

ONLINE MULTIMEDIA ADVERTISING

YADATI NARASIMHA KARTHIK (A0069129)

**A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE**

**DEPARTMENT OF COMPUTER SCIENCE,
SCHOOL OF COMPUTING,
NATIONAL UNIVERSITY OF SINGAPORE
2013**

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Yadati
Yadati Narasimha Karthik

21/10/2013
A0069129

Abstract

Past few years has seen a tremendous explosion in the availability of video data on the internet. A major reason for such an explosion is the rise of community video sharing websites like YouTube and the growing popularity of various social networks like Facebook. With the rise of such sources of video content, the number of users participating in such forums in different capacities as audience and content creators has also increased. Presence of a large number of people in such forums (for example, more than one billion users visit YouTube each month) provides a lucrative opportunity to advertisers in order to market their product/service. This thesis concentrates on providing a wholesome framework for computational video advertising.

Commercial video advertising strategies, like the ones available on YouTube, do not perform audio-visual content analysis for placing advertisements (examples include pre-roll/mid-roll/post-roll advertising). Contextual advertising has been studied in videos from a semantics perspective where a sports related video would have sports related advertisements. An important aspect, which has been completely ignored in current video advertising, is the emotional impact of the video and the advertisement on the user. In this thesis, we use the psychological theories on emotion and apply them to two broad areas of advertising: *content-based advertising* and *personalized advertising*.

As part of *content-based advertising*, we tackle the problems of video-in-video advertising, overlay advertising (image overlay on videos) and companion advertising (image advertisements at the side of the video). We propose and implement a scalable mathematical framework based on psychological theories on emotion. We employ a 0-1 Non-linear Integer Programming (NIP) framework to formulate the problem and then propose a genetic algorithm based solution. We compare our advertising strategies with commercial advertising strategies, like the ones present on Youtube: Pre-roll/post-roll advertising and also state-of-the-art contextual advertising. Through systematic experiments, we demonstrate better results than the existing methods in terms of user experience and assimilation of advertising content.

Personalized advertising, also known as targeted advertising, is prevalent in textual advertising where users are tracked using cookies on their computers. Personalized advertising, using the various sensors to observe the user, is still in its infant stages. We propose and implement a *personalized video-in-video advertising* strategy which takes into account, the user's emotional state to place in-stream advertisements dynamically. We demonstrate the effectiveness of the proposed advertisement placement strategy using different experiments.

Contents

1	Introduction	6
2	Related work	17
2.1	Contextual advertising	18
2.2	Emotion in advertising	26
2.3	Personalized advertising	27
2.4	Previous work	30
3	CAVVA: Computational Affective Video-in-Video Advertising	31
3.1	What to expect?	31
3.2	Background	31
3.3	Proposed Method	36
3.3.1	Step 1: Input video and advertisements	37
3.3.2	Step 2: Scene change detection	37
3.3.3	Step 3: Affective video analysis	38
3.3.4	Step 4: Optimization framework	38
3.3.5	Step 5: Output video	39
3.4	Problem Formulation	39
3.4.1	Efficiency and Quality of CAVVA	41
3.5	Experiments	43
3.5.1	Data Collection	44
3.5.2	User-study	46
3.5.3	Advertisement/Brand recall	47
3.5.4	Eye-tracking experiment: Measuring pupillary dilation as a proxy for arousal	48
3.5.5	Ground truth data	49
3.6	Results and Discussion	50
3.6.1	Subjective user experience	51
3.6.2	Advertisement/Brand recall	53
3.6.3	Eye-tracking experiment	55
3.7	Support for overlay advertising	57
3.8	Summary	59
4	Companion advertising	61
4.1	What to expect?	61
4.2	Background	61
4.2.1	Mood congruency effect	62
4.2.2	Relation between arousal and memory	63
4.3	Proposed approach	64
4.4	Problem formulation	65

4.5	Experiments	68
4.5.1	Data Collection	69
4.5.2	User-study	69
4.5.3	Advertisement/Brand recall	70
4.5.4	Eye-tracking experiment: Measuring attention related features	70
4.6	Results and Discussion	70
4.6.1	User-study	71
4.6.2	Advertisement/Brand recall	72
4.6.3	Eye-tracking experiment	72
4.7	Summary	73
5	Personalized video-in-video advertising	75
5.1	What to expect?	75
5.2	Background	75
5.3	Feature selection and fusion	78
5.4	Multimodal fusion	80
5.5	Personalized online advertising framework	81
5.6	Experimental setup	83
5.7	Evaluation and User studies	83
5.7.1	Data Collection	85
5.7.2	Experimental design	85
5.8	Results and discussion	87
5.8.1	Brand and Advertisement recall	87
5.8.2	Subjective experience	88
5.8.3	Impact on long-term recall	89
5.9	Summary	89
6	Summary and Conclusion	90
6.1	Summary	90
6.2	Contributions	91
6.3	Future work	92

List of Figures

1	An example of successful contextual advertising.	8
2	Examples of unsuccessful contextual advertising.	9
3	Example of contextual video advertising in which a person is speaking on phone and a phone related advertisement is placed in between.	10
4	Example of pre-roll/post-roll advertising with advertisements at the beginning and at the end of the video.	10
5	Example of in-stream video-in-video advertisement.	13
6	Example of image advertisement overlaid on a video frame.	13
7	Example of companion advertisement.	14
8	Example of Pre-roll advertisement on YouTube	18
9	Example of Mid-roll advertisement on YouTube	19
10	Example of overlay advertisement on YouTube	19
11	Example of Companion advertisement on YouTube	20
12	Example of advertisement insertion using AdImage [LCH08]	21
13	Architecture for AdOn	24
14	Example of (a) Pleasant advertisement and (b) Unpleasant advertisement from our dataset	32
15	Example of (a) Pleasant print advertisement and (b) Unpleasant print advertisement from our dataset	33
16	Visualizations of transition in valence. (a) A transition from low valence to high valence through the advertisement, indicating the initial inertia to come out of the negative mood. (b) Maintaining a high valence before and after the advertisement.	35
17	Affect based advertising strategy - CAVVA. (1) Input video, (2) Scene change detection, (3) Affective analysis, (4) Optimization framework, (5) Output video.	36
18	Experimental setup for the eye-tracking experiment, which involves a user (1) watching the video on a monitor (3) and the eye-tracker (2) observing the user.	49
19	Self-Assessment Manikin [BL94], used to obtain ground-truth valence, arousal data for the videos and the advertisements.	50
20	Frames from the result of applying the three different advertisement insertion strategies - PRPR (row 1), VideoSense (row 2), CAVVA (row 3), on an example video. The graph plots the valence, arousal scores for CAVVA(row 4).	51
21	Visualization for transition in valence for 15 randomly chosen advertisement insertion points.	52

22	Average ratings to (1) Uniform distribution of videos, (2) Disturbance to the flow of the video, (3) Relevance of the advertisement and (4) Overall viewing experience for each of the three different advertising strategies.	52
23	Average ratings for the most liked video (left) and the most disliked video (right).	54
24	Average immediate recall (left) measured as - (1) Uncued advertisement recall, (2) Uncued brand recall, (3) Cued advertisement recall and (4) Cued brand recall for the three different advertising strategies. Average day-after recall (right) measured as (1) Cued advertisement recall and (2) Cued brand recall.	55
25	Average pupillary dilation (arousal) during advertisements.	57
26	Block diagram for the proposed companion advertising strategy. . .	65
27	An example of a video with 4 associated banner advertisements . .	71
28	Valence-Arousal plot for (i) pre-roll, (iii) post-roll and (ii) VideoSense [MHYL07].	77
29	Affective, online advertisement insertion in <i>MyAds</i>	77
30	Personalized online advertisement insertion.	81
31	Experimental setup. (a) The user; (b) Eye-tracker; (c) Camera; (d) Stimulus monitor	83
32	Example advertisement insertion and selection - Three strategies . .	84

List of Tables

1	Comparison between state-of-the-art contextual advertising and our advertising strategy	25
2	Comparison of VideoSense and CAVVA	26
3	Comparison of previous work and current thesis	30
4	Variables used in the optimization framework	42
5	Video Data Used for Experiments.	45
6	Advertisements used in the experiments	45
7	Results from the two-sample Kolmogorov-Smirnov test for the four subjective questions: Q1-Uniform distribution of advertisements, Q2-Disturbance to the program flow, Q3-Relevance of the advertisement, Q4-Overall viewing experience	53
8	Normalized average number of fixations for Mode I (Random ad placement), Mode II (VideoSense) and Mode III (CAVVA extended to overlay advertising)	59
9	Variables used in the optimization framework	67
10	Demographic details of the participants.	69
11	Average ratings for Q1 - number of advertisements, Q2 - changing advertisements disturb the flow of the video, Q3 - relevance of advertisements, Q4 - overall viewing experience. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR	71
12	Average immediate recall values for three advertising strategies. UAR - Uncued Advertisement Recall, UBR - Uncued Brand Recall, CAR - Cued Advertisement Recall and CBR - Cued Brand Recall. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR	72
13	Average day-after recall values for three advertising strategies. LAR - Long-term Advertisement Recall, LBR - Long-term Brand Recall. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR	73
14	Average fixation frequencies for advertisement and brand across three advertising strategies. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR	73
15	Cued and uncued recall over brands and advertisement content	88
16	Subjective user-responses on a 5-point scale	88
17	Long-term (Day-after) Advertisement/Brand recall.	89

1 Introduction

Advertising is, by definition, "the act or practice of calling public attention to one's product, service, need, etc., especially by paid announcements in newspapers and magazines, over radio or television, on billboards, etc.". Advertising has been existent for a very long time and its history can be traced down to the Egyptian civilization where people used to advertise on papyrus for goods and services. Advertising has evolved through the evolution of the different types of media. Evolution of advertising has been parallel to the evolution of media. For example, invention of print media started the print media advertising and growing popularity of television also popularized television advertising, which is still a profitable option for advertisers. Conventional video broadcast (ex: Television) typically involves large spending from advertisers and professional editing for advertisement placement in the program content. Online video advertising, on the other hand, has content providers very often from amongst the viewer community and the sheer volume of uploaded video rules out any possibility of manual editing or advertisement selection and insertion. Furthermore, viewers and often program content uploaders pay nominal amounts to access the video distribution service. Our target in this thesis is to cater to such online video content and explore the placement of different types of advertisement formats.

There are three major players in any advertising scenario - the user, the advertiser and the content provider. The goals of each of these players are different and their needs are conflicting. The goal of the user is to maximize his/her engagement with the video content with minimal disturbance because of advertisements. If the advertisements are to be placed, the advertisements should be meaningful with respect

to the content of the video as well the user's current needs. Though advertising might be annoying from a user's perspective, it is one of the major reasons which makes free hosting and access to such video content possible. From an advertiser's perspective, it is a completely different challenge as the advertiser would want the user to notice the advertisement and also be able to remember it at a later point in time to enable a growth in the sales of the product/service. A successful advertising strategy should address the challenges posed by the conflicting needs of the user and the advertiser listed below.

1. Advertisement placement should result in minimal disturbance for the user and
2. The placement of advertisements should result in an increased viewer engagement.

Online advertising has evolved from random placement of advertisements to placing contextually relevant advertisements on the web pages. Contextual advertising further led to the development of semantic targeting, where the server analyzes the meaning of the keywords in the context of the entire web page before deciding on the type of advertisements to be placed. Figure 1 demonstrates an example of contextual advertising where the website of a popular news paper displaying advertisements based on the keyword "South Africa". The highlighted advertisements for travel packages to South Africa is displayed on a web page which talks about the latest news in south africa. Serving advertisements purely based on keywords can also lead to a few problems because of multiple meanings associated with a word. Figure 2 highlights a few problems in contextual advertising based on keywords alone. For example, the figure on the right side shows the title of the article



Figure 1: An example of successful contextual advertising.

with Steve Jobs' name and an advertisement related to a job is served which is not relevant to the article. All these developments have been taking place in textual advertising, where a keyword is the main component on which the advertising is based. An excellent example of such contextual advertising is Google's AdSense network, which serves contextual advertising based on another Google program called AdWords. AdWords indexes and identifies important keywords on a web page and lets advertisers bid on the keywords relevant to them.

Exponential growth in the availability of public online digital video collections has given users the flexibility to watch a video of his/her choice at any time. There is an ever increasing viewer base developing for such online video collections. Watching online videos is now a mainstream activity with 78% of people watching online videos at least once a week and 55% watching everyday. Cisco expects video to account for 57% of consumer internet traffic by 2015, nearly four times as much as regular web browsing and email ¹. These video collections cater to a variety of

¹www.youtube.com/yt/press/statistics.html



Figure 2: Examples of unsuccessful contextual advertising.

geographic and topic-wise user groups [BSW12]. As a result, video sharing websites are becoming valuable resources for people to sharing not only information, but also life experiences [BSW12] and in turn becoming lucrative markets for job advertisement of products and services.

Recent explosion of online video content calls for similar forms of intelligent advertising by exploiting the richness of information available in the audio-visual content. Though there is a lot of information to be exploited in videos, online video advertising is still in the nascent stages with simple extensions from textual advertising. An example of contextual advertising is shown in Figure 3 where a person is talking on phone and a related advertisement is placed. Such contextual advertising for videos still analyzes keywords on the web page where the video is embedded instead of the audio-visual content. Analyzing the audio-visual content for contextual video-in-video advertising has received little attention in the past and we mention a few works, in the chapter on related work, which exploit the audio-visual

features to place contextually relevant advertisements in videos. Another exam-



Figure 3: Example of contextual video advertising in which a person is speaking on phone and a phone related advertisement is placed in between.

ple of a common form of video advertising on popular video sharing websites like YouTube is called pre-roll/post-roll advertising, where advertisements are inserted at the beginning or at the end of the video. Figure 4 illustrates an example of such advertising, where we find advertisements at the beginning and at the end of the video. Insights from psychology [Ple05] suggest that the process of decision-making

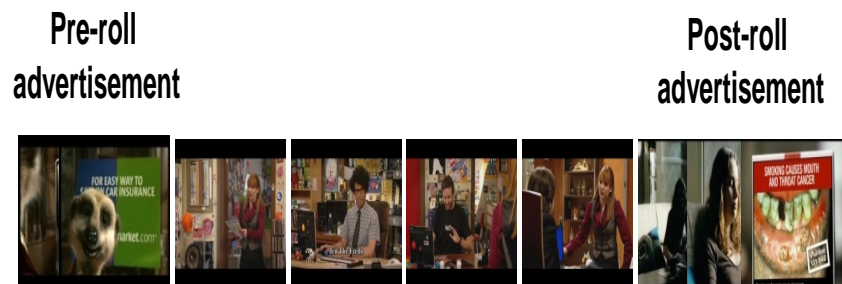


Figure 4: Example of pre-roll/post-roll advertising with advertisements at the beginning and at the end of the video.

is not just rational, but also emotional. Emotions play a major role in influencing human decision processes. Such insights have resulted in a change of mode in

advertising from a rational form, in which facts regarding the products are told to the consumer, to emotive forms which try to evoke an emotion in the consumer and make the product more compelling. Affect induced by an advertisement is considered as one of the important factors in the success of advertising campaign. Various tests have been designed, in the marketing literature, which study the relation between the affective impact of the program and the affect induced by the advertisement. The experience of an emotion is termed as affect and it can be measured in a discrete or a continuous space. One of the ways to represent affect is the circumplex model [Rus80] of affect which is a dimensional representation, where affect is measured in two dimensions - arousal, referring to the intensity of the emotion and valence, referring to the type of emotion. We choose this continuous representation of affect as it is more appropriate in the context of video analysis. Inspired from these experiments, we have constructed an automatic advertisement placement mechanism which takes into account the affective impact of the video as well as the advertisement. In addition, we also propose a personalized video advertising framework which takes into account the user's emotional state before placing advertisements dynamically in the videos.

Many existing video-oriented sites, such as YouTube, Yahoo! Video, MSN Video have tried to provide effective video advertising services. However, it is likely that most of them match the ads with videos only based on textual information and insert ads at the fixed positions, e.g., the beginning or the end of a video. Typical examples for textual relevance matching are the keyword-targeted (e.g., Google's AdWord [MSVV05]) and content-targeted advertising (e.g., Google's AdSense). In other words, contextual relevance in these sites is only based on textual information, while less intrusive insertion points are typically fixed to the beginning or the

end of videos. The following issues are important in designing an intelligent video advertising strategy based on the audio-visual content analysis.

1. Currently, advertisements are inserted at pre-defined positions which are generally at the beginning or at the end of the video. Context plays an important role in determining the effect an advertisement would have on a user and current video advertising strategies do not exploit the knowledge of context which can be obtained from the audio-visual streams of the video. In order to find appropriate points in a video, where we can place advertisements, analyzing the audio-visual content is important. For example, we can decide to place advertisements only at scene change points where there is discontinuity in the audio-visual data is expected.
2. Selection of advertisements should not be random and should be related to the surrounding video. For example, we propose a video-in-video advertising strategy where we select an advertisement which has a similar emotional tone as the preceding scene.

In this thesis, we explore different advertising strategies viz. in-stream video-in-video advertising, overlay advertising, companion advertising and personalized advertising. We give a brief description each of the advertisement formats here.

