

Image Annotation Through Gaming

Lasantha Seneviratne and Ebroul Izquierdo

Multimedia and Vision Lab, Queen Mary, University of London, Mile End Road, E1 4NS, London, UK.
{lasantha.s, ebroul.izquierdo}@elec.qmul.ac.uk

ABSTRACT

We introduce an interactive framework for image understanding, a game that is enjoyable and provide valuable image annotations. When people play the game, they provide useful information about contents of an image. In reality the most accurate method to describe the content of an image is manual labelling. Our approach is to motivate people to label imagers while entertaining themselves. Therefore if this game becomes popular it will be able to annotate most imagers on the web within a couple of months. When considering accuracy we use a combination of computer vision techniques to secure the accuracy of image labelling. By doing this we believe our system will make a significant contribution to address the semantic gap in the computer vision sector.

1 INTRODUCTION

Object recognition and semantic concepts in images is a main research topic in the computer vision sector. There are billions of imagers on the web; retrieving those using high-level semantic concepts is not accurate enough yet. Low-level feature extraction techniques are able to determine the difference and distribution between colours, textures etc; but the gap between low-level features and high-level concepts is an open issue. Over the last decade problems related to semantic gap have driven the research into several directions. The web base ESP game [1] is an “out the box” approach that provides an appealing way to annotate images. The idea behind the game is to label images on the web according to their visual integrity. As mentioned the most precise way to describe the image integrity is manual annotation. Considering billions of imagers are on the web this technique is more costly and impractical.

The main objective of this paper is to present an interactive approach to annotate imagers using manual labelling. In order to reduce the cost of manual annotation we introduce a highly enjoyable framework. When considering the label validity we use different combinations of techniques to increase the accuracy. This includes both psychological and computer vision techniques. When considering the psychological behaviour we use some simple techniques to clarify the user attitude. By doing so we were able to find whether the player is a cheater or not and treat them differently. At the same time we use computer vision techniques to increase the accuracy of labels. The goal is to classify an image according to user key-word quarry and annotate them. In real world applications images may represent a scene that may contain a number of objects. We therefore required an annotation within an object

level rather than consideration of the whole scene. We used a very simple and effective approach called ‘elementary building elements of imagers’ or simply image blocks technique [3]. The advantage of this approach is that it distinguishes the background from an object. Those image blocks are more related to objects thus objects are represented by a number of image blocks. Therefore we use object base image representation techniques in our framework.

This paper is organised as follows, section 2 describes the general view of the system; section 3 describes the performance measure and the paper ends with the conclusions and future works in section 4.

2 GENERAL VIEW AND ARCHITECTURE

We call our system ‘Tag4Fun’; an interactive game designed using a 3D graphic library call OpenGL. This game is designed to be played by a single player and meant to be played by a large number of players. The goal of the system is to annotate imagers according to their contents.

The image that requires annotation will be displayed by the Tag4Fun game. The game player comments on the contents of the image. The basic game structure is similar to a well known game ‘Tetris’. The major difference in Tag4Fun is it uses characters instead of different shapes of building blocks. To speed the annotating process Tag4Fun uses three columns of moving characters. The 3D characters move from top to bottom on the screen, the player is intended to collect them using the keyboard. For interactive purposes, Tag4Fun generates random magic characters which are subject to change to any character. The game player has to construct the key-word related to the contents of the image by collecting individual characters. The collected characters are used to select the pre-trained classifier for image classification purposes thus improving the label accuracy.

The Tag4fun visual game will entertain and motivate the player and provide valuable key-words about what is contained within the image. At the same time it helps to determine the users’ attitude by feeding imagers using 3 different databases called none-annotated, partially-annotated and fully-annotated. The game player will be fed randomly by all 3 databases. Therefore if the game player tries to annotate a fully annotated image using non related key-words the system will identify them as cheaters and treat them only with partially annotated imagers. The key-words generated by those players won’t be used for any labelling.



Figure 2: Tag4Fun game framework

2.1 When is an image annotated?

When the classifier agrees on an image it will be temporally annotated and the player will get a certain number of points to encourage them to continue playing. When an image passes through Tag4Fun it contains a number of possible labels for it. If an image describes using the same label 5 times that keyword will be associated as a taboo word for the image and won't allow players to use the taboo words for further labelling. If an image got 8 taboo words the image will fully annotate and be discarded from the database. All other information captured will be saved for future references. For integrity and language changes over time, a few months later fully annotated imagers will be loaded back for update purposes. (For example, George. W. Bush is the president of the United States and will be the former president in the future).

2.2 Low-level feature extraction

Most image retrieval systems failed to produce satisfactory results especially when a user was interested in a particular object rather than whole scene. We therefore used a simple but effective technique called 'elementary building elements of imagers', often called image blocks. This technique divides the whole image into blocks of imagers. Image blocks represent different types of objects in the image, or combinations of blocks represent a single object. We extracted low level features from each block and classified them manually to create a vocabulary of training sets. The trainee model selection is directly related to the input quarry (Key-word). Using the pre-trained models we were able to classify image blocks; because of the image block concept we find the block related to a particular object. Therefore in the future we will use it as a benefit to help the game player interact by giving them a chance to collect more bonus points when they point to the location of the object. This will allow us to test the user attitude for the second time and provide more valuable information about the location of a particular object.

The low level feature extraction was completed by using three low-level descriptors, dominant colour descriptor (DCD), colour layout descriptor (CLD) and edge histogram descriptor (EHD). Those descriptors are defined and used by MPEG-7 (Moving picture experts group). Using those descriptors we trained a support vector machine (SVM) [4] classifier for

classification; this helps us to increase the accuracy of annotation and in turn minimise cheating.

3 PERFORMANCE MEASURE

These types of games depend on the physiological behaviour of the game player. Therefore it is extraordinarily difficult to measure the performance of Tag4Fun unless it is being played by a large number of users. As an ongoing project Tag4Fun is not yet ready for commercialization. Its performance was analyzed in two different ways. First of all we analyzed the performance of the classification process for three different concepts and secondly we analyzed the performance of the complete frame work. For testing purposes the classifier was trained for three concepts using 10 images, equating to 320 image blocks. The concepts used are butterfly, tree and cougar. The Performance for the three concepts obtained is displayed in table 1 as follows.

Precision	Butterfly	cougar	Tree
CLD	45%	12%	65%
DCD	30%	5%	40%
EHD	45%	12%	40%

Table 1: Performance of the SVM classifier

According to table 1, we conclude the performance of our classifier is not as accurate as expected. Therefore we will keep working until we achieve a satisfactory result. However with such precision of the classifier we managed to get 71% accuracy for the complete framework. (The performance measured by using eight regular game players).

4 CONCLUSIONS AND FUTURE WORK

We introduced a computer game which is able to encourage and motivate game players to annotate imagers manually. The proposed framework was tested with eight regular game players and the performance was acceptable. As an ongoing project we will develop and improve the whole system to achieve high accuracy of labels. We will also improve the quality of our system according to physiological aspects of regular game players.

Future work will mainly focus on techniques for improving the accuracy of annotation process and combining low-level features to improve the accuracy of the classification process.

5 REFERENCES

- [1] Luis von Ahn and Laura Dabbish "Labeling Imagers with a Computer Game" Pittsburgh,PA,USA 2004.
- [2] Luis von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu and Manuel Blum "Improving Accessibility of the Web with a Computer Game" Pittsburgh 2006.
- [3] Qianni Zhang; E.Izquierdo "Optimizing Metrics combining Low-Level Visual Descriptors for Image Annotation and Retrieval" 2006.
- [4] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin "A Practical Guide to Support Vector Classification" 2008.