Sperker, Hans-Christian; Henrich, Andreas

Feature-based object recognition : A case study for car model detection

Date of secondary publication: 26.02.2025

Accepted Manuscript (Postprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-1067084

Primary publication

Sperker, Hans-Christian; Henrich, Andreas (2013): Feature-based object recognition : A case study for car model detection, in: László Czúni, Klaus Schöffmann, und Tamás Szirányi (Ed.), 11th International Workshop on Content-Based Multimedia Indexing (CBMI), 2013 ; Proceedings, New York, NY, USA: IEEE, pp. 127–130, doi: 10.1109/CBMI.2013.6576568.

Feature-based Object Recognition – a case study for car model detection

Hans-Christian Sperker
Media Informatics Group
Bamberg University, Germany
hans-christian.sperker@uni-bamberg.de

Andreas Henrich
Media Informatics Group
Bamberg University, Germany
andreas.henrich@uni-bamberg.de

***Abstract*—Recognition of untextured as well as shiny objects introduces problems to most of the present interest point detectors and descriptors. Those problems arise from changes in brightness intensities around potential interest points, caused by reflections on such an object's surface. Since the selection of the best performing recognition algorithm heavily depends on the underlying task, an evaluation of some algorithms tested especially for the respective data collection and typical query types has to be taken into account. If the task is for example the model detection of a car, most of the existing algorithm evaluations are misleading since they use data sets with mostly textured and matt objects. The present paper introduces a data set of car images which stands for untextured, shiny objects and evaluates different algorithms namely SIFT, ASIFT, SURF, ORB, and BRISK on that data set.**

Findings are that most algorithms have a low effectiveness recognizing such objects. This mainly depends on the kind of interest point detector and is not so much dependent on the kind of descriptor. Detectors preferring points on edges and corners are more reliable than those preferring interest points on blob-like structures.

I. Introduction

In recent years local image features have emerged as effective object recognition techniques for many applications in image analysis and retrieval. A confusing variety of approaches have been presented and it remains unclear which approach is best suited for a particular scenario, especially for scenarios a bit aside from the mainstream. For a closer look, the typical three main aspects of the approaches have to be distinguished: *Feature point detection* yielding the most characteristic local areas (interest points) in the image. *Feature point description* defining expressive representations for those points and their direct neighborhood. *Feature point matching* defining a (dis-)similarity measure for descriptors in order to find related interest points in two images.

The choice for the most appropriate interest point detector and descriptor for a specific task strongly depends on the problem at hand [1]. Therefore, a lot of different data sets exist on which the different algorithms may be tested [2], [3]. This should help picking the best algorithm for a given task. However, most of these data sets have in common that depicted objects and scenes are well textured and matt and that query images show exactly the identical objects for the task of recognition, which may be sufficient for many tasks. But if the presented task is for example the identification of an untextured shiny object showing the equal and not the identical object, most data sets and algorithm evaluations fall short.

Such a scenario could be as follows: Given an image of a car seen e.g. on a parking place, a system has to determine the car model. In this scenario, obviously the car from the parking place itself will not be present in the database, but images showing cars of the equal model can be. Further sharpening our scenario, we could assume that the database with example images is made up of properly selected images of cars representing the car models which have to be recognized[1]. Since a data set appropriate for the sketched scenario showing untextured, shiny objects is currently not existing, it is hard to choose the best performing feature-based object recognition algorithm for a task like car model detection.

In this work such a data set is introduced and an evaluation of different feature-based object recognition algorithms is presented. The task on which the performance of the different algorithms is tested is the classification of car models depicted on an image, restricting ourselves to the hard case of distinguishing car models of only one make. The presented work is conducted on images of cars but should help to get an insight into the different algorithms and their effectiveness, working with all kinds of untextured and shiny objects, as well.

This paper is organized as follows. In section 2 an overview of related work on the fields of detector- and descriptor-algorithm evaluation and vehicle classification is presented. Section 3 briefly describes the different algorithms evaluated in this work. Section 4 amplifies details about the designed system and introduces the used data set as well as the performed tests. Section 5 summarizes the experimental results and section 6 gives a conclusion and points to future work.