1. In-stream video-in-video advertising (Figure 5): This form of advertising is more in line with the traditional TV advertising, where the program stops playing and an advertisement is played before the program starts again. We provide a similar mechanism for online videos, where we insert video advertisements into the video stream and hence the name .

2. Overlay advertising (Figure 6): In this form of advertising, image advertisements are overlaid on a part of the video, usually the four corners of the video in order to minimize user disturbance.
3. Companion advertising (Figure 7): Image advertisements are placed at the side of the video. For example, YouTube places companion advertisements on the right side of the video playing area.
4. Personalized advertising: This paradigm of advertising takes into account the users' preference before placing advertisements in the video. For example, we take into account the user's emotional state and insert appropriate advertisements into the video.



Figure 5: Example of in-stream video-in-video advertisement.



Figure 6: Example of image advertisement overlaid on a video frame.



Figure 7: Example of companion advertisement.

The advertisement formats are different for different strategies, where we use video advertisements for video-in-video advertising and banner advertisements for the other two advertising strategies. An important difference between the existing computational video advertising strategies and the proposed advertising strategies in this thesis is that our methods are affect-based and the existing methods are based on semantics [MHYL07]. We highlight the contributions of the thesis as follows:

1. We provide an automatic video-in-video advertisement insertion system in an offline optimization framework - Computational Affective Video-in-Video Advertising, which performs better than the state-of-the-art advertising strategies in terms of subjective user experience and sustaining user interest in the advertisements. An extension is provided to tackle the problem of overlay advertising.
2. A personalized advertisement insertion system which takes into account the emotional state of the user to place appropriate advertisements in video dynamically. Performance of the method is compared to state-of-the-art offline advertising strategies and is demonstrated to perform better.

3. A companion advertising strategy based on an offline optimization function derived from experimental results in consumer psychology. Performance of the strategy is shown to be better than the existing companion advertising strategies.

As highlighted in the contributions, we address different aspects of online multimedia advertising and we provide a detailed organization of the thesis here:

1. Chapter 2 highlights existing literature in advertising giving details about the importance of emotion in advertising, current practices in online video advertising and also a review of the state-of-the-art contextual advertising strategies.
2. Chapter 3 introduces a mathematical framework, based on experimental results from consumer psychology, to insert video advertisements in online videos in an in-stream manner. We propose video-in-video advertising strategy and compare it with existing video advertising practices, contextual advertising and demonstrate its effectiveness through a series of systematically designed experiments. The work reported in this chapter has been accepted for publication [[YKK13a](#)].
3. Chapter 4 introduces another optimization function for companion advertising, where a banner advertisement is associated with each scene of the video based on the emotion induced by the scene as well as the advertisement. Since we are not halting the video to play an advertisement, the disturbance caused by the advertisement is minimal and we focus on maximizing the long-term recall of the advertisements. We design user-study and eye-tracking experiments to demonstrate the effectiveness of the advertising strategy.

4. Chapter 5 introduces the paradigm of personalized advertising, where we take into account the user's emotional state in addition to the emotion induced by the video to place advertisements in the video on-the-fly. Detailed experimental results demonstrate that the personalized advertising performs better when compared to the state-of-the-art video advertising strategies. The work reported in this chapter has been published in a conference [[YKK13b](#)]
5. Chapter 6 provides a summary of the thesis and provides possible future directions.

2 Related work

YouTube is a rapidly expanding community video sharing website which has been on the rise since its inception in the year 2007. As per the latest statistics, more than 1 billion users visit YouTube each month with over 100 hours of video being uploaded every hour and 6 billion hours of video being watched every month. According to Nielsen, YouTube reaches more US adults ages 18-34 than any cable network ². These facts provide lucrative opportunities for advertisers to reach a wide range of audience. As a result, advertisements seen on YouTube has been increasing rapidly over the past few years. Various types of advertising have been explored, some of which are listed here:

1. Pre-roll/Post-roll advertising: In this strategy, advertisements are placed before beginning the video or at the end of the video with an option to skip the advertisement after a pre-determined time. These advertisements are video advertisements and the time interval after which the advertisement can be skipped is currently 5 seconds on YouTube (Figure 8).
2. Mid-roll advertising: Advertisements are placed at random points within the video, where the actual video stops and a video advertisement is played before the video starts playing again. A similar time interval is allowed for skipping the advertisement (Figure 9).
3. Overlay advertising: These are textual/banner advertisements which are placed over the video, generally at the bottom portion of the video player. There is an option given to the user if he/she wants to hide the advertisement (Figure 10).

²www.youtube.com/yt/press/statistics.html

4. Companion advertising: These are banner advertisements which are placed on the right side of the player and remain static throughout the playing time of the video. There is no option to hide this type of advertisements, as they do not obstruct any part of the area where the video is being played (Figure 11).

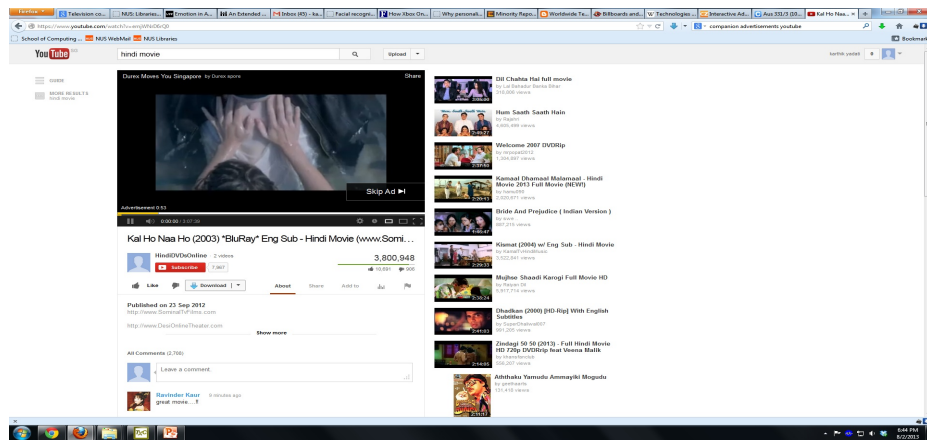


Figure 8: Example of Pre-roll advertisement on YouTube

2.1 Contextual advertising

A contextual advertising system scans the text of a website for keywords and returns advertisements to the web page based on those keywords. The advertisements may be displayed on the web page or as pop-up advertisements. For example, if the user is viewing a website pertaining to sports and that website uses contextual advertising, the user may see advertisements for sports-related companies, such as memorabilia dealers or ticket sellers. Contextual advertising is also used by search

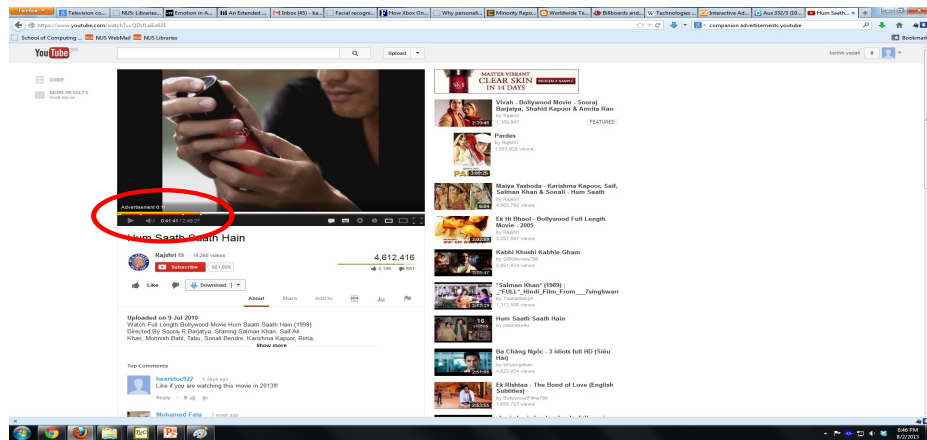


Figure 9: Example of Mid-roll advertisement on YouTube

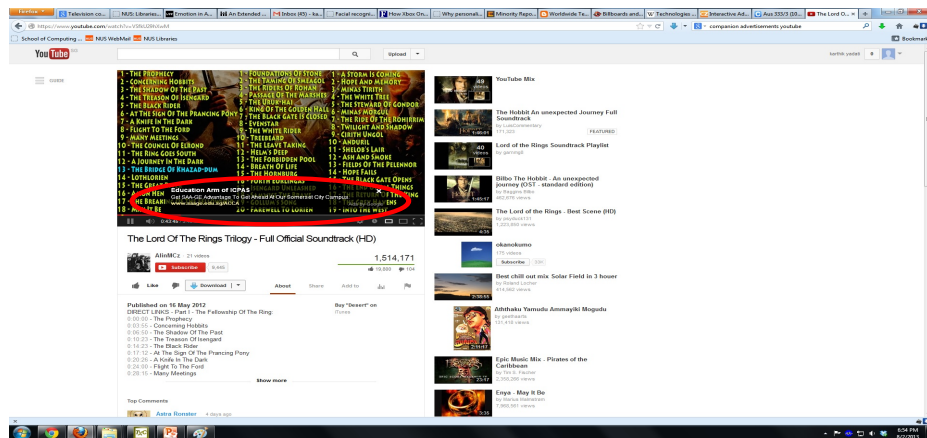


Figure 10: Example of overlay advertisement on YouTube

engines to display advertisements on their search results pages based on the keywords in the user's query. Contextual advertising is a form of targeted advertising in which the content of an advertisement is in direct correlation to the content of the

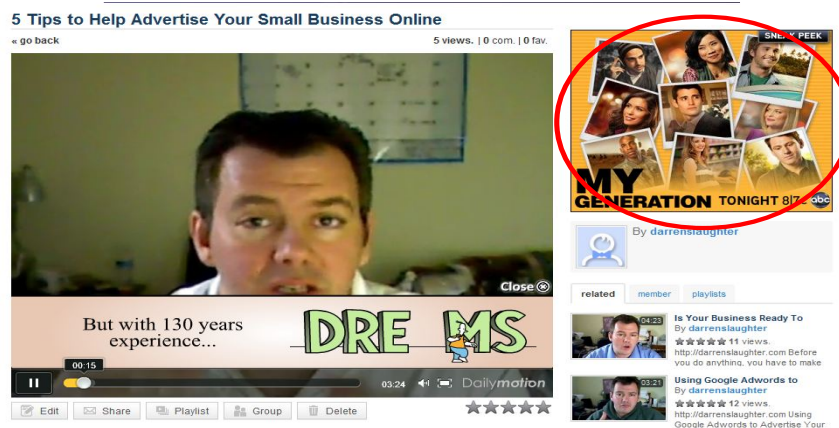


Figure 11: Example of Companion advertisement on YouTube

web page the user is viewing. For example, if you are visiting a website concerning traveling in India and see that an advertisement pops up offering a special price on a flight to Delhi, that's contextual advertising. Google AdSense was the first major contextual advertising network. It works by providing webmasters with JavaScript code that, when inserted into web pages, displays relevant advertisements from the Google inventory of advertisers. The relevance is calculated by a separate Google service, that indexes the content of a web page. Contextual advertising is popular in the textual/banner advertising on web pages and is receiving attention in the multimedia analysis community with the expansion of community video sharing websites like YouTube. Here we present previous work, which address the problem of contextual advertising in videos.

A novel contextual advertising system, *AdImage* [LCH08], which automatically associates relevant ads by matching characteristic images, referred to as adImages (analogous to adWords). AdImage provides a framework for placing contextually

relevant video and image advertisements by measuring the content based relevance and the semantic relevance of the advertisements. The framework defines two terms called *AdImage* and *AdConcept*. *AdImage* is a characteristic image for a product e.g. the logo and *AdConcept* could be a semantic concept like a car, ocean etc. or any semantic event, e.g., a sports event. The advertiser would specify a set of *AdConcepts* and *AdImages*. Now, given a video the framework tries to locate the *AdConcept/AdImage* in the video and places the relevant advertisement at the point where the concept/image is located. In order to select the appropriate advertisement, among various competing advertisements, the framework computes a scheduling score based on the bid placed by the advertiser, unspent budget of the advertiser and the relevance of the advertisement to video. An advertisement with the highest scheduling score is selected. An example advertisement insertion is illustrated in Figure 12. The proposed framework consists of two different modules:

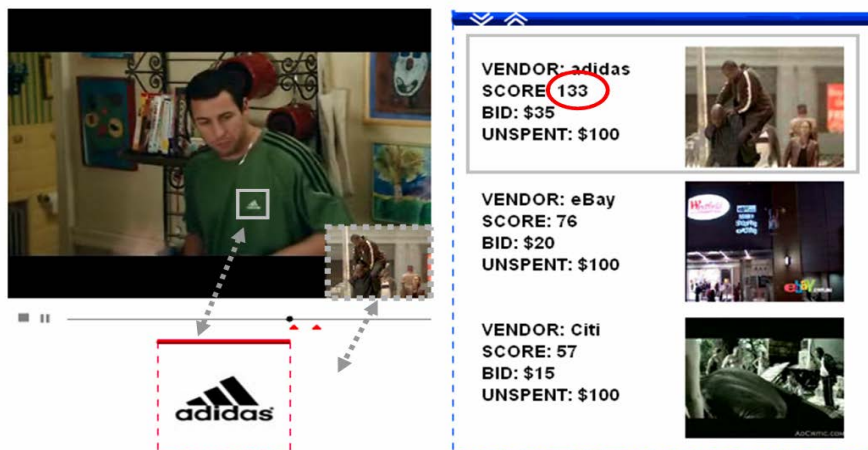


Figure 12: Example of advertisement insertion using AdImage [LCH08]

1. Image matching: The proposed approach for image matching includes three

major parts. First, for each image frame in the video, we adopt Lowe’s Difference of Gaussian method to detect feature points and scale-invariant feature transform [Low04] to represent properties within these feature points. For efficiency, we then adopt an approximate nearest neighboring indexing method (i.e., [AMN⁺98]) to locate matched feature points between the adImage and the inspected frame. Finally, spatial constraints by approximating applicable affine transform between matched feature points are used to remove the outliers, which are the matched feature points not complying with the estimated affine transform. A candidate video frame contains an adImage once the number of the matched inliers is larger than a threshold. Note that all adImages are matched for each frame. This creates a problem for large video collection (ex: YouTube), as this method is difficult to scale.

2. Ad scheduling: In AdWord [MSVV05], each advertiser places bids on a number of keywords and specifies a maximum daily budget. As queries arrive during the day, certain ads will be displayed for their relevance. The objective is to maximize the total revenues while respecting the daily budgets and ad relevance. In our framework, after matching each adImage to the uniformly-sampled frames in the video, we will get a sequence of adImage matches with computed fitting scores. Motivated by AdWord [MSVV05] and given a ranking list of candidate ads, online ad insertion is formulated as an optimization problem, which aims at selecting a subset of ads to maximize not only contextual relevance but also total revenues. We also need to schedule the competing ads in a temporal order since some ad videos are likely to overlap in the viewed videos since ad videos are generally a few seconds long and two adImages might be temporally nearby.

Another work in the area of video contextual advertising is *AdOn* [MGHL10] and Figure 13 illustrates the system architecture of AdOn. A video is represented by the combination of visual track (video stream), audio track (script from closed caption and automatically recognized characters embedded in the key frames), as well as ancillary text (title and keywords) which is provided by the content owner. Meanwhile, the video stream is decomposed into a series of shots. A key frame is extracted from each shot to represent the shot content. The overlay captions are obtained by a caption detection and an OCR engine based on these key frames. The ancillary text and scripts are aligned with shots and further used for selecting a list of relevant ads by text-based search techniques. Then, we detect a set of overlay ad locations based on content intrusiveness and importance.

Intuitively, the overlay ads would appear at the non intrusive (e.g., visually smooth without any significant texture) region in the video highlights (i.e., the shots with the most exciting stories). The shot intrusiveness is based on the combination of face, caption, and image saliency maps, while shot importance is measured by the duration and motion intensity in each shot. Given the expected number of ads in the video, as well as a candidate list of ads and overlay positions, a matching module will associate each ad with the most suitable location. The work which addresses the problem of video-in-video contextual advertising is titled *VideoSense* [MHYL07], where advertisement insertion points are identified as points having high discontinuity and low attractiveness from the viewer’s perspective. Discontinuity measures the *dissimilarity* between consecutive shots of a video and attractiveness is defined as the *importance* or the *interestingness* of the video shot from a user’s perspective. The contextually relevant advertisements are identified using two aspects: global and local relevance. Global relevance is obtained by calculating the textual

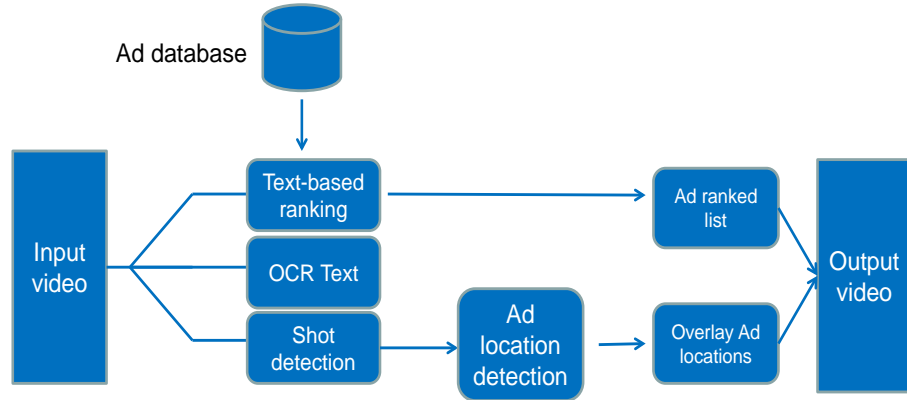


Figure 13: Architecture for AdOn

relevance between the web page containing the video and the keywords associated with the advertisement. Local relevance is computed by measuring the similarities between the video and the advertisement at the insertion point with respect to the following features – motion content, tempo, color. The insertion point selection and advertisement placement is then treated as a 0–1 non-linear optimization problem and solved using a greedy approach.

