Proceedings of the International Multiconference on Computer Science and Information Technology pp. 19–23
ISBN 978-83-60810-27-9
ISSN 1896-7094

Advanced scale-space, invariant, low detailed feature recognition from images - car brand recognition

Štefan Badura
University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia
Email: stefan.badura@fri.uniza.sk

Stanislav Foltán
University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia
Email: stanislav.foltan@fri.uniza.sk

***Abstract*—This paper presents analysis of a model for car brand recognition. The used method is an invariant keypoint detector - descriptor. An input for the method is a set of images obtained from the real environment. The task of car classification according its brand is not a trivial task. Our work would be a part of an intelligent traffic system where we try to collect some statistics about various cars passing a given area. It is difficult to recognize objects when they are in different scales, rotated or if they are low contrasted or when it is necessary to take into count high level of details. In our work we present a system for car brand recognition. We use scale space invariant keypoint detector and descriptor (SURF – Speeded-up Robust Features) for this purpose.**

I. Introduction

THIS paper presents a model for car brand recognition. Not many works have been published for such problem yet. We discuss a model for scale, space invariant car brand detection and recognition. The goal of our work is to provide information about car type approaching monitored region. For example car brands as “Ford”, “Kia”, “Volkswagen” etc. are considered. Recently, some promising approaches to detect foreground objects from images have been published: SIFT - Lowe (Scale Invariant Feature Transform) [1] and SURF - Bay et al. (Speeded Up Robust Features) [2]. Both methods detect interest points, called features, and also propose a method of creating an invariant descriptor for these features. The created descriptor is used as vector, uniquely identifying the found interest points. It has to be distinctive and robust for various scale-space deformations. Vectors can be used for matching of detected interest points even under a variety of disturbing conditions like scale changes, rotation, changes in illumination and viewpoints or image noise. The invariance is the most important ability of these keypoint detectors. The car brand can be in various representation in the image. An example is shown in the Fig.1.

The purpose of this paper is to review and investigate a method referred to as SURF. The SURF detector-descriptor method for our problem is analyzed. We use the OpenSURF library which we integrated into our software demo. In [3] we proposed a possibility to use the SIFT method for the same problem. It showed some positive results but average percentage of successful classification was not to high. The number of different car brands was 7 and the final percentage moved around 60%. Some car brands were classified with better results than others. As far as we know, no other works exist for car brand recognition problem. We focus on scale-space invariant detectors and descriptors. These seem to be a good compromise to other methods. The SIFT and SURF in various problem are used where patterns or objects from database in given scene (unknown image) are searched. Results from other types of common work are motivation for our problem. In [4] authors measure visual similarities between different visual entities with SIFT. In [5] the SIFT is used for face detection. In [6] the SURF is used for foreground detection. Comparative work done by Mikolajczyk and Schmid in [7] is a good reference for other types of interest points detectors and descriptors. The authors of SURF [2] claims SURF to be a superior to SIFT in terms of runtime execution while it is still providing good results with regards to feature point quality. The goal of this paper is not to compare SIFT against SURF or otherwise, but the goal is to analyze the possibility of using the SURF for car brand recognition from images taken in a natural outdoor environment.

Fig 1: A series of images to show various car brand invariances for the same car brand.

In the second part the SURF is discussed in more details. First the algorithm is analyzed from a theoretical standpoint, to provide a brief overview of how it works and what was the motivation of using it. In the third part a brief overview of preprocessing procedures for region of interest specifying is discussed. The 4th proposes experiments for the problem of car brand recognition. The paper is concluded in 5th part.

II. Speeded-Up Robust Features (SURF)

The SURF is scale and rotation invariant interest point detector and descriptor. It is designed to overcome relatively low computation time which SIFT has, by providing the same functionality. Much of the performance increase in SURF can be referred to the use of an special image representation known as the “Integral Image" [2]. The SURF is three step process:

- Interest points detection and localization.
- Feature vector construction – descriptor (interest point description).
- Descriptors matching.

A. Interest Point Detection and Localization

The SURF uses Fast – Hessian features detector and it is based on determinant *D(H)* of Hessian matrix *H(f)*. The Hessian matrix is the matrix of partial derivatives of two dimensional function *f(i,j)* (see eq. 1) [8][9].

$$H(f)=\begin{bmatrix} \frac{d^2 f}{di^2} & \frac{d^2 f}{didj} \\ \frac{d^2 f}{didj} & \frac{d^2 f}{dj^2} \end{bmatrix} \quad (1)$$

The determinant of this matrix *(H)*, known as the discriminant, is calculated by:

$$D(H)=\frac{d^2 f}{di^2}\cdot\frac{d^2 f}{dj^2}-\left(\frac{d^2 f}{didj}\right)^2 \quad (2)$$

