International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-1, January 2016
ISSN: 2395-3470
www.ijseas.com

Survey on Key Feature Descriptors Used In Computer Vision Applications

Pami Prakash[1]
M.Tech Scholar, Department of Computer Science, Sree Budha College of Engg,Kerala

Anil A.R[2]
Head of the Department,
Department of Computer Science,
Sree Budha College of Engg,Kerala.

ABSTRACT

Computer vision is a field that includes methods for acquiring, processing, and understanding images and, high-dimensional data from the real world to produce decisions. Many applications rely on matching key points across images. Nowadays, the deployment of vision algorithms on smart phones and other devices with low memory and computation complexity has a great importance. The goal is to make descriptors which are faster to compute, more compact while remaining invariant to any common image transformations. There are different faster and more robust key points and association algorithms such as Scale Invariant Feature Transform (SIFT), Speed-up Robust Feature (SURF), and more recently Binary Robust Invariant Scalable Key points (BRISK) etc. To address these current requirements, we need to analyze which descriptor is the most robust one.

Keywords: *Computer Vision, feature descriptors, descriptor matching keypoint descriptor*

I. Introduction

Keypoint descriptor matching and recognition are two main concepts in computer vision and image processing. Computer vision has been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception. It is concerned with theory and technology for building artificial systems that obtain information from images or multi-dimensional data. There are many applications of computer vision such as controlling process, navigation, detecting events, organizing information, modelling objects, automatic inspection etc. Moreover, many applications such as object recognition, 3D reconstruction, image retrieval, and camera localization rely on feature extraction. Local features and their descriptors constitutes the building blocks of many computer vision algorithms. Local features refer to a pattern or distinct structure found in an image, such as a point, edge, or small image patch. They are usually associated with an image patch that differs from its immediate surroundings by texture, color, or intensity. They enables these algorithms to better handle scale changes, rotation, and occlusion. Keypoint extraction is a special kind of reducing dimension. Transforming input data to a set of features is called feature extraction. If the extracted features are carefully chosen it is expected that the features which are set will extract the relevant information from the input data so as to perform a desired task by using reduced representation instead of using the entire size of the input. An important area of application is image processing, in which algorithms are used to detect various desired portions or features of a digitized image. Local features provide the ability to find image correspondences regardless of occlusion, changes in viewing conditions, or the presence of clutter.

When given two images of the same scene, most features that the detector finds

International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-1, January 2016
ISSN: 2395-3470
www.ijseas.com

in both images are the same. The features are robust to changes in viewing conditions and noise. The difficulty in extracting features from an image relies on balancing two factors: high quality description and low computational requirements. SIFT and SURF algorithms, which are the well-known leaders in the field possesses great performance under all kind of image transformations. SURF is considered as the most computationally efficient among the high performance methods to date. Object recognition, object matching, and many other vision applications is based on representing images with sparse number of keypoints. Efficiently describing key points, with stable, compact and robust representations invariant to scale, rotation, affine transformation, and noise is a real challenge.

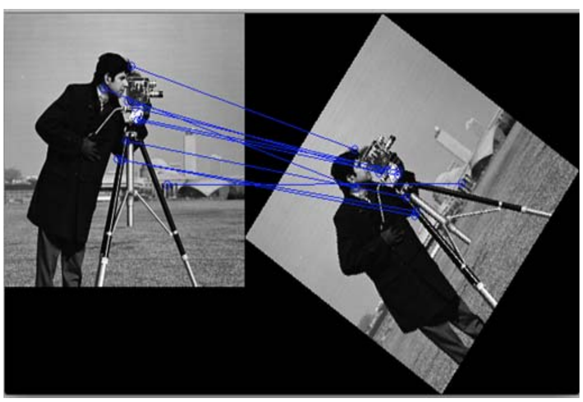

Figure 1: Illustration of feature extraction. A series of features are matched here.

II. Review of Literature

D. Lowe. proposed Scale Invariant Feature Transform[1] which is a widely accepted one with highest quality options currently available, which promises robustness and common image transformation invariance– however, the expense of computational cost is also considered here. Most of the efforts in the last decades was to perform as better as SIFT [1,7] with lower computational complexity and memory load. The SIFT descriptor seems the most ideal descriptor for practical uses, and hence most widely used nowadays. It relies on local gradient histograms. It is robust, repetitive, distinctive and relatively fast, which is most desirable for on-line applications. It is computationally intensive for use in real-time applications of any complexity. Applications such as object recognition, detection, image stitching, gesture recognition, video tracking, face identification etc. uses this descriptor due to their high performance.

Bay et al. [2] proposed Speeded up robust features [2] which is a fast, distinctive, and invariant to scale and rotation interest point detector and descriptor which is built upon the goodness of the leading detectors and descriptors which are widely used in this context. These grid histograms of gradients are concatenated into a 64-dimensional vector. The high dimensionality makes it difficult to use this in real time applications, so SURF also use a 36-dimensional vector of principle components of the 64 vector for a speedup. SURF also improvises SIFT by using a box filter approximation to the convolution kernel of the Gaussian derivative operator, it relies on local gradient histograms but uses integral images to speed up the computation. This convolution is speeded up further using integral images to reduce the time spent on this step. SURF[9] approximates or even outperforms previously proposed methods with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster. This is achieved by relying on integral images for image convolutions; by building incrementally on the strengths of the leading state-of-the-art detectors and descriptors. These algorithms can also be simplified. This leads to a combination of novel detection, description, and matching steps. This encircles a detailed description of the detector and descriptor. The relevant speed gain is due to the use of integral images, which tremendously reduce the number of operations for simple box convolutions, independent of the chosen scale. The results showed that the performance of Hessian approximation is comparable and often better than the state-of-the art key point detectors. The high repeatability is beneficial for camera self-calibration, where an accurate interest point detection has a direct influence on the accuracy of the camera self-calibration and therefore on the quality of the resulting 3D model.