II. Related Work

Evaluating interest point detection and description algorithms is not a new idea, but most of the authors are evaluating these algorithms on rather well-suited data sets. Such data sets consist of images which show scenes or objects which are well textured and have matt or semi-matt surfaces. Handling untextured objects with shiny surfaces introduces problems to feature-based algorithms. Therefore, most of the existing evaluation work is not of great help for a task like classifying untextured and shiny objects.

In [4] Gauglitz et al. evaluate feature-based algorithms and introduce a test data set consisting of images taken

[1]These images can show the cars from controlled angles and lighting conditions, with typical colors and various equipment packages. Even rendered images coming e.g. from a car configurator or CAD data are conceivable.

from pictures of textured, matt objects, or scenes like brick walls. Algorithms such as ASIFT, ORB, and BRISK are not considered.

In [3] Moreels and Perona evaluate feature-based algorithms on 3D objects. They introduce a data set consisting of 100 rotated objects. Again, most of those objects are textured and matt. From the algorithms tested in the present paper only SIFT is evaluated by the authors.

The work [5] by Mikolajczyk and Schmid evaluates many different interest point detectors and descriptors. They use a self created data set consisting mostly of matt and textured images like graffiti or brick walls. SIFT is the only algorithm tested in their and also in the paper at hand. Similar to the aforementioned studies their work and results cannot simply be applied to the recognition of untextured, shiny objects.

Identifying the model of an object, especially the model of a car or vehicle from a fully controllable index data set of images is a lesser examined topic. Although there is no unique understanding of the terms, in most cases—and also in our paper—we will use the term of *vehicle (type) classification* when vehicle types have to be distinguished, such as saloon, estate car or van. The term *car (model) classification* will be used when the different car models of a specific make have to be distinguished, such as Volkswagen Polo, Golf, Passat or Audi A3, A4, A6.

In [6] Cheung and Chu present a system designed to identify the make and model of a car based on front and rear views. To determine matching images they use SIFT. An evaluation of the system is missing, as well as a comparison with other algorithms and images taken from different views.

In [7] Ma and Grimson propose a method to determine the type of a vehicle, for example Minivan or Sedan. To deal with the problem of untextured objects an edge representation is used based on the Canny edge detector paired with a modified version of SIFT which is more robust against reflections. As with the aforementioned work, only SIFT is considered and all images are taken from the same viewpoint.

In [8] Gu and Lee propose a car model recognition system based on pose and center estimation, shape feature extraction, and template matching. One crucial step in their work is the extraction of the car from the image which is done by video frame differences. Thus this approach is not easily applicable for image based car model recognition. Also object occlusion introduces further problems since this complicates the shape feature extraction and template matching.

III. System Description

Two different kinds of feature-based descriptor algorithms are evaluated in our work: float-based descriptors namely SIFT [9], ASIFT [10], and SURF [11] as well as descriptors based on binary feature values such as ORB [12] and BRISK [13]. In the following a short overview on the used algorithms is given:

SIFT is one of the best performing detector / descriptor combinations [5], [14]. Through the use of a Difference-of-Gaussian (DoG) detector in combination with a scale-space of the image and the assignment of a main orientation per interest point, followed by a descriptor based on the distribution of gradient orientations around these points SIFT achieves an invariance to scaling, rotation, translation and partially invariance to illumination as well as viewpoint changes.

ASIFT is a fully affine invariant extension of SIFT [10]. ASIFT uses an affine map to simulate changes of the camera viewpoint. This map is applied to the original image which is then processed by SIFT to detect and describe interest points. Through this fully affine extension ASIFT can outperforms SIFT with respect to image matching and recognition rates, of course at the price of higher computation efforts [10].

SURF is a detector / descriptor combination inspired by SIFT. SURF is more efficient to compute while showing comparable performance and in some cases also outperforms SIFT [11]. To detect interest points SURF uses a method called Determinant-of-Hessian. After interest points are identified on different scales an orientation is assigned to them making them scale and rotation invariant. The descriptor is based on the distribution of gradient orientations around a keypoint.

ORB is a detector / descriptor combination which is constructed for real-time systems and is therefore faster than SIFT and SURF while similar matching performance is achieved [12]. To detect scale and rotation invariant interest points, a scale-space is constructed on which an oriented version of FAST is applied [15]. For every interest point, a descriptor based on a rotation invariant extension of BRIEF [16] is then created. This descriptor builds up on binary comparisons of intensities around the identified interest point.