Table 1 presents a brief summary of the various state-of-the-art contextual advertising strategies and our advertising strategy based on affect. From the table, we can observe that the state-of-the-art contextual advertising strategies do not handle all possible advertisement formats and more importantly, they completely ignore a significant factor in advertising - emotion.

Table 1: Comparison between state-of-the-art contextual advertising and our advertising strategy

<i>Algorithm</i>	<i>Description</i>	<i>Video-in-Video</i>	<i>Overlay</i>	<i>Personalized Companion</i>
VideoSense	Advertisements are placed at shot change points with high discontinuity and low attractiveness	Yes	Yes	No
AdOn	Advertisements are placed on least salient regions of the key frames in shots	No	Yes	No
AdImage	Advertisements are placed based on the occurrence of various company logos in the video	No	Yes	No
Our strategy	Advertisements are placed based on the emotional impact of the video and the advertisement	Yes	Yes	Yes

Another important comparison to be made is between VideoSense [MHYL07] and one of our proposed advertising strategies - CAVVA as both the strategies mainly deal with video-in-video advertising. Table 2 presents a comparison between these two strategies.

Table 2: Comparison of VideoSense and CAVVA

<i>VideoSense</i>	<i>CAVVA</i>
Optimization function based on the semantics of the video and contextual relevance.	Optimization function based on emotional impact of the video and the advertisement in addition to video-ad visual relevance.
Optimization function based on the observation of the following factors Discontinuity of the shot and attractiveness of the shot	Optimization function is supported by results from the consumer psychology literature
Experiments: Subjective user experience	Experiments: Subjective user experience, eye-tracking (pupillary dilation), qualitative advertisement/brand recall
A greedy search strategy is used to obtain a solution to the optimization function	Genetic algorithm (GA) is used to obtain a solution to the optimization function
0-1 Non-linear Programming (NIP) is used to formulate the problem	0-1 Non-linear Programming (NIP) is used to formulate the problem

2.2 Emotion in advertising

Now-a-days, commercials have evolved into short films (as short as 20secs) which convey a story to the user and strike an emotional chord with the user. Emotion evoked by an advertisement can become an important factor in the sales of the product/service. An aspect of advertising, which has been exhaustively studied in the marketing literature, is the role of program context on the effectiveness of

the accompanying commercials. According to [Kei98], the arousal and pleasantness induced by a stimulus (video) tend to interact and there is a residual amount of this affect available to influence the perception of the subsequent stimuli (Ads). Hence, the context of a video effects the arousal and pleasantness induced by an ad. The effects of program induced arousal and valence on the perception of subsequent commercials has been studied independently as well as in a combination. According to the Excitation Transfer Theory [BPW95], emotional responses to advertisements are said to be intensified if there is residual arousal from the previous program. The emotional congruence theory [BPW95] was used to study the effects of the program's valence on the perception of the subsequent commercials. In the current work, we incorporate findings from the marketing literature into our framework in finding the appropriate points of insertion for ads and also for determining the ads appropriate for the context.

2.3 Personalized advertising

Personalized advertising has been existent in online advertising and is termed as *behavioral targeting* as it tries to analyze the user's behavior and then deliver personalized and targeted advertisements. Most implementations of behavioral technology work by tracking a person by setting a cookie on his computer. Following are the steps followed when is cookie is installed:

1. When the user enters the URL in a browser window, the process begins.
2. The browser will check the computer's disk for a cookie associated with the web site whose URL the user has entered. If it finds the associated cookie, the browser will send the cookie information to the web server. If the browser

does not find a cookie, no data is sent.

3. If the web server receives the cookie data along with the page request, it can use that information to personalize the web page for the user.
4. If the web server does not receive a cookie, it knows that the user is visiting the website for the first time. The web server then creates a new ID for the user in the web server's database and sends a cookie in the header for the web page to the user's computer. The user's computer then stores the cookie on the disk. The user is identified uniquely using this ID.

As the user moves across the website or different websites, with every request from his browser this cookie data is passed onto server responsible for behavioral targeting. This server records a host of anonymous data. Typically, following types of visitor specific data can be tracked by the server: City and Country from where the user is accessing the website, local time of the user, browser type, Operating system, time spent on page etc. All this information, along with the information previously tracked as the user moved across pages, is used to make a profile of the user. This profile of the user is utilized to server targeted advertisements.

The above stated personalized advertising tracks the user behavior over websites and delivers targeted advertising. Another type of personalized advertising could be achieved by studying the user's physical and emotional state. Availability of sensors like a web camera allows the advertisers to study the physical and emotional attributes of a user viz. gender, age, ethnicity, facial expression etc. These sensors are becoming cheaper by the day and their availability is spreading wide. Analyzing the user's attributes and his/her behavior can be fruitful for better personalized advertising. An example of a futuristic personalized advertising is demonstrated

in the movie - *Minority Report*. Most of the advertising to consumers in *Minority Report* occurs when they are out of their homes. The advertisements interact in various ways; an Aquafina splashes water on its customers, Guinness recommends its products to the downtrodden to recover from "a hard day at work" etc. The advertisements not only recognize you, but recognize your state of mind to serve more targeted advertisements.

Although such personalized advertising seems very far-fetched, there are a few examples of such personalized advertising which have been deployed in the real world. For example, [Immersive Labs](#) has developed software for digital billboards which can measure the age range, gender and attention-level of a passer-by, and quantify the effectiveness of an outdoor marketing campaign. Beyond just bringing metrics to the outdoor advertisements, facial detection technology can tailor ads to people based on their features. [Plan UK](#), a children's charity group ran a bus-stop advertisement as part of their "Because I Am A Girl" campaign where women passing by would see a full 40-second clip, while if man saw the ad, it would only display a message directing him to their website. The next generation of systems could take this data collection much farther—an algorithm could judge whether you look happy, sad, sick, healthy, comfortable or nervous, and direct personalized advertisements to you.

As part of this thesis, we propose a personalized advertising framework which analyzes the user's emotional state and appropriately delivers targeted advertisements based on experimental results from consumer psychology literature. We use a camera and an eye-tracker to measure the user's emotional state and design algorithms which can select appropriate advertisements. More details about the framework and its evaluation is presented in Chapter 4.

2.4 Previous work

There is a previous work [Yad12], which was based on similar experimental results from consumer psychology and marketing literature but is different from the work reported in this thesis. Table 3 provides a comparison, which highlight the similarities as well as differences with this work.

Table 3: Comparison of previous work and current thesis

<i>Previous work [Yad12]</i>	<i>Current work</i>
Advertising strategy is based on experimental results from consumer psychology literature	Advertising strategy is based on experimental results from consumer psychology literature
The advertising strategy was based on various thresholds and was not scalable	The current advertising strategies are based on an optimization framework and are scalable to large number of videos and advertisements
The strategy tackled only video-in-video advertising	Current work handles different forms of advertisements - in-stream video-in-video, overlay and companion advertisements
Advertising strategy was based on content analysis and emotional states of a group of users	Current advertising strategy also handles personalized advertising for each user, in addition to the content based advertising
Experiments are conducted on a limited group of users	Experiments are conducted over a larger population

3 CAVVA: Computational Affective Video-in-Video Advertising

3.1 What to expect?

This chapter introduces a mathematical framework, based on experimental results from consumer psychology, to insert video advertisements in online videos in an in-stream manner. We propose video-in-video advertising strategy and compare it with existing video advertising practices, contextual advertising and demonstrate its effectiveness through a series of systematically designed experiments.

3.2 Background

As highlighted in Chapter 2, emotion plays a significant role in the success of an advertising campaign. Humans, being emotional creatures, respond to any situation emotionally. Similarly, any advertisement would elicit a definite emotional response from the user. As the author of the book *The Advertised Mind* [Ple05] explains, emotional responses are hard-wired in our brains and are essential for the survival of the human race. Emotional content in advertisements attracts the attention of the user because of his/her past experiences, as it is a common practice to associate any new experience to a similar past experience. For example, if the user sees a familiar and a pleasant event in an advertisement, it is more likely to attract his/her attention than a painful event. As advertiser's aim is to attract users' attention, emotion is an important tool which has been exploited in advertising over the past. Broadly speaking, an advertisement can elicit a positive emotion or a negative emotion in the user. Since emotion is a central factor in our advertising strategy, an

important step is the classification of advertisements based on the type of emotion elicited by them - positive or negative. Film theories suggest that content producers use different techniques to convey the mood of the video clip or the image. For example, usage of bright and saturated colors generally indicates a happy, cheerful mood (positive emotion) while usage of gloomy and unsaturated colors induces a sad mood (negative emotion) in the user. Figure 14 gives an example for a positive and negative emotion eliciting video advertisements, while Figure 15 provides an example of print advertisements eliciting positive and negative emotions in the user. Various factors, including emotion, influencing advertisements have been studied in the consumer psychology and marketing literature over the past few decades. We build our mathematical framework, for automatic advertisement insertion, based on such experimental results from literature.



	Advertisement	Theme	Description	Affect
(a)		Animated characters made of chocolate are shown having fun in a theme park full of chocolate	Since this is a cheerful advert, the color distribution is bright and the motion activity is higher to convey excitement.	Pleasant/High valence
(b)		A young girl is shown breaking an egg violently depicting how the brain gets damaged when people consume drugs	Since this advert shows negative content, the color distribution is gloomy and also high motion activity and sound energy to convey violence.	Unpleasant /Low valence

Figure 14: Example of (a) [Pleasant advertisement](#) and (b) [Unpleasant advertisement](#) from our dataset

Owing to the importance of emotion in advertising and millions of dollars spent on advertising by advertisers annually, there has been extensive research on studying the effectiveness of advertisements based on the emotion elicited by them. Since the scope of this thesis is to take into account the emotional impact of advertisements, we will limit ourselves to experiments in the area of emotional

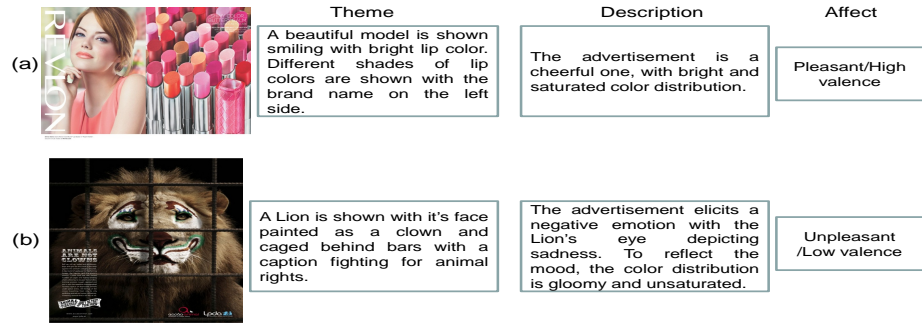


Figure 15: Example of (a) Pleasant print advertisement and (b) Unpleasant print advertisement from our dataset

advertising. These experiments are conducted in a standard setting, where the subjects are shown videos with advertisements inserted in between. The videos are edited professionally to imitate the broadcast style and quality. After watching the videos, the users answer a questionnaire which judge their responses to different types of advertisements in various emotional contexts. We summarize experimental results, in such a setting, from the consumer psychology and marketing literature in the form of rules as follows:

1. In a low arousal, low valence (unpleasant) program context, viewers treat the subsequent advertisements favorably, *opposite* to their evaluation of the program [BPW95]. This is referred to as a *contrast* effect, as the users evaluate the subsequent commercials in the *opposite* direction to their evaluation of the program.
2. In a high arousal, high valence (pleasant) program context, viewers treat the subsequent advertisements as pleasant, similar to their evaluation of the program [BPW95]. This is referred to as an *assimilation* effect as the users evaluate the commercials in the *same* direction as their evaluation of the

program.

3. A positive commercial viewed in the context of a positive program is treated as pleasant, when compared to the same commercial viewed in a negative program context [KMS91].
4. Human beings try to overcome their negative mood and they try to maintain their positive mood. [Ple05]

From rule 4, we observe that whatever the current mood of the user (pleasant or unpleasant), he/she would want to move to a pleasant mood. Advertisements are of varied types inducing both positive and negative emotions. An advertisement which can induce a positive mood in the user is termed as a positive advertisement and an advertisement which induces a negative mood in the user is termed as a negative advertisement. Negative emotion inducing advertisements are generally public service announcements like an anti-drug campaign, where the viewers are advised to stay away from drugs by highlighting the problems faced by people taking into drugs. From rule 4, we can say that advertisement insertion should help the user transition to a higher valence (pleasant) state through the advertisement, which implies that the scene following the advertisement should be of higher valence (pleasant) when compared to the scene preceding the advertisement. Our method aims to choose advertisement insertion points in the video stream so as to minimize disruption and simultaneously select advertisements that will not only be evaluated favorably, but also recalled well later.

Combining rules 2, 3 and 4, we find that a positive advertisement would be appropriate in a context where the scene preceding the advertisement is of high arousal and high valence and the scene following the advertisement falling into the category

of high valence. This would enable a high valence state being maintained. Similarly, we place a negative advertisement at a point where there is a low arousal, low valence scene preceding the advertisement and a high valence scene following the advertisement. By doing this, we achieve a gradual transition from a negative scene to a positive scene through a negative advertisement, where the user experiences an initial inertia to come out of the negative state caused because of the preceding scene. Figure 16 provides a visualization of the advertisement insertion scenarios. The X-axis in the figure refers to the scene index and the Y-axis represents the valence scores. The set of lines labeled as (a) represents a transition from a low valence scene (current) to a high valence scene (next scene) with an advertisement of valence level similar to the valence level of the current scene. Similarly, the set of lines represented by (b) demonstrate a maintenance of the high valence state with the current scene, the advertisement and the next scene having a high valence scores. From the figure, we can say that the advertisement should be emotionally similar to the scene preceding the advertisement.

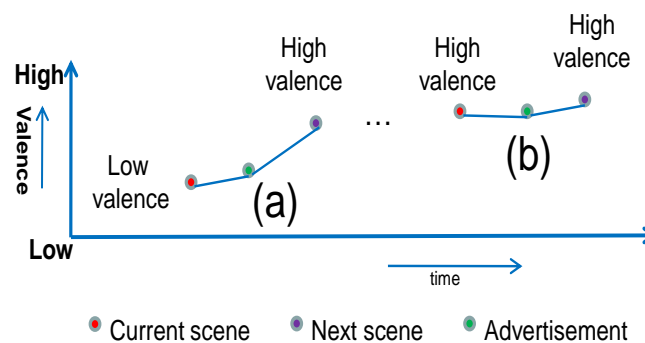


Figure 16: Visualizations of transition in valence. (a) A transition from low valence to high valence through the advertisement, indicating the initial inertia to come out of the negative mood. (b) Maintaining a high valence before and after the advertisement.

3.3 Proposed Method

The problem of video-in-video advertisement insertion is to perform in-stream advertisement insertion in videos. Figure 17 gives a schematic of the sequence of steps involved in our advertisement insertion strategy titled CAVVA (Computational Affective Video-in-Video Advertising). Given a video and a set of advertisements, we identify scene change points in the video where advertisements can be inserted and simultaneously identify a subset of advertisements which can be inserted at the identified advertisement insertion points. The input to the system is a set of advertisements and a video, while the output is an augmented video with advertisements inserted in an in-stream manner. Now, given a video and a set of advertisements, we have devised a sequence of steps which result in an output video with advertisements inserted in it.

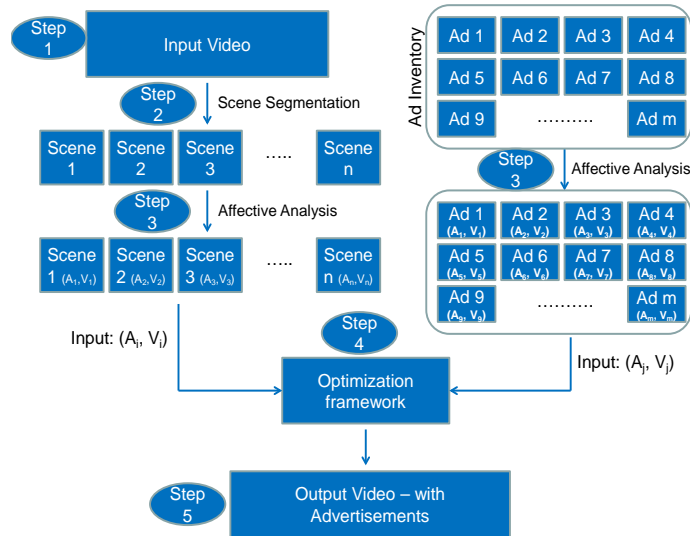


Figure 17: Affect based advertising strategy - CAVVA. (1) Input video, (2) Scene change detection, (3) Affective analysis, (4) Optimization framework, (5) Output video.

