry.

| | DCR | ADCRE | Closure loop error (m) |
|---|---|---|---|
| Harris | 0.41 | 0.42 | 1.19 |
| FAST | 0.41 | 0.39 | 0.74 |
| CenSurE | 0.43 | 0.40 | 0.72 |
| AGAST | 0.42 | 0.40 | 0.71 |
| SIFT | 0.39 | 0.36 | 0.79 |
| SURF | 0.39 | 0.39 | 0.88 |
| BRISK | 0.34 | 0.35 | 0.65 |
| BRISK(0) | 0.47 | 0.35 | 0.53 |

**Table 6**
Matching accuracy of different descriptors for stereo visual odometry.

| | MCP | Iteration numbers of RANSAC[a] | Closure loop error (m) |
|---|---|---|---|
| NCC | 0.27 | 313 | 1.55 |
| SAD | 0.28 | 315 | 1.51 |
| SURF | 0.56 | 92 | 0.79 |
| U-SURF | 0.58 | 85 | 0.72 |
| SIFT | 0.63 | 78 | 0.67 |
| BRISK | 0.64 | 71 | 0.59 |
| U-BRISK | 0.66 | 70 | 0.53 |
| BRIEF | 0.65 | 71 | 0.55 |

[a] The terminal condition for the iteration of RANSAC is when there are no more than 1% outliers removed.

[1] To evaluate the performance on challenge conditions, we did not extract continuous image pairs, some middle frames among the pairs are skipped to obtain the pairs with significant changes on scales and viewpoints.

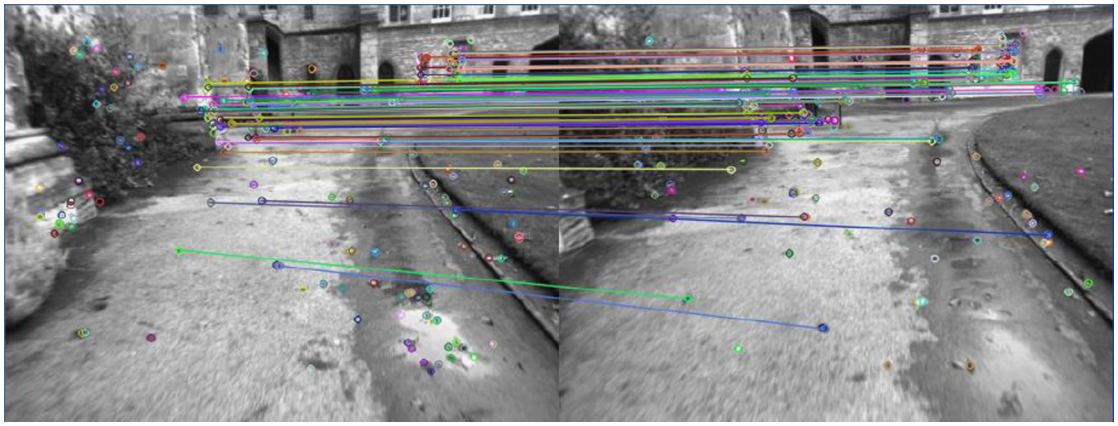

**Fig. 11.** BRISK (0) features tracked over several frames.

### 6.4. Precision evaluation

We also use the *DCR*, *ADCRE* and *MCP* to evaluate the performances of different detectors and descriptors. In experiments, the Oxford New College data sets are used, we also use the closure loop error to evaluation the accuracies of localization with different detectors and descriptors. There are two closure loops in the Oxford New College data sets, shown in Fig. 10. The closure loop error is calculated as follows: We assume the initial position of the vehicle is zero, and then estimate the motion of vehicle with the tested detector or descriptor until it returns to the end, which should be the same with the initial position in a closure loop motion. And then we can obtain the error of corresponding detector or descriptor from the bias of end.

In general, a larger *DCR* will represent that there are more available feature points used in the estimation and will lead to more accurate results in motion, a smaller *ADCRE* will represent that the projects of inliers will be more close to the ground true ones, and the detected features are more accurate. So larger *DCR* and smaller *ADCRE* will lead to less error in localization error (or closure loop error).

