Scientific
Research
Publishing

Animal Classification System Based on Image Processing & Support Vector Machine

A. W. D. Udaya Shalika, Lasantha Seneviratne

Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
Email: udayashalika@gmail.com, Lasantha.s@sliit.lk

Abstract

This project is mainly focused to develop system for animal researchers & wild life photographers to overcome so many challenges in their day life today. When they engage in such situation, they need to be patiently waiting for long hours, maybe several days in whatever location and under severe weather conditions until capturing what they are interested in. Also there is a big demand for rare wild life photo graphs. The proposed method makes the task automatically use microcontroller controlled camera, image processing and machine learning techniques. First with the aid of microcontroller and four passive IR sensors system will automatically detect the presence of animal and rotate the camera toward that direction. Then the motion detection algorithm will get the animal into middle of the frame and capture by high end auto focus web cam. Then the captured images send to the PC and are compared with photograph database to check whether the animal is exactly the same as the photographer choice. If that captured animal is the exactly one who need to capture then it will automatically capture more. Though there are several technologies available none of these are capable of recognizing what it captures. There is no detection of animal presence in different angles. Most of available equipment uses a set of PIR sensors and whatever it disturbs the IR field will automatically be captured and stored. Night time images are black and white and have less details and clarity due to infrared flash quality. If the infrared flash is designed for best image quality, range will be sacrificed. The photographer might be interested in a specific animal but there is no facility to recognize automatically whether captured animal is the photographer's choice or not.

Keywords

Image Processing, Support Vector Machine (LIBSVM), Machine Learning, Computer Vision, Object Classification

1. Introduction

Object detection and recognition based on image processing is vastly concentrating field in research. The motivation for this project is to build a system for automatically detecting and recognizing wild animals for the animal researchers & wild life photographers. The animal detection and recognition is an important area which has not been discussed rapidly. However this project is targeted to build a system which would help animal researchers and wild life photographers who devote their precious time for studying animal behavior. Technology used in this research can be modified further to use in applications such as security, monitoring purposes, etc.

Wild life photography is regarded as one of the most challenging forms of photography. It needs sound technical skills such as being able to capture correctly. Wild life photographers generally need good field of crafting skills and a lot of patience. For example, some animals are difficult to approach and thus knowledge of the animals' behavior is needed in order to be able to predict their actions. Sometimes photographers need to stay calm and quiet for many hours until the exact time. Photographing some species may require stalking skills or the use of hide/blind for concealment. A great wild life photograph also is the result of being in the right place at right time. Because of the highly consuming nature of the time it costs too much and same time life can be in danger involving in the business. Usually animal researchers often travel to remote locations all over the world. Hostile conditions are often the norm in the life of photojournalist. Some will sit for hours and hours before snapping a shot which is worthy to be sold. The photographers should be brave enough to stay in hostile environment without comfort and with great patience till animals appear. To have a perfect scene or outcome, it is always coming by fewer disturbances to the natural behavior of animal. Due to their high sensitivity animal can easily identify the presence of human being. So photographers need to be ready for facing any critical moment in the jungle because we cannot predict what will happen in next moment. The DSLR cameras which use in the industry are also quite expensive and there are limitations in shutter on-off cycle. So proper recognition must exist in such equipment. The higher the quality of the images, the more space in memory it takes. So it also encourages to having proper recognition in wild photography.

In this paper, we selected to be conservative and limited our endeavor to only three kinds of animals. This is not a random selection because it is always a challenging part to collect a set of data in animals to create proper data base. Our motivation is to achieve the recognition part, and thus we used the ACE media mpeg project for feature extraction, Support vector machine (LIBSVM) to get the probabilistic outcome. For the feature extraction we used several descriptors like Color layout descriptor (CLD), Color structure descriptor (CSD), Edge histogram descriptor (EHD), Homogeneous texture descriptor (HTD), Region shape, Contour shape. Using these descriptors, research was carried out about how the individual descriptors perform and the performances of combined descriptors.