3.3.1 Step 1: Input video and advertisements

We have an input video, which could be any video available over the internet. For example, a user clicks a video on YouTube and based on various factors, YouTube inserts advertisement into the video. We have tried to use representative videos from the internet in order to provide a realistic evaluation of our method, details of which would be provided in the section on experiments. We also have a pool of video advertisements collected from various sources on the internet.

3.3.2 Step 2: Scene change detection

We consider each scene change point as a probable advertisement insertion point and hence we perform a scene segmentation of the input video as per [RS03]. A scene is a segment of a video where a certain action takes place. Scene segmentation has been studied widely in the video analysis literature. We propose to use the method as described in [RS03], because of its simplicity, good results and applicability to the current problem. The algorithm is a two-pass approach which uses motion content, shot length and color properties of shots to detect scene boundary points. The video is initially parsed into shots by camera break detection. Each shot is represented by one or more key frames and for each shot, its length and motion contents are also estimated. In the pass one of the algorithm, a color similarity measure of shots is computed called Backward Shot Coherence(BSC). We find valleys in BSC and detect several Potential Scene Boundaries (PSB). A scene with changing contents (for example an action) may split into many scenes for not satisfying color similarities. We merge scenes during the pass two by analyzing Shot Dynamics (SD) of each scene.

3.3.3 Step 3: Affective video analysis

We perform the affective analysis on the individual scenes as well as the advertisements. We employ the popular continuous space of the circumplex model [Rus80] in order to model affect computationally. Affect is modeled in terms of two dimensions - Arousal (intensity of the emotion, varying from excitement to calmness) and Valence (type of emotion, varying from happy to sad). Different low-level audio-visual features are extracted in order to model the two components of affect [HX05]. In order to model arousal, we compute motion activity, shot change frequency and average sound energy in the audio stream. Similarly, we compute pitch in the audio stream and the HSV color histogram to model valence. Various principles from film theory are the basis for the choice of features to model the individual components of affect. For example, in order to show high levels of excitement (high arousal), the directors introduces a lot of motion in the video. Similarly, in order to show a pleasant scene, a director uses bright and saturated colors. In order to show a sad scene, the director can use a dull and unsaturated color distribution.

3.3.4 Step 4: Optimization framework

The output of the Step 3 is a set of scores for valence and arousal for the scenes as well as the pool of advertisements. These arousal and valence scores are provided as an input to an optimization function which then identifies the advertisement insertion points and the subset of advertisements to be inserted. We use a 0–1 Nonlinear Integer Programming framework to formulate the optimization function and then use a genetic algorithm to solve the function. We provide the details of the optimization function as well as the solution in later sections.

3.3.5 Step 5: Output video

The optimization framework outputs the advertisement insertion points i.e., the scene change points of the video where the advertisements can be inserted and the subset of advertisements to be inserted at the identified insertion points. We use a professional video editor to create an output video, which is the result of the video-in-video advertisement insertion mechanism.

3.4 Problem Formulation

Given a video V of duration t seconds. Let the video be segmented into n scenes $Scenes = \{s_i\}$, $1 \leq i \leq n$. As there are n scenes, there would be $m = n - 1$ scene change points (excluding the first and the last scene). We consider each scene change point as a potential advertisement insertion point and that would result in m probable advertisement insertion points. Each scene of the video has an arousal score $A_s = \{a_i\}$ and a valence score $V_s = \{v_i\}$, $1 \leq i \leq n$; $0 \leq a_i \leq 100$ and $0 \leq v_i \leq 100$ associated with it.

Additionally given are a set of p advertisements $Advertisements = \{ad_j\}$, $1 \leq j \leq p$ with an associated value of arousal $A_{ad} = \{a_j\}$ and valence $V_{ad} = \{v_j\}$ with $1 \leq j \leq p$ and $0 \leq a_j, v_j \leq 100$.

Now, we introduce the following decision variables $x \in R^m$, $y \in R^p$, $x = [x_1, \dots, x_i, \dots, x_m]^T$, $x_i \in \{0, 1\}$, and $y = [y_1, \dots, y_j, \dots, y_p]^T$, $y_j \in \{0, 1\}$ where x_i and y_j indicate whether insertion point s_i and advertisement ad_j are selected ($x_i = 1$; $y_j = 1$). We adopt the nonlinear 0-1 integer programming (NIP) paradigm to design an optimization framework to incorporate the above stated rules. The function AI defines the cost for identifying the advertisement insertion points and the function

AS defines the cost for selecting the advertisements. The function AS includes a term $Relevance(s_i, ad_j)$, which computes content based relevance between the scene s_i and the advertisement ad_j based on the following features motion, audio tempo and color. This definition of relevance is similar in spirit to local relevance in VideoSense [MHYL07]. Our final optimization equation is defined as $f(x_i, y_j)$, which is a summation of the terms involving the cost of identifying advertisement insertion points (AI) and the selection of suitable advertisements (AS). The final constraints in the optimization equation account for the uniform distribution of advertisements across the video, by ensuring that there is at least one advertisement in a group of M scene change points (M being the number of advertisements to be inserted in the video). We expect our constraint on uniform placement of advertisements to perform better with increase in the frequency of advertisements inserted into the video stream. Though this constraint performs satisfactorily for our current dataset, we might have to redesign this constraint on comparable lines as the distribution entropy measure in [MHL09]

$$AI(x_i) = \sum_{i=1}^m x_i \left[\frac{(A_s(i+1) - A_s(i))(max(A_s) - A_s(i))}{max(A_s)} + \frac{V_s(i+1)}{max(V_s)} + \frac{V_s(i+1)}{V_s(i)} \right] \quad (1)$$

$$AS(x_i, y_j) = \sum_{i=1}^m \sum_{j=1}^p x_i y_j \left[1 - \left| \frac{V_s(i) - V_{ad}(j)}{max(max(V_s), max(V_{ad}))} \right| \right] \times [Relevance(s_i, ad_j)] \quad (2)$$

$$\max f(x_i, y_j) = \alpha AI(x_i) + \beta AS(x_i, y_j) \quad (3)$$

$$(4)$$

$$s.t \sum_{i=1}^m x_i = M; \sum_{j=1}^p y_j = M ; \alpha + \beta = 1;$$

$$where \ x_1 + x_2 + \dots + x_M = 1;$$

$$x_{M+1} + x_{M+2} + \dots + x_{2*M} = 1; \dots$$

Our optimization function identifies the advertisement insertion points and also the suitable advertisements to be inserted at those points in a unified framework simultaneously. The AI term formalizes rules 1,2 and 4. Furthermore, we require advertisements that can help maintain or transition to high valence states, this is implemented in the $AS(x_i, y_i)$ term. A summary of different variables used in the optimization equation is presented in Table 4

3.4.1 Efficiency and Quality of CAVVA

Considering m potential insertion points in a video and a total of p advertisements, we need to identify M points in the video where advertisements can be inserted. As a result, the search space becomes $O(m) = \binom{m}{M} \times M!$. We employ an efficient genetic algorithm based solver [Whi94], in order to solve the optimization function. We use the global optimization toolbox available in Matlab® to formulate the above stated optimization function and obtain the solution. Though the idea of

Table 4: Variables used in the optimization framework

<i>Variable</i>	<i>Description</i>
$i; j$	Scene index; Advertisement index
$n; m; p; M$	# scenes; # probable insertion points; # of advertisements; # of advertisements to be inserted
$x_i; y_j$	Binary variable (0/1) for insertion point; advertisement
$AI(x_i)$	Function to determine the advertisement insertion point
$AS(x_i, y_j)$	Function to select the appropriate advertisement
$A_s(i);$ $A_s(i+1);$ $max(A_s)$	Arousal score of the current scene, next scene and maximum arousal value
$V_s(i);$ $V_s(i+1);$ $max(V_s)$	Valence score of the current scene, next scene and maximum arousal value

using genetic algorithm is mentioned in VideoSense [MHYL07], they implement a greedy selection strategy in their paper. The genetic algorithm works on a population, which represents a set of points in the solution space. A solution to our problem would be of the form $\{x_i, y_j\} 1 \leq i \leq m, 1 \leq j \leq p$. A typical solution vector would be of length 212 (for 200 advertisements and 12 scene change points on an average), with $x_i = 1$ and $y_j = 1$ for the chosen insertion points, advertisements and everything else as 0. An initial population of vectors is chosen randomly by the algorithm and all the subsequent generations (population at each iteration) are generated computing the value of the fitness function (Eqn (3)). An important parameter, in identifying a global optimum, is the size of the population which represents the number of individuals present in each generation. As the size of the population increases, the chances for obtaining a globally optimal solution also increase. But, the search process in a large population increases the running time of the algorithm. In order to obtain a trade-off, we used different population sizes

and found that a population size of 800 gives better performance for our data. Another important parameter is the number of generations which are produced before finding the solution. We limit the number of generations to 100 in order to keep the running time low. We also implement brute force algorithm Alg. 1 to evaluate the quality of solutions obtained from the genetic algorithm. Comparing the results of the genetic algorithm and Alg. 1, we find an 85% agreement and hence we can say that the genetic algorithm results in a near optimal solution.

Algorithm 1: Brute-force advertisement insertion

- 1) Initialize a matrix Ad_Scene_Map of size $n \times m$, where rows represent the scenes of the video and the columns represent the advertisements.
 - 2) Initialize another matrix Ad_Select of size $M \times 2$, where M is the desired number of advertisements to be inserted in the video.
 - 3) Compute $f(x, y)$ for each of the scenes and the advertisements and fill the matrix Ad_Scene_Map
 - 4) Compute $maxf(x, y)$ as the maximum value across the matrix Ad_Scene_Map and update the matrix Ad_Select with the corresponding x_i and y_j Fill in a large negative number in the rows $i \pm n/4$ of Ad_Scene_Map , to ensure uniform distribution of ads. Similarly, to avoid repetition of ads, we fill a large negative number in the column j of Ad_Scene_Map
 - 5) Repeat steps 4 & 5 until the desired number of ads (M) are selected.
-

3.5 Experiments

In this section, we describe the different experiments conducted in order to demonstrate the effectiveness of our advertisement insertion method (CAVVA). The experiments have been designed keeping in view the two main objectives of an advertising strategy – reduced disturbance for the user and increased engagement with the advertising content. For a subjective evaluation of CAVVA, we conduct a user-study which gathers user responses to a systematically chosen questionnaire. We

also perform an objective evaluation by conducting an eye-tracking experiment to assess the users' arousal during the advertisements. Additionally, we also conduct a qualitative recall based experiments which measure how well the user remembers the advertisement.

In order to compare our method with the existing strategies, we use the following different advertising strategies.

- **PRPR**: Pre-roll/Post-roll advertising. In this method, advertisements are inserted at the beginning and the end of the video. (ex: YouTube)
- **VideoSense** [MHYL07]: Contextually relevant advertisements are placed at strategically chosen points of high discontinuity and low attractiveness in users' perspective. For a fair comparison, we ignore the global relevance between the video and the advertisement as mentioned in [MHYL07].
- **CAVVA**: Our method, in which we employ the rules on affect from consumer psychology and marketing literature. Advertisement insertion points and the advertisements themselves are chosen according to the optimization equation.

3.5.1 Data Collection

We utilize the popular online video collections [Tellyads](#) (advertisements) and [Youtube](#) (videos) for compiling our dataset. We collect *15 videos* of total duration *165 minutes* with an average duration of *12 minutes* per video clip. After scene segmentation, we have a total of *185 scenes* and an average of *12 scenes* per video clip. We include videos from various genres like TV shows (3 videos with 40 scenes), movie clips (3 videos with 44 scenes), news broadcasts (2 videos with 26 scenes), animated clips (2 videos with 22 scenes), documentaries (2 videos with 32 scenes) and user

generated videos (3 videos with 21 scenes), which are a part of the videos available on YouTube. Table 5 summarizes the video data used in our experiments. In terms of advertisements, we collected 200 unique video advertisements which belong to different product categories. Each category and the percentage it constitutes in our dataset is described as follows - Sports goods (8%), Food products (29%), Life insurance (8%), Baby care products (6%), Clothing (6%), Beauty products (10%), Automobiles (10%), Consumer electronics (16%) and Public service announcement (7%). Table 6 summarizes the types of advertisements used in our experiments.

Table 5: Video Data Used for Experiments.

Type of Video	Number of videos	Duration	Number of scenes
TV Shows	3	40 min	40
Documentaries	2	25 min	32
Consumer generated videos	3	20 min	21
Movie clips	3	45 min	44
Animated clips	2	15 min	22
News broadcasts	2	20 min	26

Table 6: Advertisements used in the experiments

Type of Advertisement	Number of advertisements	Proportion
Sports goods	16	20%
Food products	58	29%
Life insurance	16	20%
Babycare products	12	6%
Clothing	12	6%
Beauty products	20	10%
Automobiles	20	10%
Consumer electronics	32	16%
Public service announcement	14	7%
Total # of advertisements	200	100%

3.5.2 User-study

In the user-study, we invite a group of users to participate in the experiment by watching a set of videos and provide subjective responses to systematically chosen questions. In total, there were 48 (male = 22 and female = 26) participants which included undergraduate, graduate students from the university in the age-group of 19–33 and an average age of 24. Each subject was compensated monetarily for participating in the experiment. The users had no knowledge about the underlying advertisement insertion method being used in the video they are watching.

After each user watches a video, he/she is presented with a questionnaire which judges the user’s viewing experience. The advertising strategies and the questions have a significant overlap or are similar to those asked in Videosense [MHYL07], these questions effectively quantify user experience and also help us compare with Videosense as a baseline. Following four questions are posed to the user at the end of each video. The users provide a rating over the scale of 1 (bad) - 5 (good) to each of the four questions.

1. Are the advertisements uniformly distributed across the video?
2. How disturbing are the advertisement insertion points to the flow of the video?
3. How relevant are the advertisements to the video at the insertion point?
4. How do you rate the overall viewing experience?

We have a total of 15 different videos on which we apply the three different advertisement insertion strategies - *PRPR*, *VideoSense*, *CAVVA* to obtain the output videos. We use a block-based design where each user watches 4 different output videos from one of the three advertising strategies. Across all the videos and users,

we have 60 ($15 \times 4 = 60$) samples which is a significant sample size for analyzing the average ratings for each of the 4 questions. As the sample size increases, the standard error of mean would decrease as it is inversely related to the square-root of the sample size. To avoid subject bias, we make sure that each of the videos is watched by atleast 4 people. The selection of the videos and also the order in which each user watches the videos is randomized so as to not induce any kind of bias in the experimental results.

3.5.3 Advertisement/Brand recall

In addition to the subjective questionnaire, the users are also asked to recall the advertisement and the brand. Advertisement/brand recall is an important metric to measure advertisement effectiveness [FBPR10]. We chose to measure the recall based metrics to check how well the user remembers the advertisement/brand, with an assumption that a user who remembers the advertisement is more likely to buy the product. We collect the following recall based metrics from the users:

- ***Cued recall***: The users are asked to recall advertisements/brands seen during the video, from a given list of advertisements/brands.
- ***Uncued recall***: The user is asked to describe the advertisement/brand (in his/her own words) which he/she has seen during the video.
- ***Immediate recall***: Immediate recall is measured immediately after the user has watched the video.
- ***Day-after recall***: This recall metric is measured a day after the user has watched the videos.

3.5.4 Eye-tracking experiment: Measuring pupillary dilation as a proxy for arousal

Pupillary dilation (PD) is a metric used in eye-tracking studies to measure the variation in sizes of the pupil in the eye when exposed to certain kinds of stimuli. Relevant studies as reported in the literature - [HP60], [BMEL08] and [DLJO08] suggest that pupillary dilation gives reliable estimates of physiological arousal. According to [KYKC11], variations in the size of pupil also measures user's interest and engagement levels with the stimuli.

We measure the pupillary dilation of the users while watching the videos using the eye-tracker, which gives us the length of the major axis of the pupil (*major_axis*) and the length of the minor axis (*minor_axis*). In order to measure the pupil size, we compute the area of the ellipse using $area = \pi \times major_axis \times minor_axis$. After computing the elliptical area, we find that the PD signal is a highly fluctuating signal. We smooth the PD signal with a moving average filter [KYKC11]. The smoothed PD signal is then normalized to get the value between 0 and 1. Higher values of pupillary dilation imply higher engagement levels.

The eye-tracking experiment is conducted in a non-intrusive manner on the same set of users who participate in the user-study. All subjects had normal or corrected to normal vision. The user is seated in front of a monitor where the video is being played and the eye-tracker is placed in front of the monitor. We use a binocular infra-red based remote eye-tracking device called SMI RED 250 from SMI, which can record eye-movement data at a frequency of 250Hz. Figure 18 demonstrates the experimental setup which involves a monitor and an eye-tracker. The experimental setup has three important components which is the user (labelled as (1)) who watches the video on a monitor placed in front of him/her (labeled as (3) in

the figure) and an eye-tracker with the given

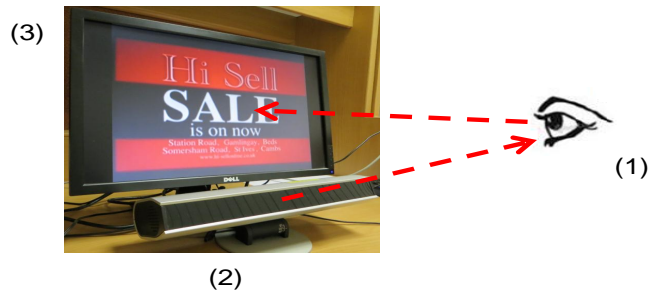


Figure 18: Experimental setup for the eye-tracking experiment, which involves a user (1) watching the video on a monitor (3) and the eye-tracker (2) observing the user.

