International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-5, June 2013

Reading Users' Minds from Their Eyes: A Method for Implicit Image Annotation

Arti R.Bhore, N.D.Kale

Abstract - This paper explores the possible solutions for image annotation and retrieval by implicitly monitoring user attention via eye-tracking. Features are extracted from the gaze trajectory of users examining sets of images to provide implicit information on the target template that guides visual attention. Here Gaze Inference System (GIS) is a fuzzy logic based framework that analyzes the gaze-movement features to assign a user interest level (UIL) from 0 to 1 to every image that appeared on the screen. Because some properties of the gaze features are unique for every user, our user adaptive framework builds a new processing system for every new user to achieve higher accuracy. The generated UILs can be used for image annotation purposes; however, the output of this system is not limited as it can be used also for retrieval or other scenarios. The developed framework produces promising and reliable UILs where approximately 53% of target images in the users' minds can be identified by the machine with an error of less than 20% and the top 10% of them with no error. As show in this paper that the existing information in gaze patterns can be employed to improve the machine's judgment of image content by assessment of human interest and attention to the objects inside virtual environments.

Keywords- Eye tracking framework, image annotation, image retrieval, Gaze Interface System.

I. INTRODUCTION

The history of eye-tracking goes back to the 19th century when scientists tried to study the reading process by direct observation. Along with the development of eye tracking equipment, experimental studies in psychology and engineering have taken advantage of this new form of implicit human feedback opting for the day that every screen will have an affordable embedded eye-tracker. In this paper we discuss the use of eye-trackers for image annotation in the field of multimedia and vision to classify images as a function of the target template that guides user visual attention implicitly, promptly and accurately.

By increasing the size of the visual databases specifically in distributed environments (such as social networks like Facebook and image sharing websites like Flickr), the necessity has risen to annotate and organize images with an undemanding, inexpensive and accurate method. There are three approaches for image annotation: In general, an image annotation task consists to assign a set of semantic tags or labels to a novel image based on some models learned from certain training data. Conventional image annotation approaches often attempt to detect semantic concepts with a collection of human labeled training images [8]. There are three approaches for image annotation [9], [10]:

1) Manual such as LabelMe [11]: Accurate but expensive, time consuming and exhaustive.
2) Automatic: Prompt and cheap but accuracy remains an important issue.
3) Semi-automatic: Performed by interaction between human and computer with a higher accuracy than automatic method and cheaper approach than the manual method (the method proposed in this paper).

These tags provide the meaningful description of images. The success of Flickr proves that users are willing to provide this semantic context through manual annotations. Recent users to annotate their photos with the motivation to make them better accessible to the general public. [1]

Automated image annotation is to assign a set of semantic tags or labels to a novel image based on some models learned from certain training data. But it is often expensive and time consuming to collect the training data. If the annotation is automatic [5], Automated image annotation conduct the process by inspection of the low level features of the images [4], classification of them and optimization of the classification results. However, regardless of how well the classification is performed, the semantic gap [6] remains a problem. Kozma*et.al.* [3] introduced GaZIR which is a gaze-based interface for image browsing and search. Now a days, the semi-automatic method is used for image annotation and retrieval with the help of the implicit feedback acquired from eye-trackers. Semi Automatic method performed by the interaction between human and computer with a higher accuracy than automatic method and cheaper approach than manual methods.

In this paper we introduce fuzzy inference based genetic algorithm which is able to assign a user interest level (UIL) score from 0 to 1 to every image that appears on the screen with more accuracy than the previous methods. The UILs used to annotate images by looking at the cluster of the images with high values of UIL.

Manuscript published on 30 June 2013.
***** Correspondence Author (s)
Arti R. Bhore*, Computer Department, Pune University/ TCOER College/ K.J.s Institutes, Pune, India.
Prof.N.D.Kale, Computer Department, Pune University/ PVPIT College/ JSPM Institutes, Pune, India.

Fig. 1. Snapshot of the video images of the eye-tracker: The white glint in the eyes are the reflection of the infra-red source, the green vectors show the gaze direction, and the red vector shows the head direction of the user.

After the fuzzy inference system is trained, it is able to assign a user interest level (UIL) score from 0 to 1 to every image that appears on the screen. UILs to annotate the images that the system has no information about by having some key information about the concept in their mind or by looking at the cluster of the images with high values of UIL.