BRISK is like ORB a binary detector / descriptor combination which aims at high-quality description with low computational complexity like SURF, achieving comparable matching performance [13]. To determine interest points, BRISK applies FAST on an image scale-space and assigns orientations to each point. Afterwards a sampling pattern around a keypoint is used to perform the intensity comparisons which build up the binary descriptor.

For image classification we used the continuous approach [17], which works directly on the high dimensional descriptors calculated by the different algorithms. The indexed images are then ranked by the number of matching keypoints and the query image gets the same class as the image at rank one. As an alternative, a voting approach based on the classes of the three top ranked images is considered.

IV. Experimental Setup

To test the effectiveness of the algorithms it was necessary to create a testbed. For this task OpenCV was selected which already implements SIFT, SURF, ORB, and BRISK. To test ASIFT the author's C++ sources[2] where integrated into the OpenCV Sources.

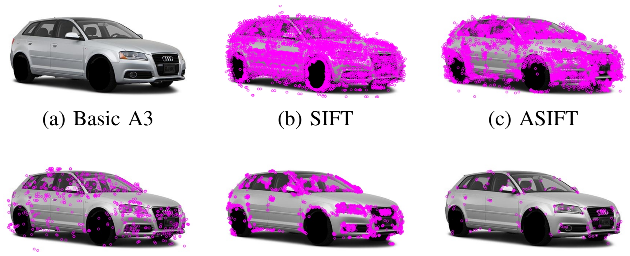
A. Data Set

The present scenario has the advantage that the data set used for creating the index is fully controllable, since there is a finite number of possible car configurations to choose a class—i.e. car model—from. Throughout this work, the

[2]http://www.ipol.im/pub/art/2011/my-asift/ (Last checked 22nd Apr. 2013)

(a) (b)

Fig. 1. *a* image from the index data set. *b* image from the query data set.



(a) Basic A3 (b) SIFT (c) ASIFT

(d) SURF (e) ORB (f) BRISK

Fig. 2. Interest points found by the different algorithms. The depicted car model is A3 as shown in (a).

index was created by using photos of different models without background clutter.

To limit the number of possible classes five car models from the German car maker Audi namely A3 to A7 in their base configurations were selected. The used index data set is comprised of 180 photos taken from those cars on a turning table without background noise, 36 images per model with a viewpoint change of 10° between images (see figure 1 (a)). For a fair comparison and in order to be able to correctly handle query images with OEM alloy rims, rims in the images in the index were occluded[3].

The used query data set consists of 75 photos, 15 images per model with roughly five different viewing directions[4], and three images per view (see figure 1 (b)). All cars are silver colored as the corresponding images from the index data set. The lighting conditions are good, without direct light and too much reflections or shadows which might change the descriptors due to their dependence on intensity changes. Obviously various extensions of the given data set with respect to the index data and the query images would be possible. We could integrate cars with various colors, we could test images of details such as headlamps or radiator grill. However, the expectation might be that the interest point detectors should be able to find expressive regions on their own. As a consequence, our tests can be seen as a first feasibility study in addition focusing on effectiveness and ignoring runtime efficiency at that point.

B. Tests Performed

For every image all interest points are detected and described (see figure 2). For every algorithm its standard parameters as proposed by the different authors were selected, except for SIFT whose detection parameters were adjusted to find as many interest points as ASIFT to ensure comparability between those two algorithms[5]. Afterwards, every query image is compared to all indexed images, and the latter ones are ranked depending on the number of matching keypoints. To compare the float-based descriptors the euclidean distance was used [9]–[11] and for the binary descriptors hamming distance was used [12], [13]. To filter out false positive matches, a distance ratio test between the 1^{st} and 2^{nd} best match for every query descriptor is performed and thresholding is applied [9]. When all matching images are ranked the class of the first or the most common class among the first three images respectively are used to determine the query image class. P@1 and P@3 are calculated and the arithmetic mean is computed per class. Hereby the distinctiveness of the interest points and descriptors is analyzed.