Calonder et al. proposed to use binary strings as an efficient feature point descriptor, coined BRIEF which is highly discriminative and distinctive even while using relatively few bits and can be computed using simple brightness comparison tests. Moreover, the descriptor similarity can be evaluated by using the Hamming distance, which is highly efficient to compute. Hence, BRIEF is faster both to build and to match. It yields a similar or better recognition performance, when compared against SURF and U-SURF on standard benchmarks. The image patches could be efficiently classified on the basis of a relatively small number of pairwise intensity comparisons to represent the image patch as a binary string. It confirms the validity of the recent trend that involves moving from the Euclidean distance to the Hamming distance for the matching purposes. They are designed to test robustness to viewpoint changes compression artifacts illumination changes and image blur. It also provides higher recognition rates, as long as invariance to large in-plane rotations is not at all a requirement.

Leutenegger et al. proposed Binary robust independent elementary features [4] which is a 512 bit binary descriptor in which weighted Gaussian average over a selected pattern of points near the keypoint is computed. BRISK is a novel method used in classic Computer Vision problems for detecting, describing and matching keypoints of the given image, when there is no sufficient prior knowledge on the scene and camera poses. It uses a deterministic sampling pattern [8] that results in a uniform sampling-point density in a given radius around the keypoint which is one of the key differences of this novel method. BRISK uses dramatically lesser number of sampling-points than pairwise comparisons. Only a single point participates in more comparisons which in turn limits the complexity of looking around for intensity values. The result obtained is a bit-string of length 512, therefore the descriptor matching will be performed fast. In contrast to many robust high-performance state-of-art algorithms, such as SIFT and SURF, this method offers a dramatically faster alternative with comparable matching performance. BRISK depends on a circular sampling pattern from which it computes the brightness comparisons to get a binary descriptor string. Hence this unique property of BRISK is useful for a wide range of hard real-time applications, with limited computation power. BRISK provides its excellent features in such time-demanding applications, which are not fault-tolerant. Amongst avenues for further research into BRISK which explore alternatives to yield higher repeatability while maintaining speed. Furthermore, BRISK pattern is analyzed both theoretically and experimentally such that it provides robustness to all image transformations of the descriptor is maximized. Given a set of keypoints, this descriptor computes a binary string by comparing the brightness of the selected pairs. In BRISK, the characteristic direction of each keypoint is observed to allow for orientation-normalized descriptors so as to achieve robustness by providing rotation transformation invariance. Careful selection of the brightness comparisons focuses on maximizing descriptiveness. In the Dual-Bootstrap [6], keypoints are instead ranked-ordered by this distinctivenessmeasure.

Rosten et al. proposed Machine learning for high speed corner detection [5] which is a corner detection method used to extract feature points and efficiently track and map objects in many computer vision applications. Corner detection is one of the first step in computer vision applications such as object tracking, recognition etc. The most desirable advantage of so called FAST (Features from Accelerated Segment Test) corner detector is its computational efficiency. It is faster than many other state-of-art feature extraction algorithms[10].

III. Conclusion

The survey shows various efficient keypoint extraction techniques. Keypoints provide the ability to find image correspondences regardless of occlusion, changes in viewing conditions, or the presence of clutter. I have looked different methods for efficient keypoint detection, but each descriptor have its own merits and demerits. We can choose our descriptor depending on the current requirements. Here provide a detailed summary of each keypoint detection technique and comparative study.

References

[1] Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. Computer Vision and Image Understanding 20 (2004) 91–110.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. Computer Vision–ECCV 2006, pages 404– 417, 2006. 1, 2

[3] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. Computer Vision– ECCV 2010, pages 778–792, 2010. 1, 2, 3

[4] S. Leutenegger, M. Chli, and R. Siegwart. Brisk: Binary robust invariant scalable keypoints. 2011. 1, 2, 3, 5, 6

[5] E. Rosten and T. Drummond. Machine learning for highspeed corner detection. Computer Vision–ECCV 2006, pages 430–443, 2006. 2

[6] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai. Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence. IEEE Trans. Pattern Anal. Machine Intell., 2007. Accepted for publication. 1, 2, 3, 7

[7] M. Brown and D. Lowe. Recognising panoramas. In Proc. ICCV, 2003. 1, 2, 3, 4

[8] C. Harris and M. Stephens. A combined corner and edge detector. In Proceedings of The Fourth Alvey Vision Conference, pages 147–151, Manchester, UK, 1988.

[9] T. Kadir and M. Brady. Saliency, scale and image description. Int. J. Comp. Vis., 45(2):83–105, 2001. 2

[10] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell., 27(10):1615–1630, 2005. 1