3.5.5 Ground truth data

In CAVVA, we identify scenes with high/low arousal/valence to insert advertisements. In this experiment, we generate ground-truth labels for arousal and valence for the scenes of a video, from an independent group of users, using the SAM [BL94]. As illustrated in Fig 19, SAM is a 9-point scale for arousal and valence. A scene with an average score of above 7 on the two dimensions is labeled as a high arousal/valence scene. Similarly, a scene with an average score of below 3 on the two dimensions is labeled as a low arousal/valence scene. We compare the ground-truth arousal labels with the arousal labels generated using the pupillary dilation signal [KYKC11] and obtain an agreement level of 69%. We also compare the ground-truth arousal and valence labels with the labels generated using the content-based arousal and valence scores [YKK13b] and obtain an agreement level of 74% for arousal and 71% for valence. These annotators do not participate in the experiments for evaluating CAVVA.

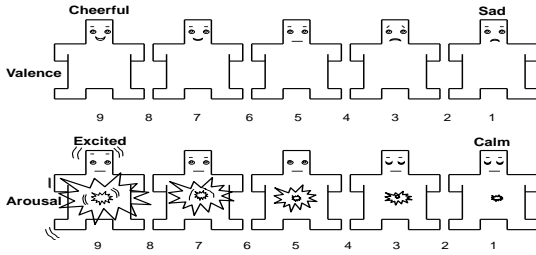


Figure 19: Self-Assessment Manikin [BL94], used to obtain ground-truth valence, arousal data for the videos and the advertisements.

3.6 Results and Discussion

This section provides the results for each of the experiments conducted and also discuss the interesting findings from the experiments. The optimization framework identifies the advertisement insertion points and also selects the appropriate advertisements simultaneously. Figure 20 depicts the example advertisement insertion points and the corresponding advertisements for the three different advertising strategies - Pre-roll/post-roll advertising, VideoSense [MHYL07] and CAVVA, applied on a video from our dataset. For CAVVA, we plot the arousal and valence scores for each of the scenes and the advertisements. Observing the valence-arousal plot, we can say that our advertisement insertion strategy indeed makes a transition from a low valence state to a higher valence state. In order to extend the observation, we provide a visualization of valence scores at 15 randomly chosen advertisement insertion points, from different videos, in Figure 21. The y-axis in the figure represents the valence score and the x-axis represents the current scene, an advertisement followed by the next scene. A maximum valence of 65 (y-axis limit in Figure 21) is observed in our dataset. We see in Figure 21, that our optimization framework brings about the desired lower to higher valence transition in accordance with the rules 1-4 described in the beginning of this chapter. In the

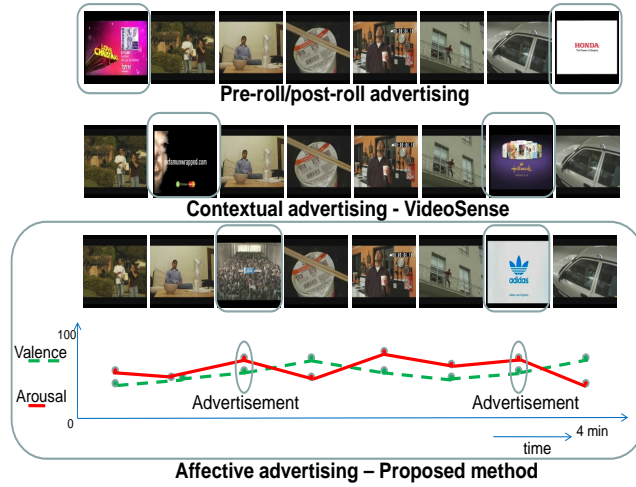


Figure 20: Frames from the result of applying the three different advertisement insertion strategies - PRPR (row 1), VideoSense (row 2), CAVVA (row 3), on an example video. The graph plots the valence, arousal scores for CAVVA(row 4).

following sub-sections, we demonstrate how CAVVA fares when compared to the existing advertising strategies. The results record the subjective user experience, advertisement/brand recall and results from the eye-tracking experiments.

3.6.1 Subjective user experience

User experience is recorded in terms of the ratings for the following systematically chosen subjective parameters: uniform distribution of advertisements, disturbance to the program flow, relevance of the advertisement and the overall viewing experience. Figure 22 presents the average ratings along with the standard deviations from the user-study based on the above parameters, where the X-axis represents the label for each of the subjective questions and the Y-axis represents the average rating to each of the questions with 1 representing a bad rating and 5 representing the best rating. Each color code is for one of the three advertising strategies -

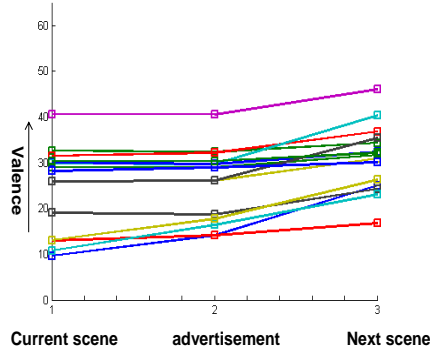


Figure 21: Visualization for transition in valence for 15 randomly chosen advertisement insertion points.

PRPR, *VideoSense* and *CAVVA*. From Figure 22, we observe that *CAVVA* performs better, in terms of average ratings, when compared to the pre-roll/post-roll advertising and *VideoSense* [MHYL07]. In order to establish the statistical signifi-

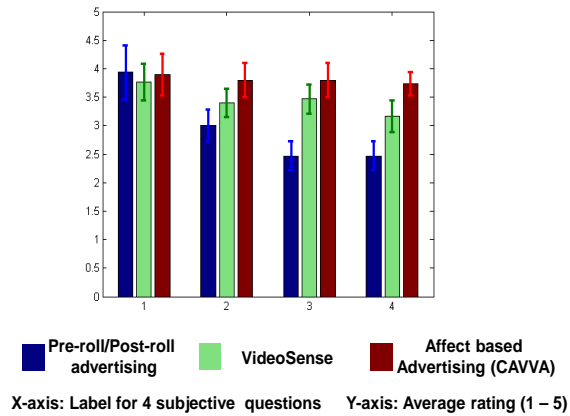


Figure 22: Average ratings to (1) Uniform distribution of videos, (2) Disturbance to the flow of the video, (3) Relevance of the advertisement and (4) Overall viewing experience for each of the three different advertising strategies.

cance of the results, we perform a two-sample Kolmogorov-Smirnov test (KS test) at three different significance levels - $p = 0.01, 0.05, 0.1$. We chose KS test as it does not assume an underlying normal distribution for the data. Table 7 presents the

results from the significance test, where columns 2-5 present the significance levels for each of the subjective questions and column 1 gives the two distributions being compared. The null hypothesis for the test is that the responses, to each question, for different advertising strategies are from the same distribution and the alternate hypothesis is that they are from different distributions. We can reject the null hypothesis at the significance levels mentioned in Table 7. We have used a variety of

Table 7: Results from the two-sample Kolmogorov-Smirnov test for the four subjective questions: Q1-Uniform distribution of advertisements, Q2-Disturbance to the program flow, Q3-Relevance of the advertisement, Q4-Overall viewing experience

Combination	Q1	Q2	Q3	Q4
Pre-roll + VideoSense	p=0.01	p=0.1	p=0.01	p=0.01
Pre-roll + CAVVA	p=0.01	p=0.01	p=0.01	p=0.01
VideoSense + CAVVA	p=0.1	p=0.1	p=0.05	p=0.05

videos and different users would like different videos. Program liking is considered to be an important factor in determining the user’s reactions to advertisements as mentioned in [Cou98]. In order to ensure that liking for a video did not influence the ratings given to the subjective questions, we asked each user to give a rating to the videos based on their liking. Based on this information, we selected top 20% most liked videos and least 20% liked videos and present the average ratings to the questions. Figure 23 presents the results for the most liked and the most disliked video. Observing Figure 23, we find that there is a trend which shows that CAVVA performs on par or better than the other two advertising strategies.

3.6.2 Advertisement/Brand recall

We measure the various dimensions [AKO13] of advertisement/brand recall (cued/uncued, immediate/day-after), which quantifies how much of the advertisement content was

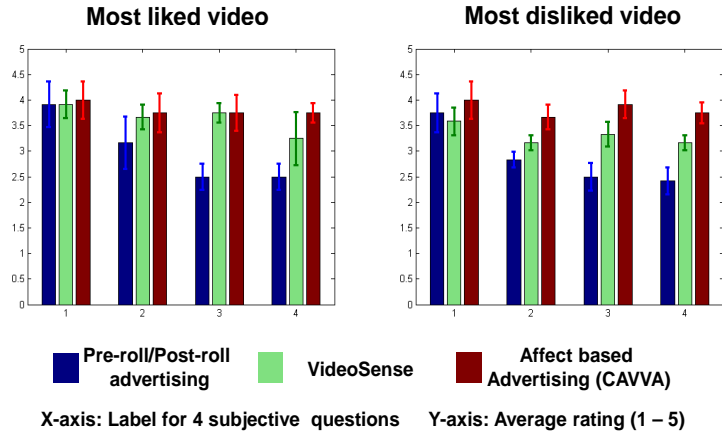


Figure 23: Average ratings for the most liked video (left) and the most disliked video (right).

assimilated by the user and how well the user remembers the advertisement/brand immediately after the session and after some time has passed (day-after recall). Figure 24 presents the results for immediate recall (left) and day-after recall (right). Each group of bars in the figure, on the left, represents a dimension of the immediate recall - uncued advertisement recall, uncued brand recall, cued advertisement recall, cued brand recall in the same order. Each group of bars in the figure, on the right, represents a dimension of day-after recall - cued advertisement recall and cued brand recall in that order. The Y-axis represents the average recall in each of the figures, with 1 representing complete recall of the advertisements where the user remembers all the advertisements and 0 representing a situation where the user does not recall any of the advertisements shown. For the immediate recall, we observe that the recall values are greater or on par with VideoSense [MHYL07] and greater than the pre-roll/post-roll advertising (left of Figure 24). We also observe that there is only a slight difference between the different advertising strategies, in terms of average ratings, for immediate cued advertisement/brand recall. This

behavior is because the user can associate the given list of advertisements/brands and remember them better as he/she has just seen the video. The error lines on top of the bars represents the standard deviation of the recall scores and we observe that it is low. We see a clear increasing trend across the three advertising strategies for the day-after recall - cued advertisement/brand recall. We do not measure the uncued recall metrics a day after the users have watched the video because it would be difficult for the user to write a description on his/her own without help.

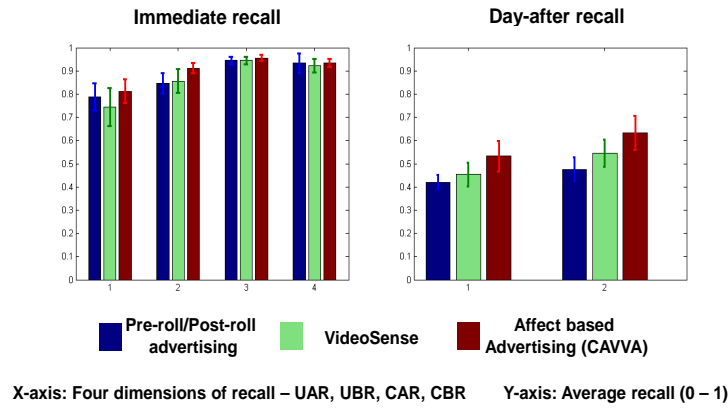


Figure 24: Average immediate recall (left) measured as - (1) Uncued advertisement recall, (2) Uncued brand recall, (3) Cued advertisement recall and (4) Cued brand recall for the three different advertising strategies. Average day-after recall (right) measured as (1) Cued advertisement recall and (2) Cued brand recall.

3.6.3 Eye-tracking experiment

In the previous sub-sections, we have seen that CAVVA improves the subjective experience of the user and also facilitates advertisement/brand recall. In the current eye-tracking experiment, we measure pupillary dilation as highlighted in Section 3.5.4. Since pupillary dilation measures user’s interest and engagement levels [KYKC11], a new advertising strategy should induce similar or enhanced engage-

ment levels, with the advertisements, when compared to the existing advertising strategies. In Figure 25, we plot the average normalized (0-1) pupillary dilation during advertisements for all the users in five different categories - All advertisements, familiar advertisements, unfamiliar advertisements, pleasant advertisements and unpleasant advertisements. The X-axis presents the categories as mentioned previously and the Y-axis represents the normalized average pupillary dilation value in the range of 0–1. In general, we observe an increasing trend across the three advertising strategies - Pre-roll/Post-roll, VideoSense and CAVVA. A higher pupillary dilation indicates a higher interest level from the user and the trend shows that the interest level of the user in the advertisements is higher when the advertisements are inserted using CAVVA.

We have strategically plotted the values of pupillary dilation for different categories of advertisements to account for various factors like familiarity, pleasantness etc. For example, we check for the values of pupillary dilation for pleasant/unpleasant advertisements in order to verify if the higher pupillary dilation is because of a bright color distribution in the pleasant advertisements. Similarly, familiarity with the advertisements can be another factor which can influence the user’s interest level in the content being presented. Another purpose of the categorization is to ensure that there is no user bias towards particular kind of advertisements. For the category of all advertisements, we observe that there is a clear trend which shows that CAVVA induces higher user engagement, when compared to the other advertising strategies. The trend is increasing from Pre-roll/post-roll advertising, VideoSense [MHYL07] to CAVVA. A similar trend is observed for unfamiliar and pleasant advertisements. A slightly different trend is visible in the average values of pupillary dilation in the case of unpleasant advertisements, where we observe that

pre-roll/post-roll advertising out-performs VideoSense. Though VideoSense performs worse than pre-roll/post-roll advertising, we see a higher value for average pupillary dilation when the advertisements are inserted using CAVVA.

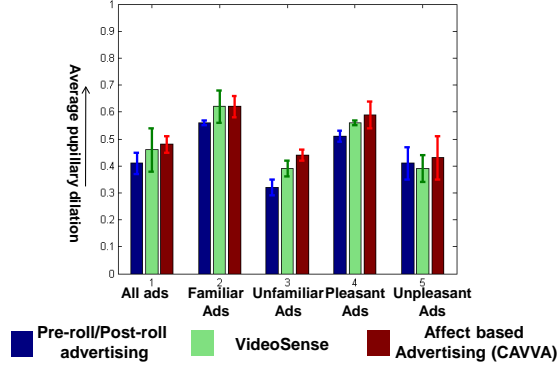


Figure 25: Average pupillary dilation (arousal) during advertisements.

3.7 Support for overlay advertising

Although advertisements are inserted in an in-stream manner in source videos in CAVVA (which is also called "in-video" advertising), we can have different advertisement displaying strategies. For example, advertisements can be displayed as overlapping windows on some spatial positions of the videos. The position can be automatically detected by adding some visually salient region detection modules to CAVVA, or indicated by the viewers through some interactive options. Overlay advertising is quite popular in video sharing websites like YouTube, which utilizes image/textual banner advertisements to be placed at a pre-defined location on the video. Such methods do not leverage audio-visual content analysis in order to identify the location to place overlay advertisements. Moreover, the advertisements often overlay at the bottom areas on the video frame without considering whether there is certain salient information such as text and caption or regions-of-interest

(ROIs) there.

Here we provide one solution taking advantage of the video content analysis techniques to overlay video advertising, based on the formulation given previously for video-in-video advertising. As we have mentioned in the beginning of the chapter, given a source video V , we can obtain a series of shots $S = s_i \ 1 \leq i \leq N_s$ and a corresponding key frame set $F = f_i \ 1 \leq i \leq N_s$, i.e., each shot s_i has one representative keyframe f_i . The task of overlay video advertising is to associate M (M is given as the number of banner advertisements to be placed in a video) ads with M spatial regions in the video frames of V . A straightforward way is two-step:

1. Select the most suitable M key frames in F ; and
2. Find one of the least intrusive regions (usually the overlay region is an image block) for each selected key frame.

In this way, each selected advertisement a_j will overlay on certain region of the selected key frame k_j , displaying for a specific amount of time. To achieve the objective in the first step, we only need to redefine some items in the optimization equation. First, the insertion point set P is now the same as F rather than shot boundaries.

The second step detects the least salient regions within the selected key frames so that the advertisements will not overlay the informative contents in these keyframes. The salient region is typically a ROI, face, or text area. We adopt face [LZZ+02] and text [CZ01] detection, as well as image saliency analysis [MZ03] to detect the most suitable region for placing the ad. Experiments are conducted to demonstrate the performance of the overlay advertising strategy when compared to the two baseline strategies - overlay advertisements at random shot change points (Mode

I) and VideoSense [MHYL07] (Mode II). We label our affect based overlay advertising strategy as Mode III. We conduct the usual experiments of gauging the subjective user experience, qualitative recall and the eye-tracking experiment. We perform better in terms subjective user experience as well as the qualitative recall studies (results similar to CAVVA). We conduct an eye-tracking experiment, which measures the number of fixations on the overlay advertisement and we present the results in table 8. We observe that we perform better than random placement of advertisements and our performance is slightly better than VideoSense [MHYL07].