II. MATH EYE TRACKING

Eye tracking studies have shown that fixations and saccades are the most important eye movement behaviors for the process of revealing user cognition. A fixation happens when the eyes seize on a single location and a saccade is the ballistic eye movement between any two fixations. The main visual processing of stimuli takes place during fixations when the eyes focus on a single location, centering the scene on the fovea for maximum fidelity of sampling the visual information present for information processing and recognition in the brain. In this experiment we used a binocular set of 60-Hz cameras with infra-red filters and the faceLAB 5.0 software package as the eye-tracking technology. In this framework, the first step is that capture video images from cameras. Next find the positions of both eyes are identified in every single image and two glints. This glint is the reflection of an infra-red source by the eyes which is positioned between the cameras. By comparing the position of the glints to the position of the pupils in each video image, the software can estimate the direction vector of the user's gaze.

III. FRAMEWORK

Finally a UIL score assigned to every image appeared on the computer screen and that UIL shows how much interest the user did show to an image during the experiment. Based on these UILs, framework classifies or cluster the images. We used the sets of images appear on the screen. The first 5 pages (called Training Pages) of the experiment are used to adjust the UIL generator engine to the user.

Fig. 2 shows the block diagram of the developed framework. We see in Fig. That image appears on the screen. Then the user's gaze intersection to the trial screen is monitored. User have intersected with any of the two images on the screen, the corresponding gaze coordinates and gaze duration are recorded and sent to the feature extraction unit. In this unit, two different feature vectors are extracted which are called the transition feature vector (TFV) and the image feature vector (IFV). Next if the feature vectors belong to the pages from the training phase (first 5 pages with fully annotated images), they are sent to the model construction unit.

For both of the feature vectors, two independent processing systems are developed (both by fuzzy logic based and neural networks based structures). As the experiment exits the training phase and continues with non-annotated images, the developed processing systems start to interpret the user's gaze movement feature vectors, and the processing system of each vector assign a UIL to the images based on the information in that vector (T-UIL and I-UIL for TFV and IFV, respectively). Finally for every image, the average of the T-UIL and I-UIL values are calculated as the final output of the system. [1]

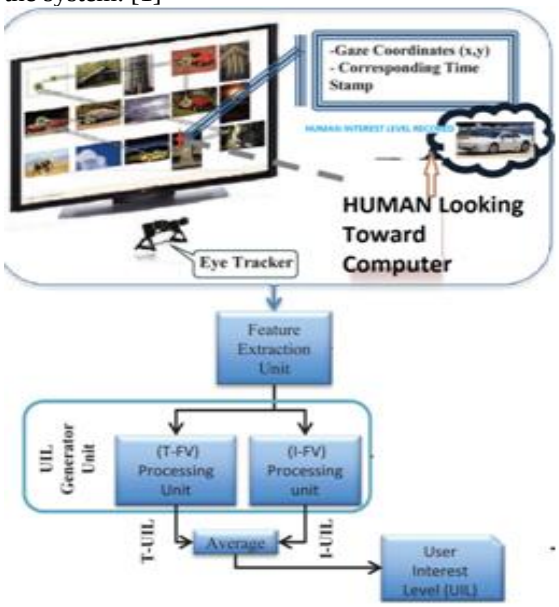


Fig. 2. Block diagram of the developed framework.

A. User Interface

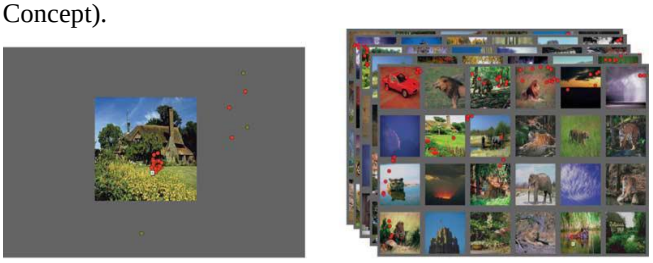
In all of the scenarios, the images appeared slightly darker on the screen, and in case of gaze intersection, they turned into their normal brightness. For example, In Google Images, When we move the mouse on Image, that Image becomes dark and come slightly front to the user. We can conclude that User more concentrate on this image.