Equations (1) and (2) are defined for continuous function f, but we work with images, so it means discrete space. Then *f(i,j)* as discrete function is considered which represents image. Point at *(i,j)* coordinates is a pixel intensity for the image. Approximated derivatives are computed by convolution with appropriate kernel (*H'* is approximation of *H*). For determinant approximation *(D')* formula is used:

$$D'(H')=D_{ii}\cdot D_{jj}-(0.9\mathrm{D}_{ij})^2 \quad (3)$$

Determinant computation is necessary for interest point detection (which are maxims in determinant matrix). In order to be able to detect interest points in scale space the notion should be introduced. A scale-space is a continuous function which can be used to find an extrema across all possible scales [10]. In SIFT scale-space as an image pyramid is used where the input image is iteratively convolved with Gaussian kernel in reduced size and so repeatedly sub-sampled. In SURF we do not need to resize images. Instead the convolution mask is re-sampled to ensure size invariance. This allows for multiple layers of the scale-space pyramid to be processed simultaneously and negates the need to subsample the image hence providing performance increase [8]. The idea is shown in Fig. 2. For appropriate and accurate interest point selection some other steps are processed. Firstly a treshold is applied, secondly non maximum supression is done and next step is interpolation. Fo more details see [2],[8].

B. Interest Point Descriptor

The SURF descriptor describes how the pixel intensities are distributed within a scale dependent neighborhood of each interest point detected by the Fast-Hessian [8]. The process of descriptor extraction consists of two parts. At first a reproducible orientation to each interest point is assigned. Secondly around each interest point a square window in appropriate scale is constructed. In this square a 64 value vector is computed which is the descriptor for given interest point. For more details see the Fig. 3. The SIFT extracts 128 values for one interest point compared to the SURF.

C. Descriptors matching

If descriptors of database items and new query are known, they are matched each other. We use simple principle and its modification. For each item in database find all keypoints. For each keypoint from the query image find the keypoint from database with the smallest distance (MSE). Two keypoints are considered as the same if the error of the best match is at least a factor of α smaller than the distance to the next closest descriptor. If not take another. Then count the number of descriptors that successfully matched.

III. Image preprocessing

In this part a brief introduction to preprocessing is discussed. The main purpose is to reduce image to a size which contains all the necessary information but the image remains as small as possible. The car brand represents just a small part compared to the whole image. Reducing an input image to the region around car brand has positive effect namely in two aspects: Lower number of detected interest points leads to more accurate computation (less error rate) and it improves the execution time. From the Fig. 4 is clear that we use license plate for image reducing into region of interest. The license plate is found and according its position a car brand is searched. We suppose that the car brand is above the license plate. It is not a rule but almost all cars satisfy this condition. In next part experiments are discussed.

IV. Experiments

In this part, we study whether SURF features are suitable and robust enough for our problem. Since we have chosen a typical pattern classification approach for solving the problem of car brand recognition, it is necessary to have a suffi-

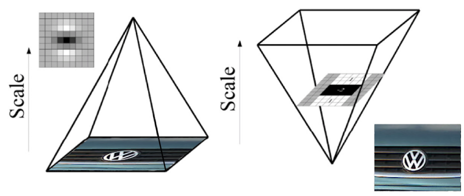

Fig 2: Pyramid. On the left - traditional approach (SIFT), on the right approach used in SURF.

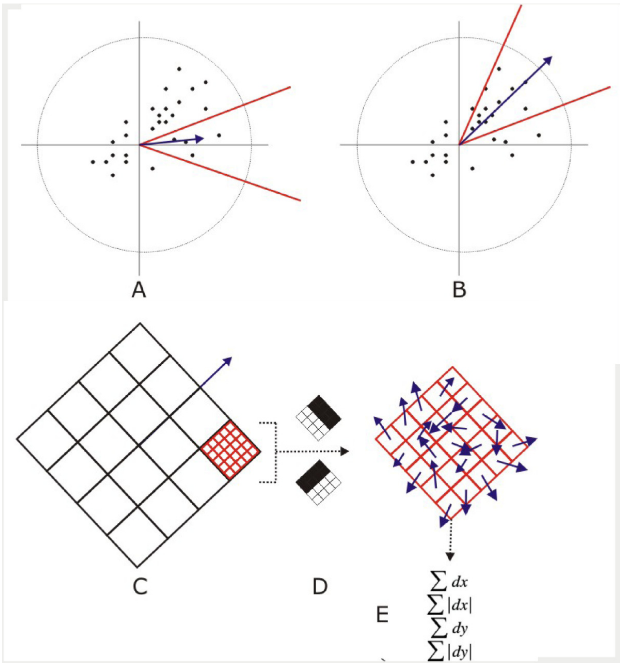

Fig 4: The SURF - descriptor. A, B represent orientation assignment. C is square around interest point with orientation. D and E show 64 value descriptor computation.

Fig 3: Process of initialization region of interest. At first license plate is found (1). Then above the license plate region of interest which holds car brand is set (2). Next the matching process is initialized (3,4) after descriptors are extracted.

cient data set. The data corpus of altogether 189 different samples was collected in the entrance to a car park. Table 1

Fig 5: Samples in database. Each image represents one class. These samples are considered as templates for matching.

TABLE I
POSPIS

Car brand	Num. Of samples
Audi	21
BMW	22
Citroen	8
Ford	16
Kia	12
Mercedes	16
Opel	24
Peugeot	14
Renault	19
Skoda	8
Volkswagen	29

shows number of samples for each brand. Our database consisted of 11 different car brands. Original image size was 640x480. When the region of interest was reduced the size moved around 200x200 pixels.