Table 8: Normalized average number of fixations for Mode I (Random ad placement), Mode II (VideoSense) and Mode III (CAVVA extended to overlay advertising)

Mode	Advertisement	Brand
Mode I	0.4	0.3
Mode II	0.53	0.34
Mode III	0.6	0.36

3.8 Summary

This chapter introduces experimental results from consumer psychology and marketing literature in the form of rules, which can be summarized as follows:

1. Scenes of high arousal, high valence and low arousal, low valence are good candidates for advertisement insertion
2. If the inserted advertisements are of similar emotional tone as the preceding scene, it leads to a good advertising strategy
3. Humans would want to come out of a negative emotional state and would like

to maintain a positive emotional state

These rules are formulated in a mathematical framework and an optimization function is defined in the 0-1 Non-linear Integer Programming paradigm. We then propose a genetic algorithm based solution, which associates scenes of the video with advertisements to be placed after the scene. We compare our advertisement insertion strategy (CAVVA) with state-of-the-art contextual advertising [MHYL07] and commercial existing advertising strategies -pre-roll/post-roll advertising and through a systematic analysis, demonstrate better results for CAVVA in terms of user experience and assimilation of the advertising content. We also propose an extension to overlay advertisements, where we allow placement of banner advertisements on the least salient regions of the video.

4 Companion advertising

4.1 What to expect?

Chapter 4 introduces another optimization function for companion advertising, where a banner advertisement is associated with each scene of the video based on the emotion induced by the scene as well as the advertisement. Since we are not halting the video to play an advertisement, the disturbance caused by the advertisement is minimal and we focus on maximizing the long-term recall of the advertisements. We design user-study and eye-tracking experiments to demonstrate the effectiveness of the advertising strategy.

4.2 Background

It is a form of video advertising, called "companion advertising", where there is a banner/text advertisement by the side of the video. For example, YouTube has companion advertisements on the right hand side of the video. Following are the technical specifications of the companion advertisements on YouTube. These advertisements are generally images (jpg/gif) or flash animations (swf) with dimensions of 300x60 and a maximum size of 50kb. They do not include any audio track and have a maximum animation time of 30 seconds. The user is not given an option to hide companion advertisements. Currently, YouTube has companion advertisements for the corresponding in-stream advertisements or independent banner advertisements. These companion advertisements remain static throughout the playing time of the video and are not related to the content of the video. In this chapter, we propose a novel companion advertising strategy based on the audio-visual content analysis of the video and a set of banner advertisements. We utilize experimental

results from the psychology literature on how emotion influences memory.

Emotion can have a powerful impact on memory. One of the most common frameworks in the emotions field proposes that affective experiences are best characterized by two main dimensions: arousal and valence. The dimension of valence ranges from highly positive to highly negative, whereas the dimension of arousal ranges from calming or soothing to exciting or agitating [Rus80]. Most studies so far focused on the arousal dimension of emotion as the critical factor contributing to the emotional enhancement effect on memory [CM95]. Different explanations have been offered for this effect, according to the different stages of memory formation and reconstruction. In this sub-section, we discuss two important effects which are used in designing the optimization function.

4.2.1 Mood congruency effect

Mood-congruent processing, which means that a person's mood can sensitize the person to take in mainly information that agrees with that mood. Material that is congruent with the mood becomes salient so that the person attends to it more deeply than to other material. The person thinks about that material more deeply and associates it more richly with other information (an activity we call associative elaboration). The result is that the person learns this material better than non-mood-congruent material. Thus, when happy, people will attend and respond more to pleasant than unpleasant parts of their environment and learn more about them; when sad, they'll attend and respond more to its unpleasant than to its pleasant parts and learn more about them.

Though we have not explicitly stated the mood congruency effect, we have been using it through all our advertising strategies - CAVVA and personalized affective

advertising. For example, in CAVVA and the personalized advertising, we place a negative advertisement (an advertisement of low valence) after a scene which induces a negative mood in the user. Similarly, we place a positive advertisement (an advertisement of high valence) after a scene which induces a positive mood in the user. We will employ a similar strategy for companion advertising as well, where we can place an advertisement of valence value similar to that of the current scene. We will use this fact in the optimization function, which will be discussed later.

4.2.2 Relation between arousal and memory

In addition to its effects during the encoding phase, emotional arousal appears to increase the likelihood of memory consolidation during the retention (storage) stage of memory (the process of creating a permanent record of the encoded information). A number of studies show that over time, memories for neutral stimuli decrease but memories for arousing stimuli remain the same or improve [LP98], [Bad82], [SN90]. Others have discovered that memory enhancements for emotional information tend to be greater after longer delays than after relatively short ones [SN90]. This delayed effect is consistent with the proposal that emotionally arousing memories are more likely to be converted into a relatively permanent trace, whereas memories for non arousing events are more vulnerable to disruption.

A few studies have even found that emotionally arousing stimuli enhance memory only after a delay. The most famous of these was a study by Kleinsmith and Kaplan [SN90] that found an advantage for numbers paired with arousing words over those paired with neutral words only at delayed test, but not at immediate test. It has long been known that when individuals process items in an elaborate fashion,

such that meaning is extracted from items and inter-item associations are formed, memory is enhanced. Thus, if a person gives more thought to central details in an arousing event, memory for such information is likely to be enhanced. Results from these experiments show that emotional arousal influences the long-term recall of stimuli, while low arousal impact short-term recall.

We have a situation similar to the experiment reported in [SN90], where each stimulus of different arousal levels are associated with different numbers and the users better recall number which were associated with highly arousing stimuli. In our case, we have the scenes as the stimuli and the number can be comparable to the banner advertisement associated with a scene. The result of the number association experiment suggest that numbers associated with highly arousing stimuli were remembered better over a long period of time. Since we are not disturbing the flow of the video, by placing a companion advertisement, we can focus on creating a strong impression on the user about the advertisement. And this impression can be measured in terms of the long-term recall. The optimization function which would be discussed later looks for highly arousing scenes to associate different kinds of companion advertisements.

4.3 Proposed approach

The input to the system is a video and a set of banner advertisements and the output is a video with different companion advertisements placed on the right side of the video. Figure 26 depicts a block diagram of the proposed companion advertising strategy. The optimization framework provides us with a scene index and the corresponding advertisement index, which we place it in YouTube style web page containing the video and the banner advertisement. The first step is to

divide the video into scenes by performing a scene change detection as proposed in [RS03]. We then perform an affective analysis on each of the scenes and the pool of advertisements to obtain the pair of arousal and valence scores for each scene and the advertisement. These arousal and valence scores are input to an optimization framework which utilizes the experimental results from psychology to associate the scene of the video with a corresponding advertisement.

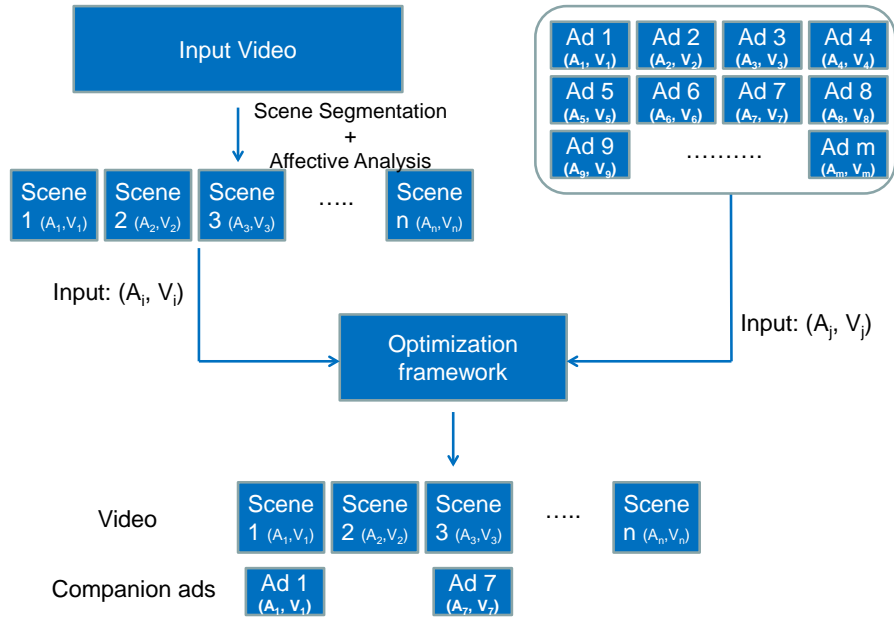


Figure 26: Block diagram for the proposed companion advertising strategy.

4.4 Problem formulation

Given a video V of duration t seconds. Let the video be segmented into n scenes $Scenes = \{s_i\}$, $1 \leq i \leq n$. As there are n scenes, there would be $m = n$ scene change points. We consider each scene change point as a potential advertisement change point and that would result in m probable advertisement placement loca-

tions. Each scene of the video has an arousal score $A_s = \{a_i\}$ and a valence score $V_s = \{v_i\}$, $1 \leq i \leq n$; $0 \leq a_i \leq 100$ and $0 \leq v_i \leq 100$ associated with it. Additionally given are a set of p advertisements $Advertisements = \{ad_j\}$, $1 \leq j \leq p$ with an associated value of arousal $A_{ad} = \{a_j\}$ and valence $V_{ad} = \{v_j\}$ with $1 \leq j \leq p$ and $0 \leq a_j, v_j \leq 100$.

Now, we introduce the following decision variables $x \in R^m$, $y \in R^p$, $x = [x_1, \dots, x_i, \dots, x_m]^T$, $x_i \in \{0, 1\}$, and $y = [y_1, \dots, y_j, \dots, y_p]^T$, $y_j \in \{0, 1\}$ where x_i and y_j indicate whether insertion point s_i and advertisement ad_j are selected ($x_i = 1$; $y_j = 1$). We adopt the nonlinear 0-1 integer programming (NIP) paradigm to design an optimization framework to incorporate the above stated rules. The function AI defines the cost for identifying the advertisement insertion points and the function AS defines the cost for selecting the advertisements. Our final optimization equation is defined as $f(x_i, y_j)$, which is a summation of the terms involving the cost of identifying advertisement insertion points (AI) and the selection of suitable advertisements (AS).

$$AI(x_i) = \sum_{i=1}^m x_i \left[\frac{(-A_s(i))}{\max(A_s)} \right] \quad (5)$$

$$AS(x_i, y_j) = \sum_{i=1}^m \sum_{j=1}^p x_i y_j \left[1 - \left| \frac{V_s(i) - V_{ad}(j)}{\max(\max(V_s), \max(V_{ad}))} \right| \right] \quad (6)$$

$$\max f(x_i, y_j) = \alpha AI(x_i) + \beta AS(x_i, y_j) \quad (7)$$

$$(8)$$

$$s.t. 1 \leq \sum_{i=1}^m x_i \leq m; 1 \leq \sum_{j=1}^p y_j \leq m ; \alpha + \beta = 1;$$

Our optimization function associates the scenes of the video to a suitable banner advertisements. The AI term maximizes the arousal of the scene, by identifying a scene of high arousal in order to improve the long-term recall of the advertisement. The second term $AS(x_i, y_i)$ looks for a pair of (scene s_i , advertisement ad_j) which have similar valence scores (mood congruence effect). A summary of different variables used in the optimization equation is presented in Table 9

Table 9: Variables used in the optimization framework

<i>Variable</i>	<i>Description</i>
$i; j$	Scene index; Advertisement index
$n; m; p; M$	# scenes; # probable insertion points; # of advertisements; # of advertisements to be inserted
$x_i; y_j$	Binary variable (0/1) for insertion point; advertisement
$AI(x_i)$	Function to determine the advertisement insertion point
$AS(x_i, y_j)$	Function to select the appropriate advertisement
$A_s(i);$ $A_s(i+1);$ $\max(A_s)$	Arousal score of the current scene, next scene and maximum arousal value
$V_s(i);$ $V_s(i+1);$ $\max(V_s)$	Valence score of the current scene, next scene and maximum arousal value

4.5 Experiments

This section describes the different experiments conducted to demonstrate the effectiveness of the proposed companion advertising strategy. Current companion advertising has banner advertising displayed on one side of the video and it remains static throughout the playing time of the video. This forms one of our baseline strategies against which we compare our companion advertising strategy. The optimization function allows for both maximizing and minimizing recall and hence we divide our strategy into two strategies which minimize or maximize recall. Emotionally arousing stimuli enhance long-term recall while less arousing stimuli enhance short-term/immediate recall. An example use case for enhancing the short-term recall would be a place, like a shopping mall, where there are bill-boards placed outside the store and depending on the user’s emotional arousal levels (measured using sensors on the bill-boards), we can show corresponding advertisements for products being sold in the store. Following will be the three advertising strategies, for which the results of the various experiments will be compared:

1. Static Companion Advertisement (SCA): This is the state-of-the-art companion advertising strategy used in YouTube
2. Affective Companion Advertisement - Short-term Recall (ACA-SR): We minimize the scene arousal, in order to favor short-term recall
3. Affective Companion Advertisement - Long-term Recall (ACA-LR): We maximize the scene arousal, in order to favor long-term recall

4.5.1 Data Collection

We use the same video data as we had used in CAVVA - *15 videos* of total duration *165 minutes* with an average duration of *12 minutes* per video clip. The advertisements would be different as we are targeting banner advertisements. We collect a total of 200 unique print advertisements, examples of which are shown below.

4.5.2 User-study

In the user-study, we invite a group of 30 users to participate in the experiment by watching a set of videos and provide subjective responses to systematically chosen questions. Following table 10 gives the demographic details of the participants. Each participant was compensated monetarily for taking part in the experiment. The users had no knowledge about the underlying advertisement placement strategy being used in the video they are watching. After each user watches a video,

Table 10: Demographic details of the participants.

# of subjects	# of female subjects	# of male subjects	Age group	Average age
30	13	17	20-29	22

he/she is presented with a questionnaire which judges the user's viewing experience. Following four questions are posed to the user at the end of each video. The users provide a rating over the scale of 1 (bad) - 5 (good) to each of the four questions.

1. What is the number of advertisements of you saw during the video? How do you rate the number of advertisements?
2. Did the changing advertisements disturb the flow of the video?
3. How relevant are the advertisements to the video content?

4. What is your overall viewing experience?

4.5.3 Advertisement/Brand recall

In addition to the subjective questionnaire, the users are also asked to recall the advertisement and the brand, similar to the previous advertising strategies. We collect the following recall based metrics from the users: *Cued recall*, *Uncued recall*, *Immediate recall*, *Day-after recall*.

4.5.4 Eye-tracking experiment: Measuring attention related features

We use a setup similar to the one used during the evaluation of CAVVA, where we have a user watching a video on the monitor and an eye-tracker is observing the user. We do not measure pupillary dilation in order to gauge the user's interest levels as the video keeps playing when the advertisements are displayed and we cannot differentiate between the user's engagement level with the video and the advertisements. Hence, we measure different attention related metrics viz. number of fixations on the advertisement, the ratio of the number of fixations to the amount of time the advertisement is displayed and similar metrics for the bounding box surrounding the brand.

4.6 Results and Discussion

In this section, we demonstrate the results of the above mentioned experiments. Figure 27 illustrates an example video with four corresponding banner advertisements.



Figure 27: An example of a video with 4 associated banner advertisements

4.6.1 User-study

Users were asked to provide a rating for each of the four chosen questions and the average ratings are reported in the following table 4.6.1. Mode I, II and III refer to the three advertising strategies for comparison. From the table, we observe that Mode I i.e, Youtube style static companion advertising performs better than the other two strategies for the Q1 and Q2, as there is only one advertisement throughout the video. For the remaining two questions, Mode II and Mode III (proposed companion advertising) perform better than Mode I.

Table 11: Average ratings for Q1 - number of advertisements, Q2 - changing advertisements disturb the flow of the video, Q3 - relevance of advertisements, Q4 - overall viewing experience. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR

Strategy	Q1	Q2	Q3	Q4
Mode I	4	5	1.3	3.16
Mode II	3.65	4.36	2.46	3.4
Mode III	3.72	4.3	2.7	3.53

4.6.2 Advertisement/Brand recall

Following table (Table 4.6.2) demonstrates the results for immediate recall - Uncued Advertisement Recall (UAR), Uncued Brand Recall (UBR), Cued Advertisement Recall (CAR) and Cued Brand Recall (CBR) across the three advertising strategies. We observe that the YouTube style companion advertising results in 100% recall of the advertisement, as there is only one advertisement to be remembered. The proposed method (Mode II and Mode III) have slightly lesser recall values as there are more advertisements to be remembered and these values are close to each other as the user is being tested for immediate recall. Particularly, we do not find any influence of low arousal on short-term recall of advertisements. Following table

Table 12: Average immediate recall values for three advertising strategies. UAR - Uncued Advertisement Recall, UBR - Uncued Brand Recall, CAR - Cued Advertisement Recall and CBR - Cued Brand Recall. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR

Strategy	UAR	UBR	CAR	CBR
Mode I	1	1	1	1
Mode II	0.75	0.9	0.9	1
Mode III	0.9	0.8	0.8	0.9

(Table 4.6.2) demonstrates the day-after advertisement/brand recall values for advertisement across the three advertising strategies. From the table, we clearly see that the Mode III out-performs the other two modes in terms of day-after recall. This further explains the relation between emotional arousal and the long-term recall of advertisements.

