Classification of wild animals situated in Slovak Country

Slavomir Matuska, Rober Hudec, Univerzity of Zilina

Abstract
Object classification is extensive and difficult task, which consist of three main steps – feature extraction from training database, training classifier and evaluation of query image. This presented paper proposed about local key point detectors and descriptors, bag of key points method in combination with SVM.

1. Introduction
In computer vision, based idea of object recognition creation a representation of particular classes, which characterized appearance of object from given class and in this way determine unknown objects class. In this paper, 5 classification classes from wild animals situated in Slovak country were chosen to classify. Namely: wolf, fox, wild boar, deer and brown bear.

Success rate of object recognition depends especially on good object representation. Moreover, object representation depends on good object characterization, which can be achieved by visual descriptors. Therefore, this paper proposes some different visual descriptors in combination with SVM and their performance in object classification [1, 2].

2. Object recognition process
Object recognition process is shown in Fig. 1 and can be divided into two parts: Training and Testing. Task of training part is to create a classification model from the training data. Training data contain a collection of images of each class. The extraction of primary images features are extracted at their low-level by different methods. Most common used methods are SIFT (Scale Invariant Feature Transform), SURF (Speeded Up Robust Features), MSER (Maximally Stable External Regions), GFTT (Good Features to Track), FAST (Features from Accelerated Segment Test) etc. These methods will be detailed described in section 3. Moreover, a low-level features extracted from images are used to creation a classification model.

To the input of the testing part enters an images and their still picture objects designated to the classification. Moreover, these objects have the same metadata description like data in training part. Based on these data, the classifier is able to regarding to classification model successfully evaluate an unknown objects to the appropriate class [1, 2].

3. Visual Descriptors
Visual descriptors are used to capture the local appearance of objects. They are calculated from the neighbor pixels and result of this process is feature vector. Creation of feature vector consist of two steps: key point detection and key point description.

3.1 Key points detectors
Task of detector is to find key points in the image. There are many methods to detect key points. In this part will be described SIFT, SURF, FAST, MSER and GFTT methods for key points detection.

SIFT: the difference of Gaussians operator is applied to an image at different scales to identify features of potential interest – key point. Then the precise position of key points is dedicated [4, 5].

SURF: detector is based on the determinant of the Hessian matrix. The discriminant value is used to classify the maximum and minimum of the function by second order derivative test [4, 7].

MSER: instead of detecting key points, MSER detects regions which are darker or brighter than backgrounds. It is affinely-invariant and robust to change of illuminations [3, 10].
FAST: is high-speed corner detector, which use corner to capture object appearance. This algorithm was evolved for real-time frame-rated application [7].

GFTT: finds the most prominent corners in the image or in the specified image region [9].

3.2 Key points descriptors

Task of key points descriptor is to describe key point by the n-dimensional feature vector. In this paper were used these descriptors: SIFT, SURF and Opponent colour descriptors.

SIFT: is the most widely used local visual descriptors. It has reasonable invariance to changes in illumination, rotation, scaling, and small changes in viewpoints. The SIFT descriptor of key point is obtained by first computing the gradient magnitudes and orientations of pixels in the neighborhood region of the key point, using the scale of the key point to select proper Gaussian kernel to blur the image. The orientation of histograms within the sub-regions around the key point are computed and combined into 128 dimensional SIFT feature vector. Produced vector is normalized to improve the invariance to changes of illumination. More detailed information about SIFT can be found in [4, 5, 6].

SURF: Results of SURF descriptor is feature vector of length 64 and is invariant to rotation, scale, brightness and after reduction to unit length or contrast. More detailed information about SURF can be found in [4, 7].

Opponent Color Descriptors: Opponent colour descriptors (OpponentSIFT and OpponentSURF) describe all of the channels in the opponent colour space using common descriptors. Opponent histogram is a combination of three 1D histograms based on the channels of the opponent colour space and this space is given by (1):

\[
\begin{pmatrix}
O_1 \\
O_2 \\
O_3
\end{pmatrix} = \begin{pmatrix}
\frac{R-G}{\sqrt{2}} \\
\frac{R+G-2B}{\sqrt{6}} \\
\frac{R+G+B}{\sqrt{3}}
\end{pmatrix}
\]

(1)

where O1 and O2 contain red-green and yellow-blue opponent pairs and describe the color information in the image. The intensity information is represented by channel O3. R, G and B are channels of RGB colour space: red, green and blue. All three channels are described by SIFT or SURF and therefore they are called OpponentSIFT and OpponentSURF descriptors [8].