1.1. Literature Review

The low level audio visual feature extraction was used for retrieval & classification. In this paper we present an overview of a software platform that has been developed within the ace Media project, termed the ace Toolbox that provides global and local low level feature extraction from audio-visual content. The toolbox is based on the MPEG-7 experimental Model (XM), with extensions to provide descriptor extraction from arbitrarily shaped image segments, thereby supporting local descriptors reflecting real image content. This paper will describe the architecture of the toolbox as well as provide an overview of the descriptors supported to date. It will also briefly describe the segmentation algorithm provided. Then will demonstrate the usefulness of the toolbox in the context of two different content processing scenarios: similarity-based retrieval in large collections and scene-level classification of still images [1].

This paper will concern on observing animal behavior in wild life using face detection and tracking. It will present an algorithm for detection and tracking of animal faces in wildlife videos. As an example the algorithm is applied to lion faces. The detection algorithm is based on a human face detection method, utilizing Haar-like features and Ada Boost classifiers. The face tracking is implemented using the Kanade-Lucas-Tomasi tracker and by applying a specific interest model to the detected face. By combining the two methods in a specific tracking model, a reliable and temporally coherent detection/tracking of animal faces is achieved. In addition to the detection of particular animal species, the information generated by the tracker can be used to boost the priors in the probabilistic semantic classification of wildlife videos [2].

Rapidly detecting animals in natural scenes, evoked potential studies indicate that the corresponding neural signals can emerge in the brain within 150 msec of stimulus onset (S. Thorpe, D. Fize, & C. Marlot, 1996) and

eye movements toward animal targets can be initiated in roughly the same timeframe (H. Kirchner & S. J. Thorpe, 2006). Given the speed of this discrimination, it has been suggested that the underlying visual mechanisms must be relatively simple and feed forward, but in fact little is known about these mechanisms. A key step is to understand the visual cues upon which these mechanisms rely. Here we investigate the role and dynamics of four potential cues: two-dimensional boundary shape, texture, luminance, and color. Results suggest that the fastest mechanisms underlying animal detection in natural scenes use shape as a principal discriminative cue, while somewhat slower mechanisms integrate these rapidly computed shape cues with image texture cues. Consistent with prior studies, little role for luminance and color cues throughout the time course of visual processing, even though information relevant to the task is available in these signals [3].

Face identification method can be used for non-native animals to be used for intelligent trap. They developed a face identification method to distinguish target non-native alien animals from other native animals using camera captured images. When a camera recognizes a targeted animal walked in to the cage, it traps the animal in the cage. Here, we set raccoon as target non-native animal, and detect its face region by using HOG features. However, the raccoon face detector often confused by raccoon dog, which is a native animal to be preserved. So, after detecting raccoon face candidates, we distinguish them by several features and SVM. Some experimental results show that we can completely distinguish raccoon and raccoon dog from camera captured images [4].

Object recognition system is capable of accurately detecting, localizing, and recovering the kinematic configuration of textured animals in real images. Deformation model of shape automatically from videos of animals was built and an appearance model of texture from a labeled collection of animal images, and combine the two models automatically. We develop a simple texture descriptor that outperforms the state of the art. We test our animal models on two datasets; images taken by professional photographers from the Corel collection, and assorted images from the web returned by Google. It demonstrates a quite good performance on both datasets. Comparing the results with simple baselines, it was evident that for the Google set, can recognize objects from a collection demonstrably hard for object recognition [5].

Detecting the heads of cat-like animals, adopting cat as a test case, we show that the performance depends crucially on how to effectively utilize the shape and texture features jointly. Specifically, study has proposed a two-step approach for the cat head detection. In the first step, two individual detectors were trained on two training sets. One training set is normalized to emphasize the shape features and the other is normalized to underscore the texture features. In the second step, we train a joint shape and texture fusion classifier to make the final decision. A significant improvement can be obtained by the two step approach. In addition, study also propose a set of novel features based on oriented gradients, which outperforms existing leading features, e.g., Haar, HoG, and EoH. Study evaluates its approach using a well labeled cat head data set with 10,000 images and PASCAL 2007 cat data [6].