The experimental results of Table 5 show that the BRISK (0) could achieve the best DCR and lowest ADCRE, and the closure loop error is also lowest. The closure loop error of Harris is worst. We also find that the performance of BRISK (0) is better than BRISK, the reason may be that the octaves of brisk feature are formed by progressively half-sampling the original image, while half-sampling will decrease the accuracies of features in motion estimation. Fig. 11 shows the BRISK (0) features tracked over several frames.

The experimental results on different descriptors are shown in Table 6. The results show that the performances of brightness comparison based descriptors may be superior to others, the performance gradient histogram based descriptors is less than brightness comparison based descriptors and better the image patch pixels based descriptors. The results of Table 6 may be determined by the characteristics of stereo visual odometry: there are small deformations among two continuous frames, the gradient histogram based descriptors may suffer with weak distinctiveness, but it is strong at anti-deformation; Image patch pixels based descriptors is good at distinctiveness, but the performance on anti-deformation is weak; And the brightness comparison based descriptors are good compromises of the anti-deformation and distinctiveness, thus they are more proper for the cases of visual odometry.

The results on Table 6 also show that performances of descriptors with rotation invariance are less than the descriptors with rotation variance; the reason may be that there are small in-plane rotations in visual odometry cases, thus the bias on rotation introduced by calculating the rotation invariance in descriptors may have more influence on the distinctiveness.

Table 6 also shows the iteration number of RANSAC for different descriptors. SAD and NCC need the largest iteration, it shows the SAD and NCC based stereo visual odometries spend much time on removing outliers.

As the above experimental results suggest the BRISK (0) and U-BRISK may be the best detector and descriptor respectively, we also carried out an experiment to compare the trajectory of ground truth and the results implemented by BRISK (0) and U-BRISK. Experimental results are shown in Fig. 12. The ground truth localization information of Oxford New College data sets is not available, so we only present the closure loop error shown in Fig. 12(a), (b); Fig. 12(c), (d) show the results on Karlsruhe dataset, which can provide ground truth localization information. The results show the approach with BRISK (0) and U-BRISK may achieve impressive accuracy in localization.

### 6.5. Executing time evaluation

As the visual odometry is a real-time application, we also compare the executing time for different detectors and descriptors using the same hardware and software platform[2] . In this experiment, all the detectors and descriptors are implemented to process the same video sequences with about 100 frames, the resolution of image is $512 \times 384$ and the maximal number of features[3] is 550. We record the average executing time of each detector and descriptor in one frame. The results are shown in Tables 7 and 8.

The experimental results show SIFT detector spends much more time in computation than other detectors, and the computation time of BRISK (0) is approximate to AGAST, which is the faster detector in experiments.

According to Table 8, the SIFT descriptor spends much more time in computation than other descriptors, and the calculation costs of SAD, NCC, U-BRISK, BRISK and BRIEF are approximate. The experimental results suggest the BRISK detectors and descriptors may be a proper solution for stereo visual odometry when considering the robustness, accuracy and executing time in all.

[2] We use a laptop with Intel Core2 Duo, T6600, 2.2 GHz CPU, 2 G memories, and all the detectors and descriptors are implemented with C++.

[3] In stereo visual odometry system, it often sets a maximal number of features in feature detection and matching to make a stable computation in each frame.