4.6.3 Eye-tracking experiment

We measure the following two properties related to attention - Average fixation frequency per advertisement per advertising strategy and Average fixation frequency

Table 13: Average day-after recall values for three advertising strategies. LAR - Long-term Advertisement Recall, LBR - Long-term Brand Recall. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR

Strategy	LAR	LBR
Mode I	0	0.6
Mode II	0.2	0.4
Mode III	0.4	0.7

on the brand per advertisement per advertising strategy, summarized in Table 4.6.3. Though the average number of fixations is higher for an advertisement in Mode I, the average fixation frequency is lower as a single advertisement is shown for a long period of time. From the table, we observe that advertisements placed during emotionally arousing scenes attract more attention. The average values in the table are normalized in the range 0-1.

Table 14: Average fixation frequencies for advertisement and brand across three advertising strategies. Mode I - SCA, Mode II - ACA-SR, Mode III - ACA-LR

Strategy	Avg. Fixation Frequency (Ad)	Avg. Fixation Frequency (Brand)
Mode I	0.2	0.1
Mode II	0.53	0.31
Mode III	0.62	0.33

4.7 Summary

In this chapter, we proposed an alternative strategy to place companion advertisements for online videos. Existing companion advertising strategies involve placement of random banner advertisements which remain static throughout the playing time of the video. In our companion advertising strategy, we propose an optimization function (based on theories from psychology literature for emotion and

memory), which changes the companion advertisements based on the arousal score of the scene and the mood congruence between the scene and the advertisement. We conduct experiments and analyze the results, which demonstrate the effectiveness of the advertising strategy in enhancing the long-term recall of the advertising content.

5 Personalized video-in-video advertising

5.1 What to expect?

This chapter introduces the paradigm of personalized advertising, where we take into account the user’s emotional state in addition to the emotion induced by the video to place advertisements in the video on-the-fly. Detailed experimental results demonstrate that the personalized advertising performs better when compared to the state-of-the-art video advertising strategies.

5.2 Background

Successful advertising needs user engagement during the advertisement delivery and it is intuitive that successful personalization of advertising strategy can come about by involving the viewer in a human-in-loop manner with the advertisement delivery system. The effectiveness of interaction is also well supported by work in attentive user interfaces [VSCM06] that sense and reason on viewer’s attention and regulate subsequent interaction. In this chapter, we bring forth and address two important complementary aspects of affect and interaction for the online video-in-video advertising problem. We focus on the placement of video advertisements in a video program stream (video-in-video advertising).

As mentioned in the previous chapter, marketing and consumer psychology literature suggests that advertising that can connect to people at an emotional level, is more compelling for the viewer and plays an important role in consumer decision making. The primary objective of advertising is to be informative about the product and influence the user towards an action of buying the product being advertised. In this regard, emotive advertising has been shown to be more effective

than its rational counterpart ³. Apart from the advertisements being emotive, studies [For02] suggest the affective context in which an advertisement can be placed, in order to extract favorable reactions from the users. We identify some important results from experiments in consumer psychology and marketing (repeated from the previous chapter):

1. In a low arousal, low valence (unpleasant) program context, viewers treat the subsequent advertisements favorably, *opposite* to their evaluation of the program [BPW95]. This is referred to as a *contrast* effect, as the users evaluate the subsequent commercials in the *opposite* direction to their evaluation of the program.
2. In a high arousal, high valence (pleasant) program context, viewers treat the subsequent advertisements as pleasant, similar to their evaluation of the program [BPW95]. This is referred to as an *assimilation* effect as the users evaluate the commercials in the *same* direction as their evaluation of the program.
3. A positive commercial viewed in the context of a positive program is treated as pleasant, when compared to the same commercial viewed in a negative program context [KMS91].
4. Human beings try to overcome their negative mood and they try to maintain their positive mood. [Ple05]

We now realize these observations into thumb rules for identifying insertion points as well as selecting appropriate advertisements. The rules (1) and (2) are used to identify the advertisement insertion points and the rules (3) and (4) are used to

³<http://www.neurosciencemarketing.com/blog/articles/emotional-ads-work-best.htm>

identify the appropriate advertisements to be placed at given insertion points in the video. Figure 28 and 29 show valence (pleasantness / unpleasantness) and arousal (excitability) of the video program (spiderman 3) from our evaluation dataset. Pre/post-roll, our method MyAds and VideoSense [MHYL07] are used to insert 2 advertisements into this program. The right panel shows choices made by MyAds, Rule (2) (low arousal and low valence) is followed for the first advertisement insertion point (Ad 1) and rule (1) (high arousal and high valence) is followed for the second advertisement insertion point (labeled Ad 2, right panel; Figure 29). We also observe that for the first insertion point, a negative advertisement is selected (rule (4)) and a moderately positive advertisement (rule (3)) is selected for the second insertion point. The left panel (Figure 28) shows the advertisement insertions made by the affect agnostic pre/post-roll strategies as well as VideoSense [MHYL07]. Mehrabian and Russel [Rus80] have proposed the representation of

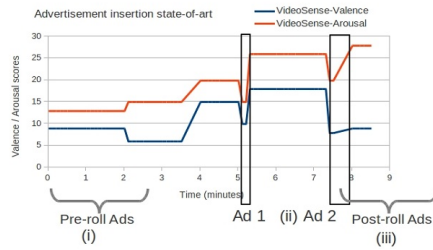


Figure 28: Valence-Arousal plot for (i) pre-roll, (iii) post-roll and (ii) VideoSense [MHYL07].

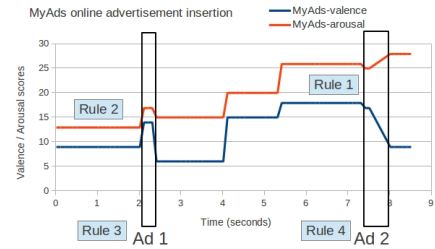


Figure 29: Affective, online advertisement insertion in *MyAds*.

affect in terms of two continuous dimensions - arousal (intensity of the emotion) and valence (type of emotion). Traditionally, arousal and valence corresponding to any audio-visual stimuli are measured using subjective user-studies, where users are shown the stimuli and asked to give their subjective rating for valence and arousal using the self-assessment manikin (SAM) [BL94]. Computational models

for arousal and valence, using relevant low-level features in an image, audio and video content [HX05] have also been proposed, as mentioned in the previous chapter. Affective state of a viewer is typically manifested in various behavioral signals viz., pupillary dilation, facial expression, gestures etc.. For example, eye-tracking has been used in studying human emotions in [DLJO08], where pupillary dilation is used as a measure of physiological arousal. With recent technological developments, modalities such as eye-tracking and face expression analysis have become nearly non-intrusive to the user. We propose a fusion strategy to combine the two channels of information - content-based features and behavioral signals (pupillary dilation and facial expression in the context of this paper), which provides a better assessment of the user’s affective state.

5.3 Feature selection and fusion

We first discover affect related features from video content [HX05] and behavioral information such as pupillary dilation and facial expression and collect valence and arousal ground-truth for each scene in 9 video programs with ratings of 1(0) for high(low) arousal and 1(0) for high(low) valence, given by 5 annotators. Pupillary dilation and facial expression data are recorded for an independent group of subjects as they free-view the chosen videos. We then evaluate behavioral signal based features and content based features [HX05] with human annotated valence and arousal ground-truth. Facial expressions result in characteristic deformations in the face and can convey the emotional state of an individual. We use six canonical emotions (neutral, happy, surprise, anger, sad, fear and disgust)[Ekm96] for our analysis. The out of the box *eMotion* emotion analyzer is employed to give continuous probability scores on the chosen canonical emotions. High(low) valence

is defined as follows, If p_{pos} and p_{neg} are the sum of probability scores over positive emotions (happy, surprise, neutral) and negative emotions (angry, sad, disgust) respectively, over the time window of a video segment (scene)

$$valence(p_{pos}, p_{neg}) = \begin{cases} 1 & , p_{pos} \geq p_{neg} \\ 0 & , p_{pos} < p_{neg} \end{cases} \quad (9)$$

Pupillary dilation information needs additional pre-processing to discover high(low) arousal segments of the video, we extend the method from [KYKC11] for this purpose. Signal variance σ over the smoothed pupillary dilation signal is computed and video segments having significant proportion τ of points with deviation greater than $k \times \sigma$, $k \in [0, 1]$, $\tau \in [0, 1]$, are identified as high arousal. We explore the space spanned by k and τ and find that $k = 0.5$, $\tau = 0.4$ are good choices for affective video content and yield an agreement of upto 76% with high arousal segments of the video. Given a fraction $\delta_{PD(s)} \in [0, 1]$ of the video found to have corresponding pupillary dilation value greater than $k \times \sigma$, we define the indicator function for high(low) arousal in videos,

$$arousal(PD_s) = \begin{cases} 1 & , \delta_{PD(s)} \geq \tau \\ 0 & , \delta_{PD(s)} < \tau \end{cases} \quad (10)$$

We find an average agreement of 74%(71%) between content-based arousal (valence) [HX05] and the ground-truth arousal (valence) and define indicator functions

for content based valence and arousal for a video scene s as follows,

$$content_arousal(s) = \begin{cases} 1 & , \delta_{CArousal(s)} \geq \tau_{CArousal} \\ 0 & , \delta_{CArousal(s)} < \tau_{CArousal} \end{cases} \quad (11)$$

$$content_valence(s) = \begin{cases} 1 & , \delta_{CVal(s)} \geq \tau_{CVal} \\ 0 & , \delta_{CVal(s)} < \tau_{CVal} \end{cases} \quad (12)$$

The values $\delta_{CArousal(s)}$, $\delta_{CVal(s)}$, $\tau_{CArousal}$ and τ_{CVal} are obtained from content-based affect model [HX05]. Similarly, we label the advertisement as positive or neutral by applying the equations 3 and 4 over advertisements.

5.4 Multimodal fusion

An appropriate fusion strategy is an important step, which fuses information from multiple modalities viz., content, eye-tracking. In our method, we have facial expression based valence and pupillary dilation based arousal scores which need to be combined with content based valence and arousal scores respectively. We evaluate a variety of fusion strategies including $argmin(x, y)$, $argmax(x, y)$, $(x \oplus y)$, $product(x \otimes y)$ over pairs of $content_valence(s)$, $valence(p_{pos}, p_{neg})$ and $content_arousal(s)$, $arousal(PD_s)$ values, \oplus and \otimes denote point wise addition and multiplication between two signals. We compare the labels (arousal and valence) given to each video segment using each of the fusion strategies and select the strategy which compares best to the ground-truth labels for arousal and valence. In this case, we find the best fusion strategy to be point wise multiplication \otimes and realize it as a logical *AND* between indicator functions defined earlier. Contrary to previous research [PCC+10], we find comparatively poor correlation between

eye blink rate and eye movement related parameters such as fixation frequency and saccades and valence or arousal ground-truth. Hence we do not include these parameters in our model.

5.5 Personalized online advertising framework

We propose an online advertisement insertion algorithm *MyAds*, to insert advertisements on the fly, as the user is watching the video. We propose an advertising framework taking into account, the affective impact of the video on the user and place affectively similar advertisement in videos, the overall schematic of *MyAds* is shown in Figure 30. Given a video V containing a set of scenes

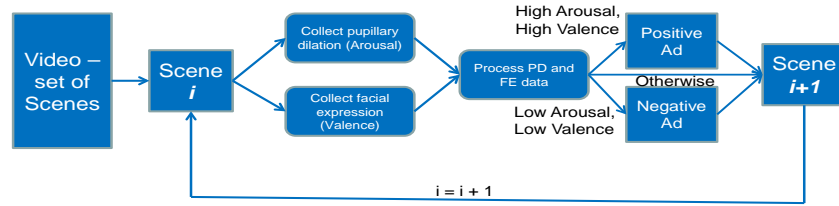


Figure 30: Personalized online advertising insertion.

$S = \{s_1, s_2, \dots, s_n\}$ obtained using automatic segmentation [RS03] and a set of advertisements $A = \{a_1, a_2, \dots, a_m\}$, video scenes s_i and advertisements a_j have varying arousal and valence values computed from low level features. Let PD_i and FE_i be the pupillary dilation and facial expression values buffered over a time window T that ending at the last frame of scene $s_i \in S$ and assuming a recommended number of advertisements N to be placed in the video stream. We aim to place emotionally neutral advertisement in a negative program context [Bol06] and a positive advertisement in a positive program context [BPW95]. We select 9 videos and 50 advertisements for our evaluation, each video yielding an average of 12 scenes.

The insertion of upto 3 advertisements per video stream gives a solution space of $C^3_{12} \times C^3_{50} \times 3!$, the proposed *MyAds* algorithm has linear time complexity $\|S\|$ and is described in Alg. 2.

Algorithm 2: MyAds:

```

input : Video  $V$ , Set of advertisements  $A$ 
 $S \leftarrow$  scenes from  $V$  [RS03];
 $i \leftarrow 1$ ; /* skip_count, minimum inter-advertisement interval */;
skip_count  $\leftarrow 0$ ; while  $i \leq \|S\|$  do
    Playback  $s_i$ ;
    if skip_count  $> 0$  then
        skip_count  $\leftarrow$  skip_count  $- 1$ ;
        Continue to next scene  $s_{i+1}$ ;
    /* Rule (1) */
    if content_arousal( $s$ ) AND arousal( $PD_s$ ) AND
    content_valence( $s$ ) AND valence( $p_{pos}, p_{neg}$ ) then
        /* Rule (3) */
         $a_j \leftarrow$  unused ad  $\in A$  : content_arousal( $a_j$ ) AND
        content_valence( $a_j$ );
        Playback  $a_j$ ;
        skip_count  $\leftarrow \frac{\|S\| + 1}{N}$ ;
    /*  $\sim$  is the complement operator */;
    /* Rule (2) */
    if  $\widetilde{\text{content\_arousal}}(s)$  AND  $\widetilde{\text{arousal}}(PD_s)$  AND
     $\widetilde{\text{content\_valence}}(s)$  AND valence( $p_{pos}, p_{neg}$ ) then
        /* Rule (4) */
         $a_k \leftarrow$  unused ad  $\in A$  :  $\widetilde{\text{content\_arousal}}(a_k)$  AND
         $\widetilde{\text{content\_valence}}(a_k)$ ;
        Playback  $a_k$ ;
        skip_count  $\leftarrow \frac{\|S\| + 1}{N}$ ;
     $i \leftarrow i + 1$ ;

```

5.6 Experimental setup

As we can see in Figure 31, our setup includes 3 different components (labeled in Fig. 31) with which the user interacts. The primary equipment is an eye-tracker which tracks the user’s eye movements as well as the pupillary responses. The eye-tracker used in the setup is a binocular infra-red based remote eye-tracking device SMI RED 250 (<http://www.smivision.com>), which can record eye-movement data at a frequency of 250Hz. In addition to the eye-tracking device, the setup also consists of a camera which observes the user’s face and tracks the different facial expressions shown by the user using an out-of-the-box software *eMotion*. The videos are shown to the user on a stimulus monitor labeled as (d) in Figure 31.

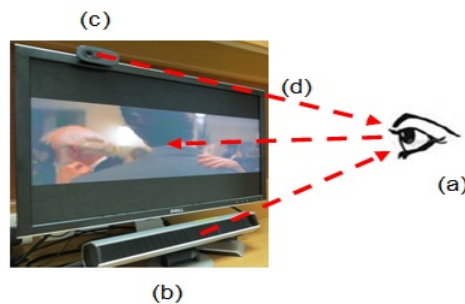


Figure 31: Experimental setup. (a) The user; (b) Eye-tracker; (c) Camera; (d) Stimulus monitor

5.7 Evaluation and User studies

We now evaluate our multimodal, interactive, online affective placement strategy *MyAds* using a thorough user-study. In order to check the effectiveness of the advertisement campaign, we need to quantify how well the user has noticed the advertisements. We find that conventional metrics such as Cost-Per-Impression (CPI) and Click-Through Rate (CTR) do not suit video-in-video advertising so

well, instead the suitable measure is *recall* which assesses how well a viewer remembers advertisement content. The assumption here is that the user is more likely to purchase a product whose advertisement is recalled at the time of purchasing a commodity. We evaluate our interactive online affective advertisement insertion strategy against the standard pre-roll/post-roll advertising and relevance and attention based method in [MHYL07]. Participants are shown a video program with advertisement insertion in one of 3 modes (a) pre-roll and post-roll advertising (*Mode I*) (b) VideoSense [MHYL07] (*Mode II*) and (c) *MyAds* (*Mode III*). Figure 32 provides an illustration of the three different ad-insertion strategies on an example video clip. Three interlinked dimensions of recall are investigated across



Figure 32: Example advertisement insertion and selection - Three strategies

the three advertising strategies,

1. Advertisement/Brand recall: How well does the user recall advertisement content and product brands ?
2. Cued/Uncued recall: After the session, the subject is presented a list contain-

ing some of the brands(cued) or asked to recall purely from memory(uncued)
?

3. Short-term/Long-term memory: The user is asked to do cued / uncued brand / advertisement recall (a) immediately after viewing the video program (b) a day after.