Detection and identification of animals is a task that has an interest in many biological research fields and in the development of electronic security systems. We present a system based on stereo vision to achieve this task. A number of criteria were being used to identify the animals, but the emphasis is on reconstructing the shape of the animal in 3D and comparing it with a knowledge base. Using infrared cameras to detect animals is also investigated. The presented system is work in progress [7].

1.2. Existing Systems & Draw Backs

The wild life camera DVR utilizes infrared technology which will capture great footage at any time of the day or nights as well as being supplied in a sturdy weatherproof and camouflaged box. Record full color video footage or 8 MP photographs produced by the wild life camera is transferred to a SD card and then review on a PC (using USB cables). The built in rechargeable battery can last (depending on the activity) up to 2 weeks. Existing technology has following features

   -Motion triggered and adjustable infrared (PIR) sensitivity
   -Rechargeable battery life 2 weeks
   -2.5" LCD Screen
   -Auto switch color images in day/B&W night images
   -SD card slot (2 GB)
   -Multi shot of 1 - 3 pictures
   -Programmable video length

-Programmable 10 sec to 990 sec delay between triggers
-No flash uses 54 IR LEDs to illuminate the coverage area
-Water proof housing
-Dimensions $160 \times 120 \times 50$ mm [8]

The major drawback of this is the no detection of animal in different angles. Whatever passes the camera will be automatically captured. Night time images are black and white and have less details and clarity due to infrared flash quality. If the infrared flash is designed for best image quality, range will be sacrificed. The photographer might be interest about a specific animal and there is no facility to recognize automatically whether captured animal is the photographer's choice or not.

2. Methodology

As the project contains both hardware & Software, need to focus on both parts separately. But this research paper mainly focuses on classification and recognition. Considering completion of the study hardware side and motion tracking has been focused. For the detection, camera rotation and communication the initial circuit had been tested on a bread board. The 16f877a microcontroller is the heart of hardware part and it has connected via max232IC over serial to USB cable to carry data to PC. When communication setting baud rate is 9600 bits/sec PIR sensors had been tested separately by applying proper voltage. The servos which controlling the camera movement is needs to program on microcontroller to provide suitable pulse widths in appropriate time (**Figure 1**).

Above pulse (**Figure 2**) period is 20 ms and to get the 20 ms repeat rate Timer 0 generates an interrupt at regular intervals. Timer 0 is driven (in this case) from internal oscillator. This is further divided, inside the pic, by 4 (Fosc/4). Prescaler value is useful since using the maximum setting 1:256 can get longer timer overflow *i.e.* when the timer passes through the values 255 to 0.

The 256 denominator is there because Timer 0 only overflows after 256 counts because it is an 8 bit timer.

Figure 3 shows the final pcb with components. Supply voltage was send through the 7805 regulator to maintain supply voltage of 5 V. Suitable capacitors were connected for smoothing purposes. 20 MHz oscilator, max 232 ic and other resistors and capacitors connected as in the design.