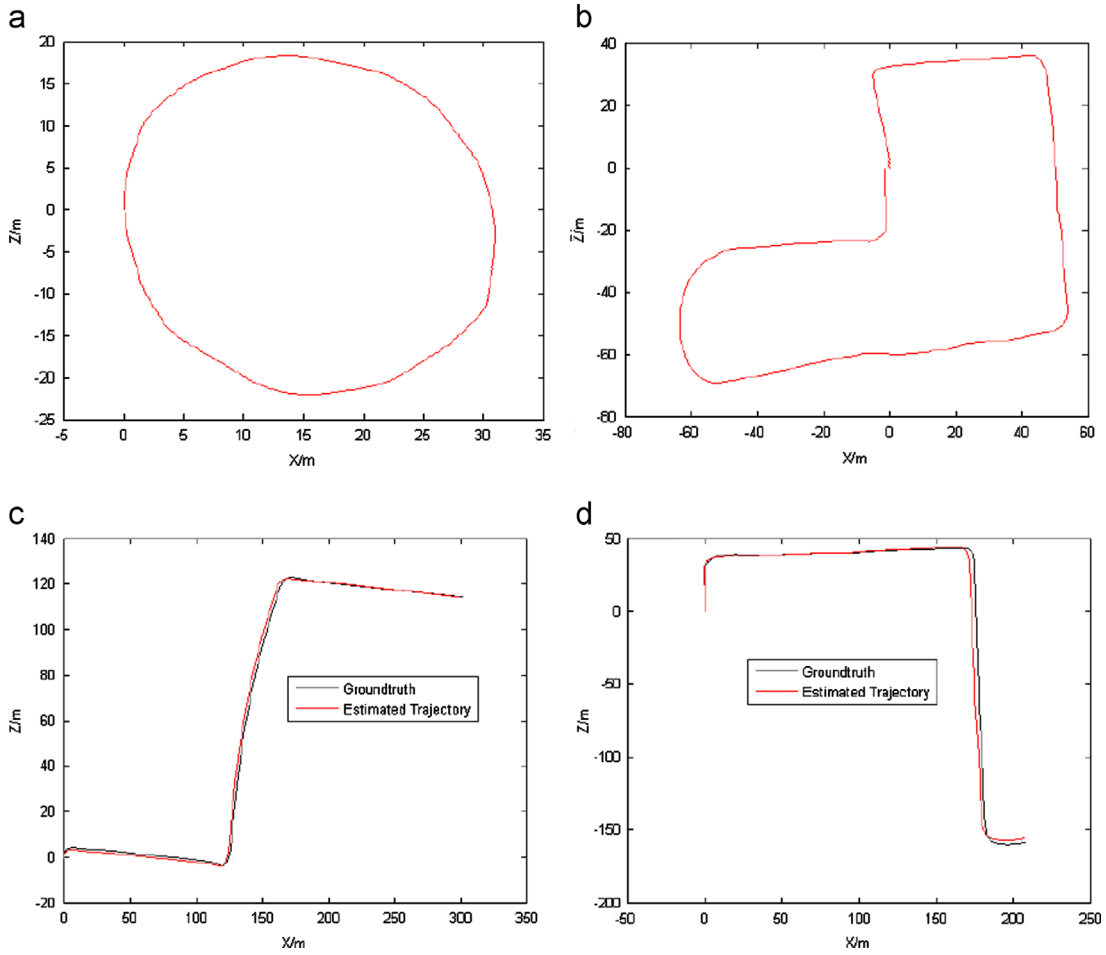


**Fig. 12.** Localization results of different dataset with BRISK (0) detector and U-BRISK descriptor. (a) Oxford New College dataset 1, (b) Oxford New College dataset 2, (c) Karlsruhe dataset 1 and (d) Karlsruhe dataset 2.

**Table 7**
Average run-time of different feature detectors in one frame.

| | Harris | FAST | AGAST | CenSurE | SIFT | SURF | BRISK | BRISK(0) |
|---|---|---|---|---|---|---|---|---|
| Time (ms) | 15 | 8 | 5 | 22 | 1640 | 197 | 21 | 6.5 |

**Table 8**
Average run-time of different descriptors computing and matching in one frame.

| | NCC | SAD | SURF | U-SURF | SIFT | BRIEF | BRISK | U-BRISK |
|---|---|---|---|---|---|---|---|---|
| Time (ms) | 10 | 9 | 138 | 93 | 1225 | 26 | 20 | 15 |

## 7. Discussion and conclusion

In brief, the corner based detectors such as FAST may be a superior solution in stereo visual odometry projects, for their low cost on computation. Among those detectors, the BRISK (0) shows an overall good performance. Generally, the detectors with scale invariance may be more stable than sample corner detectors, so they have less failure rate on motion estimation. However, multiple scales based detectors do not achieve better performance when considering the localization accuracy in visual odometry. It may because those scale-space features are not well localized at higher levels in an image pyramid. Obviously, features at high levels have less accuracy relative to the original image. So the lower scale in features may lead to more accurate localization results when considering the kind of features in visual odometry. Besides, multiple scales based detectors need additional computation for those scales. So we may use the policy in stereo visual odometry: the detectors with multiple-scales are adopted only when there are some extremely motions occurring, in normal times, the simple detectors such as corner based approaches are implemented.

The brightness comparison based descriptors are performed well in overall robustness, precision, and computation, thus are proper for implementing in visual odometry. The experimental results also show that the rotation variance descriptors perform better than those rotation invariance ones. Thus the rotation invariance descriptors may be not necessary in stereo visual odometry.