B. Clarified User Directed (CUD) Scenario

This scenario is when a user has a specific concept, target concept (TC). CUD classifies these images into two classes, either Favored by the user or Not-Favored by the user based on his/her eye movement attention. [1] In this scenario, every time that the user clicks on a provided Next button or clicks on an image while looking at it, he/she is provided with the next page of 24 images. Each image appears only once during the experiment and it is chosen randomly from the database. For performance measurement of the framework, the users were told to imagine that they are responsible for selecting an image for the cover of a magazine from the images that appear on the screen.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

This image had to contain the same concept as a randomly selected image that appeared at the beginning of the experiment in the start page. The key concept of this image considered to be the TC in user's mind. The start page with the "garden of a house" key concept as the TC(Targeted Concept).



(a) (b)

Fig. 3. Framework scenario screen-shots: The image of the start page is regarded as the target concept in the user's mind. (a) Start Page. (b) Scenario.

IV. FEATUREEXTRACTION UNIT

The Gaze Coordinate, fixation duration (that's how much time human stay on every page of the scenario) and ID are the output of Eye tracker. This outputs nothing but inputs for FEU. These three factors are sent to the FEU after the user finished exploring the page and requested a new page.
We extracted two different feature vectors where one of them contains the properties of eye-movements and the other one contains information about the properties of gaze intersection with the image.[1]

A. Feature Evaluation

The output of developed framework, we tested the training data of every user by calculating the mutual information between the samples of every feature and their corresponding class variable, where the class variable shows the real state of an image with regards to the TC (e.g., if the TC is the tiger and the image contains "tiger" as a concept, then its class variable is 1; otherwise, it is 0).

V. MODEL CONSTRUCTION UNIT

This unit is responsible for constructing two fuzzy logic based systems for the two feature vectors. By using the training (input: output) pairs for the training pages of the experiment where:
1. The "input" is formed of the extracted feature vectors for every image and the corresponding transition that they belong to.
2. The "output" is determined as and if the image belongs to TC or TC classes, respectively, and is known as the class variable (CV).

VI. CONCLUSION

In this paper, the key for image annotation and retrieval by monitoring user attention via eye-tracking is proposed. Here, the features are extracted from the gaze trajectory of users to provide implicit information.

A fuzzy based adaptive framework is capable of measuring the interest of the users to images that appear on the screen by tracking their eyes. This framework assigns a UIL score to every observed image that can be used for image annotation. This framework was able to be trained for every user and produce results with high accuracy at an acceptable rate. The chosen gaze features and the form of output of the framework make it flexible where by some trivial changes it can be used for retrieval purpose and measurement of the user's interest in other forms of visual objects on screen.

REFERENCES

[1] L. Kozma, A. Klami, and S. Kaski, "*Gazir: Gaze-based zooming interface for image retrieval*" in Proc. 2009 Int. Conf. Multimodal Interfaces, New York, 2009, pp. 305–312, ser. ICMI-MLMI'09, ACM.
[2] SHYNI.G1,SHANMUGAPRIYA.K2 "*An Adaptive Fuzzy Logic Based Technique To Read User's Mind for Image Annotation*" (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3, March 2013)
[3] M. Ames and M. Naaman*," Why we tag: motivations for annotation in mobile and online media"* In Proc. of CHI '07, pages 971–980, SanJose, California, USA, 2007.
[4] R. Shi, H. Feng, T.-S. Chua and C.-H Lee, "*An adaptive image content representation and segmentation approach to automatic image annotation in Image and Video Retrieval*" ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2004, vol. 3115, pp. 1951–1951.
[5] H.Ma, J. Zhu, M.-T. Lyu, and I. King, "*Bridging the semantic gap between image contents and tags"* IEEE Trans. Multimedia, vol.12,no.5, pp.462–473,2010.
[6] J. Fan, Y. Gao, H. Luo, and R. Jain, *"Mining multilevel image semantics via hierarchical classification*" IEEE Transactions on Multimedia, 10 (2):167–187, 2008.
[7] A. L. Yarbus, "*Eye Movements and Vision*" New York: Plenum,1967, translated from Russian by Basil Haig,, Original Russian edition published in Moscow in 1965..
[8] "*Gazir: Gaze-based Zooming Interface for Image Retrieval*" by laszlokozma, Artokl, Samuelkaski, Department of Information and Computer science, Helsinki University of Technology Finland.J. Jones. (1991, May 10). Networks (2nd ed.)

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)