2.1. Classification

For the feature extraction process ACE media project as used which was capable of handling several feature extraction processes. The design was based on the architecture of MPEG-7 experimentation Model (XM) the official reference software of the ISO/IEC MPEG-7 standard. In addition to a more "light weight" and modular design, a key advantage of the ACE toolbox over the XM was the ability to define the process regions in the case of image input. Such image regions can be created either via a grid layout that partitions the input image in
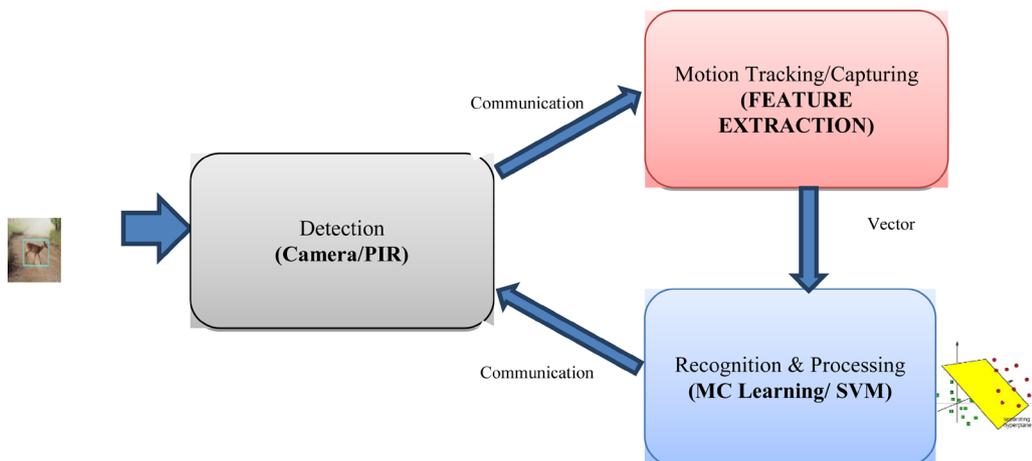
Figure 1. Whole process in nutshell. When the animal is presence detection, tracking, communication and recognition take place.
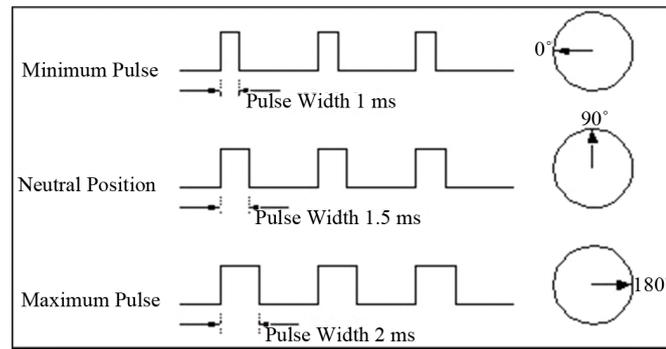
Figure 2. Standard servo controlling pulses [9].

Figure 3. Final PCB with hardware.

to user defined square regions or a segmentation tool that partitions the image in to arbitrary shaped image regions, that reflect the structure of the object present in the scene.

The current version of the ace tool box supports the descriptors listed in **Figure 4**. The visual descriptors are classified in to four types: color, texture shape & motion (for video sequences). Currently there is only a single video based and audio descriptor supported, but this will be extended in the future. The output of the ace Toolbox was a XML file for each specified descriptor, which for image input relates to either the entire image or separate areas within the image. An example of typical output is shown in **Figure 5**. The tool box adopts a modular approach, whereby APIs were provided to ensure that the addition of new descriptors was relatively straightforward. The system has been successfully complied and execute on both windows-based and Linux based flatforms [1].

So initially with a aid of C programming file handling techniques and the extracted features need to be in format in **Figure 5** in to a text file. Reason is that it can be use in LIBSVM.

Formats should be as per the **Figure 6** there for use in LIBSVM. +1 in the above dataset determines that it is positive dataset thus for negative dataset then it need to be begin with −1. Since its necessity to extract features of thousands of images C code was used. It facilitated numbering the total images in sequence and accessing them in iteration by time to time and features were saving in one particular text file.

After the feature extraction concern was on developing a proper training model. So LIBSVM had been used and it was pretty straightforward. LIBSVM is a library for support vector machines (SVM). LIBSVM has gained wide popularity in machine learning and many other areas. The followed process as below

-Conduct simple scaling on the data
-Consider the RBF kernal
$$K(x, y) = e^{-\gamma \|x - y\|^2}$$
-Use the cross validation to find the best parameter C and γ
-Test