V. Experimental Results

Table I shows the mean precisions at rank 1 and 3 for different algorithm combinations and model classes. The used naming convention for the different tests states the used detector and the used descriptor, separated by a slash. The pure ASIFT setup shows its superiority against pure SIFT and pure SURF which is attributed to the simulation of changes in the camera axis. Also remarkable is the performance of pure ORB, which is nearly as good as pure ASIFT and thus ranks second. All classifications are better than picking the correct class by chance. While the performance of pure SIFT is consistently weak and favors the A7, other algorithms show better predictions for some models than for the others (see pure SURF and pure BRISK) whereas still others perform good on many classes and only fall short on some others (see pure ASIFT, pure ORB, SURF / SIFT, and ORB / SURF). Comparing table I considering the predictability per class also shows that no single class is preferred by all algorithms.

Table II shows two exemplary confusion matrices for pure ORB and SURF / SIFT combination. It can be seen, that the probability for the correct class is in most cases much higher than for a single other class.

Table III shows the average number of interest points per image found by the different detectors. It shows that a higher number of interest points does not necessarily yield a higher precision.

VI. Conclusion and Future Work

The circumstance that pure ASIFT performs better than pure SIFT is attributable to the enhanced interest point detection. Since ASIFT uses SIFT for interest point description this step is the same for both algorithms and does not explain the performance gain.

ORB and BRISK detectors both rely on an extended version of FAST. The adjustments done in case of ORB generate over ten times more interest points than in case of BRISK. To cover the influence of the two detectors on the descriptors, the algorithms were evaluated crosswise. Comparing the results of pure ORB, pure BRISK, ORB / BRISK, and BRISK / ORB shows, that the detector as well as the descriptor of ORB performs better than BRISK.

Pure ORB and pure ASIFT show similar performance even if ORB only uses nearly one-third of the interest points

[3]Occlusion of rims in the indexed images in some cases leads to a difference of 30% in mean precisions.

[4]The viewing directions are front, rear, side, 3/4-front and -rear.

[5]This adjustment does not significantly affect the performance of SIFT in comparison to the standard parameters.

TABLE I. MEAN PRECISIONS AT RANK 1 AND 3 FOR THE DIFFERENT ALGORITHMS.

SIFT / SIFT	ASIFT / ASIFT	SURF / SURF	ORB / ORB	BRISK / BRISK	SURF / SIFT	ORB / BRISK	BRISK / ORB	ORB / SURF	SURF / ORB	
0.13	0.73	0.13	0.80	0.13	0.47	0.60	0.07	0.60	0.40	A3
0.13	0.87	0.33	0.53	0.20	0.47	0.13	0.27	0.47	0.27	A4
0.07	0.80	0.27	0.60	0.47	0.60	0.73	0.53	0.60	0.27	A5
0.20	0.20	0.67	0.20	0.33	0.67	0.13	0.07	0.07	0.40	A6
0.67	0.40	0.27	0.40	0.20	0.33	0.27	0.20	0.40	0.27	A7
0.24	0.60	0.36	0.51	0.27	0.51	0.37	0.23	0.43	0.32	Mean

(a) Mean precision at rank 1.

SIFT / SIFT	ASIFT / ASIFT	SURF / SURF	ORB / ORB	BRISK / BRISK	SURF / SIFT	ORB / BRISK	BRISK / ORB	ORB / SURF	SURF / ORB	
0.42	0.47	0.11	0.69	0.13	0.58	0.60	0.13	0.58	0.27	A3
0.04	0.62	0.29	0.44	0.13	0.51	0.24	0.13	0.56	0.22	A4
0.17	0.62	0.31	0.58	0.44	0.62	0.67	0.31	0.47	0.20	A5
0.27	0.27	0.51	0.27	0.27	0.58	0.24	0.13	0.11	0.38	A6
0.51	0.47	0.27	0.56	0.27	0.36	0.42	0.16	0.49	0.29	A7
0.28	0.49	0.30	0.51	0.25	0.53	0.44	0.17	0.44	0.27	Mean

(b) Mean precision at rank 3.

TABLE II. CONFUSION MATRICES FOR TWO ALGORITHMS AND P@1, THE ABSCISSA SHOWS THE ACTUAL CLASS AND THE ORDINATE THE PREDICTED CLASS.