## Acknowledgment

This research is based upon work supported in part by National Natural Science Foundation of China (61173123), and Natural Science Foundation of Zhejiang Province (R13F030003, Y1101237, LY12F02012).

## References

[1] D. Nister, O. Naroditsky, J. Bergen. Visual odometry. in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 652–659, 2004.

[2] K. Mikolajczyk, C. Schmid., A Performance Evaluation of Local Descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[3] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest point detectors, Int. J. Computer Vision 37 (2) (2000) 151–172.
[4] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, A comparison of affine region detectors, Int. J Comput. Vision 65 (2005) 43–72.
[5] P. Moreels, P. Perona, Evaluation of features detectors and descriptors based on 3D objects, Intl. J. Comput. Vision 73 (3) (2007) 263–284.
[6] C. Harris, M. Stephens, A combined corner and edge detector. in: Proceedings of the 4th Alvey Vision Conference, pp. 147–151, 1988.
[7] Yuan Yuan, Yanwei Pang, Xuelong Li, Gabor-based region covariance matrices for face recognition, IEEE Trans. Circuits aSyst. Video Technol. (T-CSVT) 18 (no. 7) (2008) 989–993.
[8] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features. in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511–518, 2001.
[9] Yanwei Pang, Xuelong Li, Yuan Yuan, Dacheng Tao, Jing Pan, Fast haar transform based Feature Extraction for Face Representation and Recognition, IEEE Trans. Inf. Forensic. Secur. (T-IFS 4 (no.3) (2009) 441–450pp 4 (2009) 441–450.
[10] E. Rosten, T. Drummond, Machine learning for highspeed corner detection. in: Proceedings of the European Conference on Computer Vision (ECCV), 2006.
[11] E. Mair, G.D. Hager, D. Burschka, M. Suppa, G. Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. in: Proceedings of the European Conference on Computer Vision (ECCV), 2010.
[12] Gabe Christopher Mei, Mark Sibley, Paul Cummins, Newman and Ian Reid. RSLAM: A System for Large-Scale Mapping in Constant-Time using Stereo, Intl. J. Comput. Vision 94 (2) (2010) 198–214.
[13] L. Stefan, C. Margarita, S. Roland, BRISK: Binary Robust Invariant Scalable Keypoints. in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2011.
[14] D. Lowe, Object recognition from local scale-invariant features. in: Proceedings of the International Conference on Computer Vision, 1999.
[15] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision (IJCV) 60 (2) (2004) 91–110.
[16] S. Se, D.G. Lowe, J. Little Vision-based mobile robot localization and mapping using scale-invariant features. In: Proceedings of IEEE International Conference on Robotics and Automation, 2001.
[17] Yanwei Pang, Mianyou Shang, Jing Pan, Scale invariant image matching using triplewise constraint and weighted voting, Neurocomputing 83 (no.5) (2012) 64–71.
[18] Yanwei Pang, Qiang Hao, Yuan Yuan, Tanji Hu, Rui Cai, Lei Zhang, Summarizing Tourist Destinations by Mining User-Generated Travelogues and Photos, Comput. Vision Image Understanding 115 (no. 3) (2011) 352–363.
[19] Y. Ke, R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. in: Proceedings of the Conference Computer Vision and Pattern Recognition, pp. 511–517, 2004.
[20] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886–893, 2005.
[21] Yanwei Pang, He Yan, Yuan Yuan, Kongqiao Wang, Robust CoHOG Feature Extraction in Human Centered Image/Video Management System, IEEE Trans. Syst., Man, Cybern, Part B: Cybern. 42 (no. 2) (2012) 458–468.
[22] Y. Pang, Y. Yuan, X. Li, J. Pan, Efficient HOG Human Detection, Signal Process. 91 (no.4) (2011) 773–781, April.
[23] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded up robust features. in: Proceedings of the European Conference on Computer Vision (ECCV), 2006.
[24] Wei Yanwei Pang, Yuan Li, Jing Pan. Yuan, Fully affine invariant SURF for image matching, Neurocomputing 85 (May 2012) 6–10.
[25] M. Agrawal, K. Konolige, M.R. Blas, CenSurE: Center surround extremas for realtime feature detection and matching. in: Proceedings of the European Conference on Computer Vision (ECCV), 2008.
[26] M. Calonder, V. Lepetit, C. Strecha, P. Fua, BRIEF: Binary robust independent elementary features. in: Proceedings of the European Conference on Computer Vision (ECCV), 2010.
[27] K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors, Int. J. Comput. Vision 60 (1) (2004) 63–86.
[28] G. Carneiro, A.D. Jepson. Phase-based local features. in: Proceedings of the Seventh European Conference Computer Vision, pp. 282–296, 2002.
[29] T. Randen, J.H. Husoy, Filtering for texture classification: A comparative study, IEEE Trans. Pattern Anal. Mach. Intell. 21 (no. 4) (Apr. 1999) 291–310.
[30] M. Varma, A. Zisserman, Texture classification: Are filter banks necessary?. in: Proceedings of the Conference Computer Vision and Pattern Recognition, pp. 477–484, 2003.
[31] H. Alismail, B. Browning, M.B. Dias, Evaluating pose estimation method for stereo visual odometry on robots. in: Proceedings of the 11th International Conference on Intelligent Autonomous Systems, 2010, pp. 101–110.
[32] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[33] M.I.A. Lourakis, A.A. Argyros, SBA: A software package for generic sparse bundle adjustment, ACM Transactions on Mathematical Software 36 (2009) 1–30.
[34] M. Agrawal, K. Konolige, Rough terrain Visual odometry. in: Proceedings of the International Conference on Advanced Robotics (ICAR), South Korea, 2007.
[35] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, P. Sayd. 3D reconstruction of complex structures with bundle adjustment: an incremental approach[C]. in: Proceedings of IEEE International Conference on Robotics and Automation (ICRA ), 2006, pp. 3055–3061.
[36] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, P. Sayd, Generic and real-time structure from motion using local bundle adjustment, Image Vision Comput. 27 (8) (2009) 1178–1193.
[37] N. Sunderhauf, P. Protzel, Towards using bundle adjustment for robust stereo odometry in outdoor terrain. in: Proceedings of the Towards Autonomous Robotic Systems TAROS06, Guildford, UK, 2006, pp. 206–213.
[38] M. Maimone, Y. Cheng, L. Mattheis, Two years of visual odometry on the mars exploration rovers, J. Field Robotics 24 (3) (2007) 169–186.
[39] G. Dubbelman, F.C.A. Groen, Bias reduction for stereo based motion estimation with applications to large scale visual odometry. in: Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Miami, USA, 2009.
[40] A. Cumani, A. Guiducci, Fast stereo-based visual odometry for rover navigation, WSEAS Trans. Circuits Syst. 7 (7) (2008) 13–17.
[41] A. Howard. Real-time stereo visual odometry for autonomous ground vehicles. in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008, pp. 3946–3952.
[42] ⟨www.robots.ox.ac.uk/NewCollegeData/⟩.
[43] B. Kitt, A. Geiger, H. Lategahn, Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme, IEEE Intelligent Vehicles Symp. (2010) 486–492.