The original data maybe too huge or small in range, thus it was rescaling to a proper range so that training and predicting will be faster. The main advantage of scaling was to avoid attributes in greater numeric ranges.
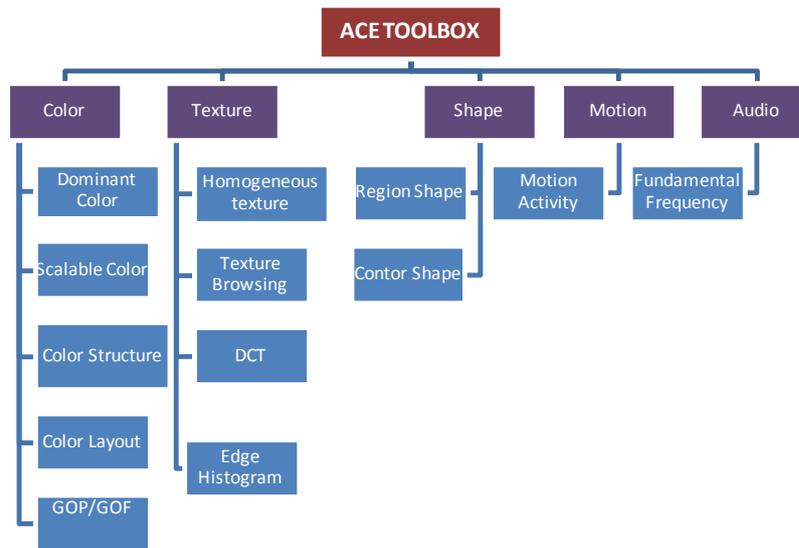
Figure 4. Supporting descriptors.
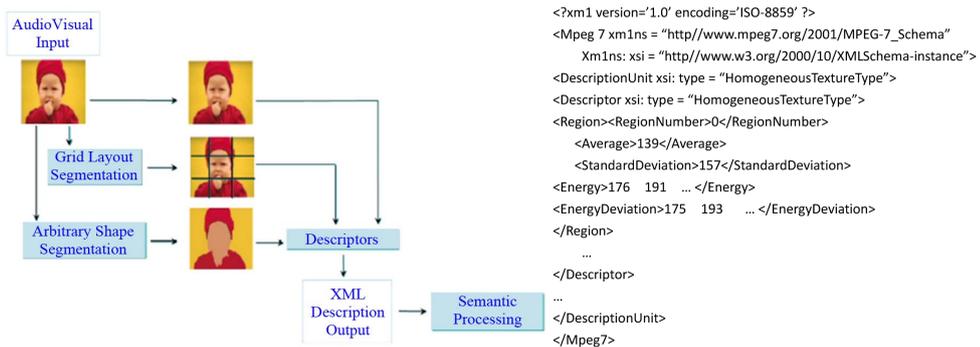
Figure 5. Overview of ace tool box & output.
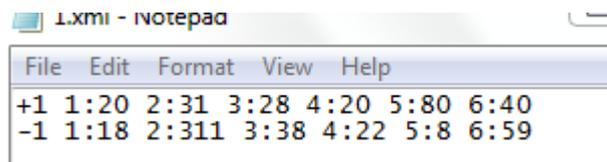
Figure 6. LIBSVM readable format.

Another advantage was to avoid numerical difficulties during the calculations. Linearly scaling for each attribute to the range [−1 to +1] or [0 to 1] was done. Below code should be run in command prompt for scaling the training set.

After scaling the dataset, kernel function was chosen for creating the model. Four basic kernels are linear, polynomial, radius basis function and sigmoid. In general RBF kernel is reasonable first choice. A recent result shows that if RBF kernel is used with model selection, then there is no need to consider linear kernel. The kernel matrix using sigmoid may not be positive definite and in general its accuracy is not better than RBF. Polynomials kernels are acceptable but if a high degree is used, numerical difficulties tend to happen. In or case also RBF kernel has been used. There are two parameters for RBF kernel: C and γ. linear kernel has a penalty parameter C. It is not known before beforehand which C and γ are best for given problem. Consequently some kind of model selection (parameter search) must be done. The goal was to identify good C and γ value, so that the classifier can accurately predict unknown data (*i.e.* testing data). For selecting the best parameter value grid.py in the libsvm-3.11 tools were used directory. Grid.py is a parameter selection tool for C-SVM classification us-

ing RBF (radial basis function) kernel. It uses cross validation (CV) technique to estimate the accuracy of each parameter combination in the specific range and helps you to decide the best parameters. To run the grid.py python interface is needed and a gun plot. In python code needed to do some modification to access the directories in LIBSVM windows and gunplot.exe.

