**VideoSense: A Contextual In-Video Advertising System, IEEE , 2009**

This paper forms the baseline of modern in-video advertising solutions. With a novel approach to insert ad-videos within original videos based on content discontinuity and content attractiveness and relativeness, this paper also achieves remarkable results during the experimentation setup.

The paper claims that ad-insertion is a non-linear integer programming problem and this is something we agree with. It highlights the differences between object-level advertising and segment-level advertising and proposes solutions to tackle them both. Our key takeaway is the approach taken for object level advertising wherein images or products or brands to be advertised are placed on static non-contributing objects in frames within videos(in our case, images).

The approach taken is to first identify insertion points in videos with respect to lesser intrusiveness and attractiveness. The context of the video is analysed through several different methods, including analysing the surroundings of the webpage containing the source video. Contextual relevance is also analysed based on local visual-aural relevance between insertion points. The extracted context is then compared with the context of said given ad-videos and the most appropriate one is chosen for in-video advertising. This forms the basis for the segment level advertising approach. In object-level advertising, at the insertion-point, the frame of the video is taken into consideration and the least desirable object is overlayed with the image of the ad that matches the context.

The paper also proposed a novel ‘multi-modal’ approach for evaluation. The evaluation was set up for different aspects of the proposed solution. For instance, different evaluation strategies were used for measuring the contextual relevance, measuring the content intrusiveness at the different insertion points, measuring the relevance of the inserted ads and so on.

The limitation is this paper proposes object-level advertising as simply an extension to their original solution and does no justice in highlighting the technical details for the same. They also claim to need better matching between ads and videos based on context.

**CAVVA: Computational Affective Video-in-Video Advertising , IEEE, 2014**

This paper highlights the drawback of traditional outdated ad-insertion methods and the need for improved approaches to tackle the advertising solution. This paper builds on the idea ad approaches suggested by Videosense. They claim to identify insertion points in videos based on genetic algorithms capable of identifying the emotional state of the user viewing the video. They also claim that with their fine-tuned model; they obtain a good balance between conflicting goals of minimizing user disturbance due to ad-insertion while enhancing the user engagement with the ad itself. They also propose a novel solution of eye-tracking during the process of watching videos to identify points of distraction and low-interest of the user towards the source video.

Aside from that, this paper also proposes certain parameters to be included while developing the recommendation system to give better results for the task at hand. The experimental evaluation setup as explained in this paper is also pretty unique and worth considering. They analyse the significance of their approach by measuring subjective user experience and the ability for a user to recall a brand later on.

While this paper builds upon the baseline Videosense paper, it still does not provide us with a concrete solution for object-level advertising. The eye-tracking approach is also something that isn’t feasible.

**Embedding user preferred advertisement in video playback in ecommerce, International Journal of Pure and Applied Mathematics , 2018**

The key-takeaway from this paper is ad-insertion techniques and benefits of such techniques from an ecommerce perspective. The implementation and the model proposed by them is a real-time framework that is based on mining data based on user’s behaviour on ecommerce websites.

Although this approach is widespread and can be seen in implementation already with the larger ecommerce companies, this paper gives an idea of how such an approach can be implemented for our project.

**Context-Dependent Sentiment Analysis in User-Generated Videos, Association for computer linguistics, 2017**

This paper talks about multimodal sentiment analysis in videos. They talk about the utilization of a LSTM- based model that enables to capture contextual information from surroundings. This is beneficial to us to understand the context of an image based on objects within the image and surrounding text(if any). They claim to have 5-10% performance improvement over state of the art and high robustness to generalizability. The key takeaways of this paper are the multimodal approach and the characteristics of the 3d-CNN(Visual feature extraction) and text-CNN they have implemented for context analysis.

The paper reports the performance of the different type of LSTM architectures available. The best performing model is the bc-LSTM model scoring the highest in each test case as generated by the authors.

**Patent : INTERACTIVE PRODUCT PLACEMENT SYSTEM AND METHOD THEREFOR**

The patent describes the process of static ad placement techniques in videos at pre-defined cue points. These ads are mostly in the format of images that are similar to the Video object level advertising approach. This patent talks about placement of such images at these cue-points without having any contextual analysis method. These ads however are capable of rerouting users to other landing pages where these products are being sold or brand marketing pages. They take away from this patent is the cue point approach. However, the major difference is while these cue points are static and are pre-defined, out approach involves these cue points being dynamic. The images inserted as ads itself shall be recommended based on a certain degree of contextual analysis.