A3	A4	A5	A6	A7	
0.80	0.27	0.13	0.27	0.33	A3
0.00	0.53	0.13	0.07	0.13	A4
0.13	0.07	0.60	0.40	0.13	A5
0.07	0.13	0.07	0.20	0.00	A6
0.00	0.00	0.07	0.07	0.40	A7

(a) ORB / ORB

A3	A4	A5	A6	A7	
0.47	0.07	0.07	0.13	0.07	A3
0.07	0.47	0.27	0.07	0.13	A4
0.27	0.13	0.60	0.07	0.07	A5
0.13	0.20	0.07	0.67	0.40	A6
0.07	0.13	0.00	0.07	0.33	A7

(b) SURF / SIFT

TABLE III. AVERAGE NUMBER OF INTEREST POINTS FOUND PER IMAGE FOR THE DIFFERENT DETECTORS.

	$SIFT$	$ASIFT$	$SURF$	ORB	$BRISK$
$Keypoints$	5.007	4.526	507	1.656	120

compared to ASIFT. This is caused by the different kinds of detection approaches used by those algorithms. While ASIFT uses the DoG to find blob like structures, ORB uses an extended version of FAST, preferring interest points on corners which are more stable on untextured, shiny objects. Both algorithms also use different approaches for descriptor creation. While ASIFT creates float-based descriptors relying on the gradient orientations around interest points, ORB is a binary descriptor using comparisons of intensities around interest points. The good performance of ORB leads to the assumption that descriptors performing binary comparisons combined with interest points on edges and corners outperform computationally complexer float-based algorithms like SIFT or ASIFT if the task is to identify untextured, shiny objects.

To see the influence of a descriptor favoring interest points on corners and edges against one favoring interest points on blob-like structures, a crosswise combination of SURF and ORB was tested. The combination of the ORB detector with the SURF descriptor to leads better performance than a pure SURF approach. Whereas a combination of the SURF detector with the ORB descriptor leads to poorer performance than a pure ORB approach. This also leads to the assumption that interest point detectors preferring corners and edges perform better on the given task.

Comparing pure SIFT, pure SURF, and SURF / SIFT, a performance gain of SIFT descriptors when used with interest points provided by the SURF detector can be observed. This shows that interest points computed by SURF are more reliable than the ones from SIFT while descriptors computed by SIFT are more reliable than the ones from SURF.

Exploring the impact on recognition effectiveness of ASIFT's viewpoint simulation on other algorithms like ORB is further work and may lead to even better effectiveness of the latter mentioned algorithm.

REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, vol. 65, no. 1-2, Nov. 2005.

[2] V. R. Chandrasekhar, D. M. Chen, S. S. Tsai, N.-M. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeszczuk, J. Bach, and B. Girod, "The stanford mobile visual search data set," in *Proc. of the second annual ACM conference on Multimedia systems*, 2011.

[3] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," in *Tenth IEEE International Conference on Computer Vision, 2005.*, vol. 1, Oct.

[4] S. Gauglitz, T. Höllerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *Int. J. Comput. Vision*, vol. 94, no. 3, Sep. 2011.

[5] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proc. of IEEE Computer Society Conference on CVPR 2003.*, vol. 2, June.

[6] S. Cheung and A. Chu, "Make, model, and color recognition of cars," Department of Computer Science and Engineering University of California, San Diego, Tech. Rep., 2008.

[7] X. Ma and W. Grimson, "Edge-based rich representation for vehicle classification," in *Proc. of the Tenth IEEE International Conference on Computer Vision - Volume 2*, vol. 2, Oct.

[8] "Car model recognition by utilizing symmetric property to overcome severe pose variation," *Machine Vision and Applications*, vol. 24, 2013.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, Nov. 2004.

[10] J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM J. Img. Sci.*, vol. 2, no. 2, Apr. 2009.

[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, Jun. 2008.

[12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *IEEE International Conference on Computer Vision, ICCV 2011*, nov. 2011.

[13] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. of the 2011 International Conference on Computer Vision*, nov. 2011.

[14] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vision*, vol. 60, no. 1, Oct. 2004.

[15] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Procee. of the 9th European conference on Computer Vision - Volume Part I*, ser. ECCV'06. Springer-Verlag, 2006.

[16] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: binary robust independent elementary features," in *Proc. of the 11th European conference on Computer vision: Part IV*, ser. ECCV'10. Springer-Verlag, 2010.

[17] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Inf. Retr.*, vol. 11, no. 2, Apr. 2008.