To have higher cross validation accuracy can use 5 fold cross validation. After finding proper C and γ values need to create training model. After creating training model can see the predicted output (precision) for testing data by running SVM-predict.

2.2. Results

Based on above commands for three animals, using individual descriptors study was carry out to test how it will be the outcome. Two color descriptors, two texture descriptors, and two shape descriptors were used separately to check the ability of recognition of animals. The results gathered are summarized in **Table 1** [10].

2300 of positive images of tigers 960 of negative images like other animals, vehicle and etc. and 422 of tigers of testing were used. 2027 of positive images of dogs 752 negative images like other animals, vehicles and etc. and 1000 images of dogs of testing set was used. Cats images also taken around same amounts. As per the results on **Table 1** though it provide some good results of recognition +ve images, recognition of −ve images are very poor. When the individual descriptors were considered edge histogram descriptor performed somewhat satisfactory level compare to other descriptors.

As per the **Figure 7** it was clear that the variation between recognition of negative and positive images of individual descriptors. Most of the descriptors recognition rate was quite higher but unable to differentiate negative from images.

As per the **Figure 8** which was plotted with cross validation accuracy can identify the difference very clearly as positive and negative image efficiencies. So overall efficiency fails and it was not suitable for proper system.

Table 1. Individual descriptor performances.

Animal	Descriptor	C	γ	Cross validation accuracy	Accuracy for recognition (−) images %	Accuracy for recognition (+) images %
Tiger	Color Layout	8	0.5	85.1786	22.222	88.3581
	Color Structure	2	0.5	93.6786	44.4444	78.4615
	Edge Histogram	8	0.125	92.7143	44.44	93.0756
	Homogeneous Texture	8	0.5	97.7857	25.763	76.1538
	Region Shape	2	0.5	82.3445	0	99.74
	Contour Shape	2048	8	86.9643	0	76.6154
Dog	Color Layout	8	2	77.5285	38.913	93.3
	Color Structure	8	0.125	82.2958	27.812	91.8
	Edge Histogram	32	0.007825	82.6916	53.871	94.2
	Homogeneous Texture	8	0.5	77.7258	32.2222	92.9
	Region Shape	8	0.03125	75.2429	11.80	96.5
	Contour Shape	32,768	8	76.7182	0	97.6
Cat	Color Layout	8	0.5	80.8973	28.893	89.713
	Color Structure	8	0.5	83.6957	37.777	76.888
	Edge Histogram	8	0.125	90.9532	39.99	92.865
	Homogeneous Texture	8	0.5	80.2558	25.6771	93.25
	Region Shape	2	0.5	80.3445	11.5973	90.597
	Contour Shape	2560	8	84.9248	0	78.653
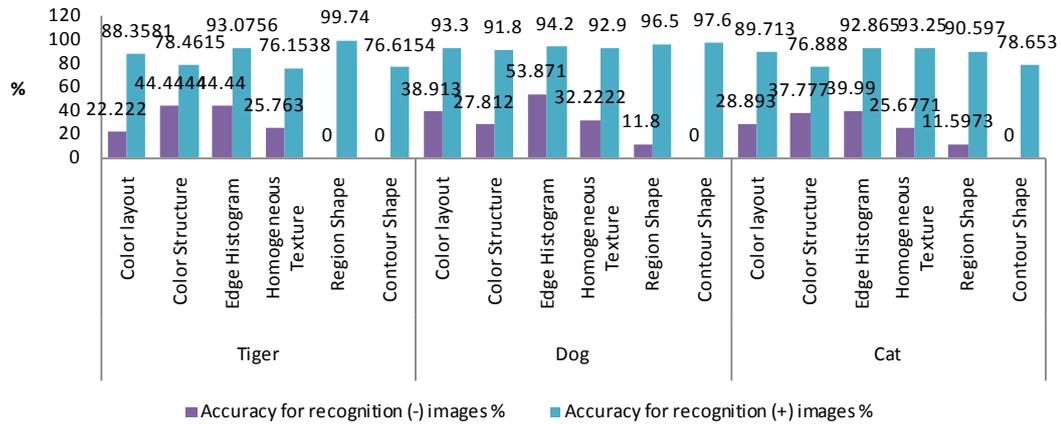
Figure 7. Performances of individual descriptors.
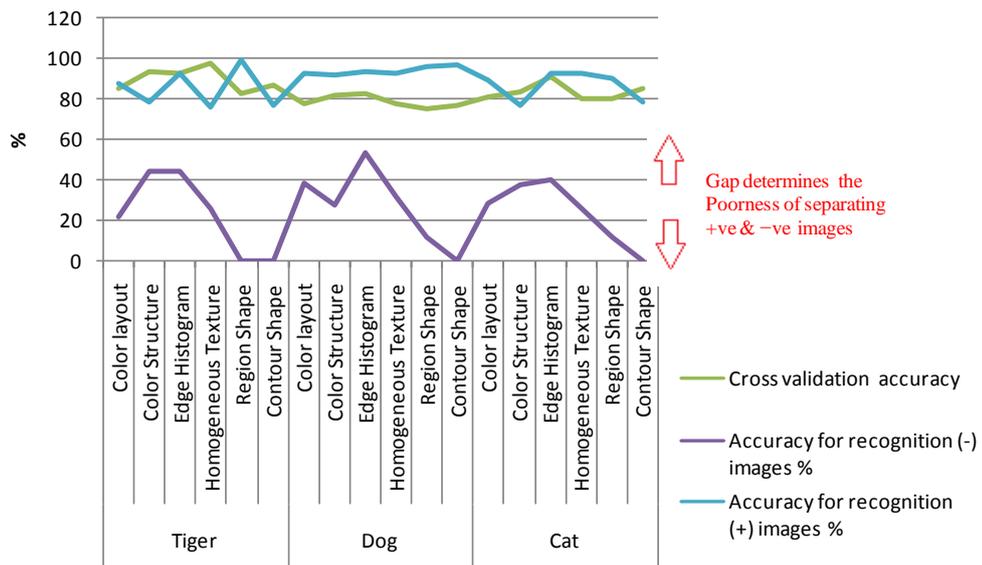
Figure 8. Performances of individual descriptors.