**Xu Yunxi** received the MS degree in communication and information system from Soochow University in 2004 and now is a Lecturer in the School of Information & Engineering at Huzhou Teachers College. He is also currently a Ph.D. candidate in the Department of Information Science & Electronic Engineering at Zhejiang University. His current research interests include visual navigation, real-time video 3D reconstruction, intelligent video surveillance.

**Liu Yong** received his B.S. degree in computer science and engineering from Zhejiang University in 2001, and the Ph.D. degree in computer science from Zhejiang University in 2007. He is currently an associate professor in the institute of Cyber-Systems and Control, Department of Control Science and Engineering, Zhejiang University. He has published more than 30 research papers in machine learning, computer vision, GIS system, intelligent CAD system and granular computing. His latest research interests include machine learning, robotics vision, information processing and granular computing.

**Jiang Yunliang** is a Professor of the School of Information and Engineering of Huzhou Teachers College, China. He received the B.S. degree in Mathematics from Zhejiang Normal University in 1989, the M.E. degree in Computer Science and Technology from Zhejiang University in 1997, and the Ph.D. degree in Computer Science and Technology from Zhejiang University in 2006, respectively. His research interests include geographic information system, artificial intelligence and information fusion. He has published over 20 papers in journals such as Lecture Notes in Artificial Intelligence, Lecture Notes in Computer Science, Pattern Recognition and Artificial Intelligence (in Chinese), and others.