



INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI  
SHORT ABSTRACT OF THESIS

Name of the Student : **Vidyapu Sandeep**

Roll Number : **136201004**

Programme of Study : **Dual Degree (M.Tech. + Ph.D.)**

Thesis Title:  
**Computational Modeling of Free-viewing Attention on Multimodal Webpages - A Machine Learning Approach**

Name of Thesis Supervisors : **Dr. Vijaya Saradhi Vedula &  
Dr. Samit Bhattacharya**

Thesis Submitted to the Department : **Computer Science and Engineering (CSE)**

Date of completion of Thesis Viva-Voce Exam : **24-August-2020**

Key words for description of Thesis Work : **Attention modeling, Eye-tracking, Machine Learning, Webpages**

---

**SHORT ABSTRACT**

With the progressive expansion of competitive e-commerce and Web resources, attention modeling is essential for Web authors, information creators, advertisers, and Web-designers to understand and predict the user attention on webpages. State-of-the-art models often overlook the design-oriented visual features of constituent web elements, including text and images. The **bottleneck** was to incorporate the elements' heterogeneous features into the model as texts are represented using features such as `text-size' and `text-color' whereas images are represented using `brightness', `intensity' and `color histograms'. This thesis work is predominantly centered around overcoming the *heterogeneity bottleneck* to predict the user's *free-viewing* attention on multi-modal webpages, precisely consisting of text and image modalities.

Owing to the prominence of position, primarily, the **position**-based free-viewing attention allocation is investigated and computationally modeled, separately for text and image elements. The analyses revealed: (i) the elements positioned in the *Right* and *Bottom* regions of a webpage are not always ignored; (ii) *Space*-related (*column-gap*, *line-height*, *padding*) and *font Size*-related (*font-size*, *font-weight*) intrinsic text features, and Mid-level Color Histogram intrinsic image features are informative, while *position* and *size* are informative for both the types; (iii) the informative visual features predict the ordinal visual attention on an element with 90% average accuracy and 70% micro-F1 score; (iv) For the prominent images, the visual features also help in predicting the weighted-voting-based, kernel-based, and multiple-levels of user attention.

Leveraging the prominence of web elements' visual features, Canonical Correlation Analysis (CCA) based computational approach is proposed to unify both the modalities and to predict the user attention at the granularity of web elements as well as webpages. The results reveal: (i) text and images are unifiable if the *interface* idiosyncrasies alone or along with *user* idiosyncrasies are constrained; (ii) The *font-families* of text are as influential and comparable to image *color histogram* visual features in achieving the unification. The achieved unification also outperforms the random baseline in predicting the user attention on individual web elements as well as overall webpages. This thesis work finds applications in user attention prediction, web-designing, and user-oriented webpage rendering.