Then by considering above results test was carried out to identify how the output would looks like. The combination was followed by taking most efficient descriptors and all the time concerned to have single color descriptor, single texture descriptor and single shape descriptor together.

According to **Table 2** enhanced results of combined descriptors were evident and outcome had more efficiency than earlier case.

Figure 9 shows the tabulated results of combined descriptors performances.

3. Conclusion

When compared with performance of individual descriptors, recognition of positive & negative image separately was not successful. But when few descriptors combined together, very fair outcome was evident yet accuracy varied around 80%. As per the graph in **Figure 8** which represented the combined descriptor results comparably, very successful results were evident which recognize both positive & negative images fairly in equal manner. The combination results of region shape, edge histogram and color structure had very low separation between positive and negative columns. The combination results of region shape, edge histogram and color layout were given average outcome compared with the above combination. The combination of region shape, edge histogram, color layout & color structure, though it recognizes a fair amount of positive images, fails to separate negative images in a great manner. So mid configuration of both data sets, which is combination of region shape,
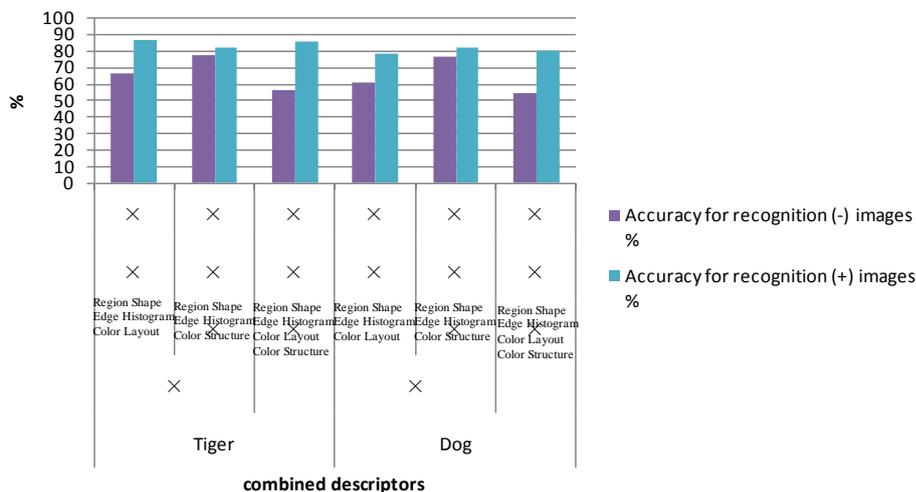
Figure 9. Performances of individual descriptors.