In addition to the recall, the users also answer subjective questions based on the following parameters (i) whether the advertisements are distributed uniformly over the video program (ii) how disturbing are ad-insertions (iii) how contextual relevance of inserted ad (iv) overall viewing experience. This part of the evaluation is similar in spirit to [\[MHYL07\]](#).

5.7.1 Data Collection

A set of 9 (of total duration 130 minutes) videos covering different kinds - movie clips, tv shows in both english and chinese (foreign language, with english subtitles), documentaries, animated clips and 50 advertisements (of average duration 20 seconds) were chosen. The advertisements cover a wide range of products including popular brands such as Coke® and also relatively unfamiliar brands. The video data used in our experiments is similar to broadcast content, but is accessed online.

5.7.2 Experimental design

We recruit our participant pool from university graduate, undergraduate and staff, there were 26 participants in total (14 male, 12 female) with mean age= 24.2 (stdev=4.2). All participants sign consent forms and receive a token payment on completion of the experiment.

The experiments follow a block-based design, where each participant sees 3 randomly chosen videos (program+inserted advertisements), all videos in a block are generated using one of (Mode I/II/III). The participants are ignorant about the strategy being used in placing advertisements. As the proposed advertising strategy *MyAds* is online and involves noticeable editing artifacts when compared to the other two strategies which are offline, users are shown videos stemming from the same advertising strategy. Each video lasts for about 12-15 minutes and the participants then answer systematically chosen subjective questions captured on a 5 point scale. Short/long term cued, uncued brand and advertisement content recall is quantified by us from participant responses using pictorial depictions of brands and text-based descriptions of the advertisements and brands. The long term advertisement/brand recall is assessed offline by the users a day after they have watched the video. All of the baseline methods including VideoSense [MHYL07], have been implemented and following are the details. We use the attention model proposed in [HKP06] to compute *attractiveness* of video scenes and *discontinuity* is computed using the merge level in visual similarity based video segmentation algorithm [RS03]. We model tempo, motion and color based similarity to compute scene-to-advertisement relevance. Feature wise distance between scene and advertisement is computed using KL-divergence between corresponding histograms and $arg_{min}(\cdot)$ over feature wise divergence scores is chosen as local relevance score. Please note that we do not model concept based relevance for fair comparison with our method *MyAds* and (a) pre-roll/post-roll advertising.

5.8 Results and discussion

We now discuss some key results from the evaluation of the proposed *MyAds* algorithm. Normality of the underlying distribution for all scores obtained from the user-study and the recall experiments is ensured using KS(Kolmogorov-Smirnov) tests. We then analyze the performance of MyAds (Mode III) using t-tests with scores from pre/post-roll insertion (Mode I) and VideoSense [MHYL07] (Mode II).

5.8.1 Brand and Advertisement recall

Does brand and advertisement recall improve with *MyAds* ? We derive cued brand recall (CBR), uncued brand recall (UBR), cued advertisement recall (CAR) and uncued advertisement recall (UAR), for short-term over all videos and users; Table 15 records the average recall scores across all the users and videos. In terms of CAR, *MyAds* performs better, when compared to pre-roll/post-roll advertising as well as VideoSense [MHYL07], in terms of a higher average recall score and a lower standard deviation. Similar is the case for UBR, where *MyAds* performs better when compared to the other two baseline methods. In the case of UAR, though the average recall score is higher, the standard deviation is also a bit higher when compared to the score of VideoSense. In this case, we can say that *MyAds* performs at the same level as VideoSense [MHYL07]. Contrary to the other recall metrics, pre-roll/post-roll advertising scores higher average recall values and lower standard deviation, for CBR, when compared to VideoSense and MyAds. This result could possibly be attributed to the fact that the user remembers the brand better as he would have seen the advertisement just before answering the questionnaire (post-roll advertising).

Table 15: Cued and uncued recall over brands and advertisement content

<i>Mode</i>	<i>CAR</i>	<i>UAR</i>	<i>CBR</i>	<i>UBR</i>
<i>Pre/post-roll</i>	0.7 (± 0.09)	0.6 (± 0.096)	0.75 (± 0.064)	0.6 (± 0.1)
<i>VideoSense</i>	0.7 (± 0.097)	0.67 (± 0.075)	0.71 (± 0.08)	0.71 (± 0.09)
<i>MyAds</i>	0.78 (± 0.075)	0.75 (± 0.08)	0.72 (± 0.08)	0.73 (± 0.1)

5.8.2 Subjective experience

Does perceived advertisement placement uniformity(Q1), program flow(Q2), advertisement relevance(Q3) and overall experience(Q4) improve with *MyAds*? We measure subjective viewer responses across all users and videos in Table 16. When compared to Mode I, *MyAds* in row 3 performs better on all parameters ($p=0.1$, 0.05, 0.01 for columns 2, 3, 4) except uniformity of advertisement placement (column 1), where its performance is similar to [MHYL07]. Since Mode I, by definition, places advertisements in a uniform manner, it outperforms the other two methods. When compared to VideoSense [MHYL07], *MyAds* performs consistently better on program flow ($p=0.1$, column 2) and advertisement relevance ($p=0.05$, column 3), where the average rating across users is equal. The scores vary from 0 (bad) to 5 (good).

Table 16: Subjective user-responses on a 5-point scale

<i>Mode</i>	<i>Uniform distribution</i>	<i>Program flow</i>	<i>Ad relevance</i>	<i>Overall viewing experience</i>
<i>Pre/post-roll</i>	4.6 (± 0.5)	3.1 (± 0.56)	2.7 (± 0.6)	3 (± 0.47)
<i>VideoSense</i>	3.63 (± 0.6)	3.27 (± 0.46)	3.45 (± 0.5)	3.54 (± 0.5)
<i>MyAds</i>	3.72 (± 0.46)	3.45 (± 0.68)	3.45 (± 0.52)	3.65 (± 0.6)

5.8.3 Impact on long-term recall

We measure short-term recall by asking the user to recall the advertisements/brands immediately after the viewing session and long-term recall is measured a day after the user has taken part in the experiment. We observe that the *MyAds* in row 3 performs better than the Mode I ($p=0.01$, 0.01: columns 2, 3) and II ($p=0.01$, 0.05: columns 2, 3), for both advertisement and brand recall respectively as illustrated in Table 17. The scores vary from 0-very poor recall to 1-complete recall.

Table 17: Long-term (Day-after) Advertisement/Brand recall.

<i>Mode</i>	<i>Advertisement Recall</i>	<i>Brand Recall</i>
<i>Pre/post-roll</i>	0.3 (± 0.05)	0.29 (± 0.05)
<i>VideoSense</i>	0.3 (± 0.06)	0.31 (± 0.09)
<i>MyAds</i>	0.36 (± 0.08)	0.39 (± 0.1)

5.9 Summary

This chapter presents a personalized video-in-video advertising strategy based on the rules derived from the consumer psychology and marketing literature. The method takes into account the emotional state of the user in terms of pupillary dilation (measured using an eye-tracker) and facial expression (measured using a camera), in addition to the emotion induced by the video to place appropriate advertisements in an in-stream manner. Systematically designed experiments and a thorough analysis of the results demonstrate the effectiveness of the personalized advertising strategy when compared to other state-of-the-art advertising strategies.

6 Summary and Conclusion

6.1 Summary

In this thesis, we explored the affective angle to computational video advertising and we tackled different types of advertisements for videos - video advertisement, overlay banner advertisements and companion advertisements.

In Chapter 3, we explored a novel and effective video-in-video advertising strategy - Computational Affective Video-in-Video Advertising (CAVVA), which takes into account the affective impact of a video and a set of advertisements. An optimization framework is proposed using 0-1 Non-linear Integer Programming (NIP) paradigm, to identify the advertisement insertion points and select suitable advertisements simultaneously. We propose a solution to the problem using the evolutionary genetic algorithms. A variety of experiments are conducted to demonstrate the effectiveness of our advertising strategy (CAVVA) and compare the results with existing video-in-video advertising strategies - Pre-roll/Post-roll advertising and VideoSense [MHYL07]. The results from the experiments suggest that CAVVA enhances user experience (subjective user-study), improves user recall (recall experiments) and an increased user engagement with the advertising content (eye-tracking experiment), when compared to the other advertising strategies. Towards the end of Chapter 3, we also provide details about extending the proposed optimization function in order to place overlay banner advertisements on identified spatio-temporal locations in videos.

As an extension to CAVVA, we demonstrated an algorithm which takes into account the user's emotional state in order to perform in-stream insertion of video advertisements in videos dynamically. The method is compared to state-of-the-art

contextual advertising and commercially viable pre-roll/post-roll advertising and was shown to perform better in terms of subjective user experience as well as qualitative advertisement/brand recall. This work is reported in Chapter 5.

Another form of video advertising strategy (reported in Chapter 4) is known as companion advertising, where banner advertisements are placed on the right side of the video without disturbing the flow of the video. Current methods do not take into account the audio-visual content of the video and also the banner advertisement remains static throughout the playing time of the video. The proposed companion advertising strategy associates scenes of a video with different advertisements based on the experimental results from psychology literature studying the relation between emotion and memory. The proposed method is shown to perform better than the existing companion advertising strategy.

6.2 Contributions

Explosion of video content on the internet has led to a surging interest in advertising on community video sharing websites such as YouTube. Existing video advertising strategies, which are available on video sharing websites such as YouTube, do not take video context into account. Video advertisements are inserted either at the beginning, at the end or at a random position in the middle of the video. Similarly, banner advertisements are overlaid at the bottom portion of the video. Another advertising strategy, called companion advertising, has banner advertisements placed at the right side of the video. Similar to contextual advertising in text, video advertising has also started exploring this advertising strategy from a semantics perspective. In this thesis, we have handled different types of advertising viz. video-in-video advertising, overlay advertising, companion advertising

and design algorithms from an alternative angle - emotion. We have designed our algorithms based on theories and experimental results from consumer psychology and marketing literature. We compare our advertising strategies to state-of-the-art contextual advertising and other strategies prevalent on video sharing websites like YouTube and demonstrate better performance with respect to user experience and assimilation of the advertising content.

As part of the thesis, we also explore the personalized advertising scenario in which we observe the user's emotional state before deciding on where and what kind of advertisements to be placed. We compare the personalized advertising strategy to other offline optimization based advertising strategies and report a better performance in different metrics - subjective user experience, advertisement/brand recall.

6.3 Future work

As part of this thesis, we explored the affective aspects of advertising and future directions include exploring the semantics side of advertising by taking into consideration the contextual relevance of the advertisement to the video and the user. Current work relies on audio-visual content analysis to perform advertisement insertion. There is a recent explosion of social media networks and other forums where the users actively participate, by posting images, videos, comments and thoughts on various topics. These channels of information could be better utilized to model user's current interests, emotional state, which could lead to a better targeted advertisements. Another aspect, which could be of significant interest, is to serve advertisements to groups of users categorized based on different parameters like age, gender, ethnicity and other abstract parameters like personality, economic status etc.

References

- [AKO13] Ron Adany, Sarit Kraus, and Fernando Ordonez. Allocation algorithms for personal tv advertisements. *Multimedia Systems*, 19(2):79–93, 2013.
- [AMN⁺98] Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM*, 45(6):891–923, November 1998.
- [Bad82] A. D. Baddeley. Implications of neuropsychological evidence for theories of normal memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 298(1089):pp. 59–72, 1982.
- [BL94] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.
- [BMEL08] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.
- [Bol06] Wendy Bolhuis. Commercial breaks and ongoing emotions: Effects of program arousal and valence on emotions, memory and evaluation of commercials, May 2006. <http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control-software/>.
- [BPW95] Jr. Broach, V. Carter, Jr. Page, Thomas J., and R. Dale Wilson. Television programming and its influence on viewers’ perceptions of com-

- mercials: The role of program arousal and pleasantness. *Journal of Advertising*, 24(4):45–54, 1995.
- [BSW12] Anders Brodersen, Salvatore Scellato, and Mirjam Wattenhofer. Youtube around the world: geographic popularity of videos. In *Proceedings of the 21st international conference on World Wide Web, WWW '12*, pages 241–250, New York, NY, USA, 2012. ACM.
- [CM95] Larry Cahill and James L. McGaugh. A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, 4(4):410 – 421, 1995.
- [Cou98] Keith S. Coulter. The effects of affective responses to media context on advertising evaluations. *Journal of Advertising*, 27(4):pp. 41–51, 1998.
- [CZ01] Xiangrong Chen and Hongjiang Zhang. Text area detection from video frames. 2195:222–228, 2001.
- [DLJO08] O lafsdottir I De Lemos J, Sadeghnia G and Jensen O. Measuring emotions using eye tracking. In *Spink A (ed) 6th International conference on methods and techniques in behavioral research.*, 2008.
- [Ekm96] P. Ekman. Basic emotions. *Handbook of Cognition and Emotion*, 1996.
- [FBPR10] Paul W. Farris, Neil T. Bendle, Phillip E. Pfeifer, and David J. Reibstein. *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance*. Wharton School Publishing, 2nd edition, 2010.
- [For02] Joseph P. Forgas. Toward understanding the role of affect in social thinking and behavior. *Psychological Inquiry*, 13(1):pp. 90–102, 2002.

- [HKP06] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In Bernhard Scholkopf, John Platt, and Thomas Hoffman, editors, *Neural Information Processing Systems (NIPS)*, pages 545–552. MIT Press, 2006.
- [HP60] Eckhard H. Hess and James M. Polt. Pupil size as related to interest value of visual stimuli. *Science*, 132(3423):349–350, 1960.
- [HX05] A. Hanjalic and Li-Qun Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143 – 154, feb. 2005.
- [Kei98] S. Coulter Keith. The effects of affective responses to media context on advertising evaluations. *Journal of Advertising*, 27(4):41–51, 1998.
- [KMS91] Michael A. Kamins, Lawrence J. Marks, and Deborah Skinner. Television commercial evaluation in the context of program induced mood: Congruency versus consistency effects. *Journal of Advertising*, 20:1–14, 1991.
- [KYKC11] H. Katti, K. Yadati, M. Kankanhalli, and T. S. Chua. Affective video summarization and story board generation using pupillary dilation and eye gaze. In *2011 IEEE International Symposium on Multimedia (ISM)*, pages 319 –326, dec. 2011.
- [LCH08] Wei-Shing Liao, Kuan-Ting Chen, and Winston H. Hsu. Adimage: video advertising by image matching and ad scheduling optimization. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 767–768, 2008.

- [Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.
- [LP98] K.S. LaBar and E.A. Phelps. Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, 9(6):490–493, 1998.
- [LZZ⁺02] Stanz Li, Long Zhu, ZhenQiu Zhang, Andrew Blake, HongJiang Zhang, and Harry Shum. Statistical learning of multi-view face detection. 2353:67–81, 2002.
- [MGHL10] Tao Mei, Jinlian Guo, Xian-Sheng Hua, and Falin Liu. Adon: toward contextual overlay in-video advertising. *Multimedia Systems*, 16:335–344, 2010.
- [MHL09] Tao Mei, Xian-Sheng Hua, and Shipeng Li. Videosense: a contextual in-video advertising system. *IEEE Trans. Cir. and Sys. for Video Technol.*, 19(12):1866–1879, December 2009.
- [MHYL07] Tao Mei, Xian-Sheng Hua, Linjun Yang, and Shipeng Li. Videosense: towards effective online video advertising. In *Proceedings of the 15th international conference on Multimedia*, pages 1075–1084, 2007.
- [MSVV05] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized on-line matching. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 264–273, 2005.
- [MZ03] Yu-Fei Ma and Hong-Jiang Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Proceedings of the eleventh ACM*

- international conference on Multimedia*, MULTIMEDIA '03, pages 374–381, New York, NY, USA, 2003. ACM.
- [PCC⁺10] Wei-Ting Peng, Chia-Han Chang, Wei-Ta Chu, Wei-Jia Huang, Chien-Nan Chou, Wen-Yan Chang, and Yi-Ping Hung. A real-time user interest meter and its applications in home video summarizing. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 849–854, 2010.
- [Ple05] E.D. Plessis. *The advertised mind: ground-breaking insights into how our brains respond to advertising*. Kogan Page, 2005. isbn-9780749443665.
- [RS03] Z. Rasheed and M. Shah. Scene detection in hollywood movies and tv shows. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.*, volume 2, pages II – 343–8, june 2003.
- [Rus80] J A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.
- [SN90] Elisabeth Schurer-Necker. Arousal and paired-associate learning. *The Pavlovian Journal of Biological Science*, 25(4):195–200, 1990.
- [VSCM06] Roel Vertegaal, Jeffrey S. Shell, Daniel Chen, and Aadil Mamuji. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior*, 22(4):771 – 789, 2006.
- [Whi94] D. Whitley. A genetic algorithm tutorial. *Statistics and Computing*, pages 4:65–85, 1994.

- [Yad12] Karthik Yadati. Affective video advertisement insertion: A computational and marketing perspective. Master's thesis, School of Computing, National University of Singapore, 2012.
- [YKK13a] Karthik Yadati, Harish Katti, and Mohan Kankanhalli. Cavva: Computational affective video-in-video advertising. *IEEE Transactions on Multimedia*, 2013.
- [YKK13b] Karthik Yadati, Harish Katti, and Mohan Kankanhalli. Interactive video-in-video advertising: A multi-modal affective approach. In *International conference on Multimedia Modelling, MMM, 2013*, Jan. 2013.