For given set of samples we tested the SURF method. Before experiments a template database was created. The database contained one sample for each brand. We chose the image which seemed to be optically the "best" (without noise, rotation, well visible etc.) for each class (see Fig. 5). From samples the SURF descriptor was computed and saved to file. Detected interest points which did not affect the car brand were deleted (see Fig 6.). The criteria used to evaluate

TABLE II
EXPERIMENTAL RESULTS.

Car brand	Result (α=0.6)	Result (α = 0.8)
Audi	0.52	0.9
BMW	0.14	0.09
Citroen	0.63	0.5
Ford	0	0
Kia	0.08	0.33
Mercedes	0.06	0
Opel	0.71	0.88
Peugeot	0.43	0.29
Renault	0.21	0.42
Skoda	0.25	0.38
Volkswagen	0.66	0.66

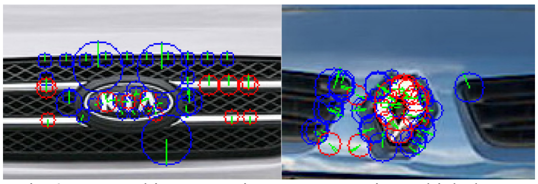

Fig 6: Detected interest points. Interest points which do not affect the car brand are deleted. Most of them describe the mask.

the performance of the detector is number of correctly matched samples from the database to samples from available testing dataset (new queries). We created tests based on average error and tests based on number of descriptor matches for certain class. The second approach provided better results but it was unfair because not all samples had the same number of detected keypoints. Summary is shown in table 2. We modified also the α constant (distance between the best and the next closest descriptor). Better results were achieved with α=0.8 (default values was 0.6) but this fact is clear because we are not so strict to coincidence. Descriptor which fitted the most produces smaller error (the same descriptors produce error equals to 0).

An average result for all samples moved around 45%. Table 1 shows interesting results. Some car brand are recognized with high percentage but the other ones are recognized poorly. The reason could be for example a small database of samples. For some car brands just low number of interest points was detected (see Fig. 7) and therefore they are recognized with insufficient result. The SURF method is quite strong tool from various aspects. Similar (the same) objects were found with almost 100%. Objects in different scales were also recognized with high percentage. Fig. 8 shows some results where the best matched keypoint as an example is depicted. Images on the left are samples from the database and images on the right are new queries.

Fig 7: An example of positive matching. The same interest points were detected in different images.

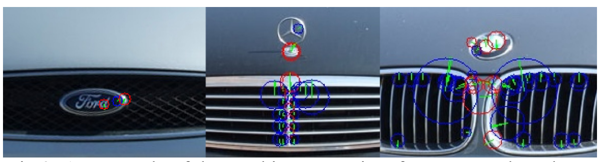

Fig 9: An example of detected interest points for some car brands. In all image low number of keypoints is detected. It is also shown that the mask is quite informative (or noisy) for some car brands (bmw and mercedes). If we would consider the mask, result could improved but the mask is not standardized.

Fig 8: An example of negative matching. Even if the images are not matched, some similarities can be found for highlighted interest point.

V. Conclusion

In this work, we investigated the usage of SURF descriptors for car brand recognition. The final percentage was not very high. The main purpose was to analyze the possibility of this method for our problem and from this standpoint we obtained some valuable results. As the future work a new strategy for creating database is necessary to build. It would be also appropriate to analyze descriptor values and that especially for car brands with small amount of detected interest points. For some brands, which were recognized very poorly, different invariant key point detector should be tested (MSER – maximal stable external regions).

References

[1] D. G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints.* In International Journal of Computer Vision, 60, 2, 2004, pp. 91–110.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. *Surf: Speeded up robust features.* Computer Vision and Image Understanding (CVIU), 2008, pp. 346–359,

[3] S. Foltán, Š. Badura. *Robust Car Brand REcognition From Camera Image.* In MENDEL - International Conference on Soft Computing, 2010.

[4] E. Shechtman, M. Irani, *Matching Local Self-Similarities across Images and Videos,* Computer Vision and Pattern Recognition, CVPR '07. IEEE Conference, 2007.

[5] Mohamed Aly. *Face Recognition using SIFT Features*, CNS186 Term Project Winter, 2006.

[6] S. Badura, A. Lieskovsky. *Intelligent traffic system: cooperation of MANET and image processing,* In First International Conference on Integrated Intelligent Computing IEEE, 2010.

[7] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors.* IEEE Transactions on Pattern Analysis & Machine Intelligence, pp: 1615–1630, 2005.

[8] C. H. Evans, *Notes on the OpenSURF Library,* CSTR-09-001, University of Bristol. January 2009.

[9] J. Bauer, N. Sunderhauf, P. Protzel. *Comparing Several Implementations of Two Recently Published Feature Detectors,* In Proc. of the International Conference on Intelligent and Autonomous Systems, IAV, Toulouse, France, 2007.

[10] A. Witkin. *Scale-space filtering,* in proc. Of Artif. Intell, conference, pp: 1019-1021, 1983.