Table 2. Combined descriptor performances.

Animal	Color layout	Color structure	Edge Histogram	Region Shape	C	ɣ	Cross validation accuracy	Accuracy for recognition (−) images %	Accuracy for recognition (+) images %
Tiger	×		×	×	2	0.0312	87.6687	66.1111	86.4103
		×	×	×	2	0.25	90.8589	77.222	82.05
	×	×	×	×	2	0.125	87.2393	55.5556	85.64
Dog	×		×	×	8	2	77.5285	60.913	78.3
		×	×	×	2	0.125	82.2958	75.812	81.82
	×	×	×	×	8	2	82.6916	53.871	80.02

edge histogram & color structure, provided good results in recognition. In a summary compared with individual descriptor performances, combined descriptors provided much accurate outcome.

References

[1] O'Connor, E., *et al.* (2005) The Acetool Box: Low Level Audio Visual Feature Extraction for Retrieval and Classification. 2*nd European Workshop on the Integration of Knowledge*, *Semantics and Digital Media Technology* (*EWIMT* 2005), 30 November-1 December 2005, 55-60. http://dx.doi.org/10.1049/ic.2005.0710

[2] Elder, J.H. and Velisavljević, L. (2009) Cue Dynamics Underlying Rapid Detection of Animals in Natural Scenes. *Journal of Vision*, **9**, 7. http://dx.doi.org/10.1167/9.7.7

[3] Kouda, M., Morimoto, M. and Fujii, K. (2011) A Face Identification Method of Non-Native Animals for Intelligent Trap. *MVA*2011 *IAPR Conference on Machine Vision Applications*, Nara, 13-15 June 2011, 426-429.

[4] Burghardt, T., Ćalić, J. and Thomas, B.T. Tracking Animals in Wildlife Videos Using Face Detection. https://www.cs.bris.ac.uk/Publications/Papers/2000186.pdf

[5] Ramanan, D., Forsyth, D.A. and Barnard, K. (2005) Detecting, Localizing and Recovering Kinematics of Textured Animals. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *CVPR*2005, **2**, 635-642. http://dx.doi.org/10.1109/CVPR.2005.126

[6] Zhang, W.W., Sun, J. and Tang, X.O. (2008) Cat Head Detection—How to Effectively Exploit Shape and Texture Features. *Lecture Notes in Computer Science*, **5305**, 802-816. http://dx.doi.org/10.1007/978-3-540-88693-8_59

[7] Levesque, O. and Bergevin, R. Detection and Identification of Animals Using Stereo Vision. http://homepages.inf.ed.ac.uk/rbf/VAIB10PAPERS/levesquevaib.pdf

[8] Flyonthewall. Ltd. Infrared CCD Wildlife Camera with Embedded DVR.

http://www.flyonthewall.uk.com/infrared-wildlife-camera-dvr-with-motion-sensor-8-0mp.html.

[9]   https://www.servocity.com.

[10]   Chang, C.-C. and Lin, C.J. (2015) LIBSVM—A Library for Support Vector Machines.
       https://www.csie.ntu.edu.tw/~cjlin/libsvm/