4 Animal Classification

The collection of features or parameters characterizing the object by classifications methods to handle classification task are used. There are two phases of creation a classification model. First, training data collections are used to set up the classification model parameters to distinguish different classes. Then, the classifier is able to regarding to classification model parameters successfully evaluate an unknown objects to the appropriate class [9, 11]. In this work, for classification model combination bag of keypoints and Support Vector Machine (SVM) methods are used.

4.1 Bags of Keypoints

Classification method called bags of keypoint is based on vector quantization of affine invariant visual descriptors of object in images. The main advantages of this method are their simplicity, computationally efficiency and invariance in affine transformation and change in illumination. The main steps of this method are:

- description of the object in images for a set of labeled training data collection,
- constructing a set of vocabularies using K-means algorithm,
- extracting bags of keypoints for these vocabularies,
- applying and training multi-class classifier using the bags of keypoints as features vectors [12].

4.2 Support Vector Machine

A SVM is classification method related to the belongs to the family of supervised learning methods that analyze data and recognize patterns. It is non-probabilistic binary linear classifier. SVM belongs to the group of model based classifiers. Training algorithm constructs the model that represents patterns as points in vector space. Task of SVM classifier is found an optimal hyperplane with maximum margin between data of two different classes. Development of the classification system includes separating data into training and testing sets. To separate data of different classes, SVM maps feature vectors into a higher dimensional space using a kernel function [13, 14]. In this work, radial basic function (RBF) kernel was used.

5. Experimental result

Training database consist of 5 classes: wild boar, wolf, brawn bear, deer and fox. The examples of images from training database are shown in Fig 2. Ten images per class were randomly chosen from training database and were used as test database.
The first step of the test process, the low-level features from training images were extracted. In next step, the extracted descriptors together with annotation record in order to create a representation of particular class were used. Thus, created data representation consists of extracted descriptor data just test database only.

These data enters the process of constructing vocabulary using k-means clustering algorithm. Then, bag of key points for vocabulary were extracted. To extract bag of key points, algorithms for matching training descriptors with cluster centre in vocabulary were used. For each feature data extracted from test image by selected descriptor, BruteForce matcher finds a cluster centre in vocabulary. To the designation of feature vector and cluster centre distance, the Euclidean distance was used. Similar approach how to find out the minimum distance of feature vector and cluster centre is called FlannBased matcher. Thus, metadata for SVM classifier to create classification model for particular classes, extracted bag of keypoints were used.

In experiments a total 6 key points detectors, namely, SURF, SIFT, STAR, FAST, MSER and GFFT were used. Moreover, to describing a keypoint by four descriptors: SIFT, SURF, OpponentSIFT or OpponentSURF and two metrics/matchers: Brute Force or FlannBased were used too. All combinations of detectors, descriptors and matchers were combined into 48 standalone runs and they were programmed in C++ language with support of OpenCV (Open source Computer Vision) library.

Average score of classification for combination SIFT descriptor, SURF, FAST, MSER and GFFT detectors and two matchers is shown in figure 3.

Average score of classification for combination SURF descriptor, SIFT, SURF, FAST, MSER and GFFT detectors and two matchers is shown in figure 4.

Average score of classification for combination OpponentSIFT descriptor, SIFT, SURF, FAST, MSER and GFFT detectors and two matchers is shown in figure 5.

Average score of classification for combination OpponentSURF descriptor, SIFT, SURF, FAST, MSER and GFFT detectors and two matchers is shown in figure 6.
6. Conclusion

In this paper, five key point detectors and two key point detectors in combination with SVM classification method were presented. From the realized experiment is evident that the highest classification success rate 84% was achieved by algorithm in combination MSER detector, SIFT descriptor and Flann Based matcher. In addition, few more combination achieved success rate higher than 80%. Those classification models can be used in combination with video segmentation methods to classify moving animals. This information can determine the migration potential of wild animals in Slovak regions and help in many areas of life.

Acknowledgement

The work presented in the paper has been supported by the Slovak Science project Grant Agency, Project No. 1/0705/13 "Image elements classification for semantic image description" and EUREKA project no. E! 6752 – DETECTGAME: R&D for Integrated Artificial Intelligent System for Detecting the Wildlife Migration.

Bibliography (Style Chapter)