

# Towards Global Optimization in Display Advertising by Integrating Multimedia Metrics with Real-Time Bidding

Xiang Chen

School of Computing

National University of Singapore, Singapore, 117417

chxiang@comp.nus.edu.sg

## ABSTRACT

Real-time bidding (RTB) has become a new norm in display advertising where a publisher uses auction models to sell online user's page view to advertisers. In RTB, the ad with the highest bid price will be displayed to the user. This ad displaying process is biased towards the publisher. In fact, the benefits of the advertiser and the user have been rarely discussed. Towards the global optimization, we argue that all stakeholders' benefits should be considered. To this end, we propose a novel computation framework where multimedia techniques and auction theory are integrated. This doctoral research mainly focus on 1) figuring out the multimedia metrics that affect the effectiveness of online advertising; 2) integrating the discovered metrics into the RTB framework. We have presented some preliminary results and discussed the future directions.

## KEYWORDS

Display advertising; real-time bidding; multimedia metrics; trade-offs optimization

### ACM Reference format:

Xiang Chen. 2017. Towards Global Optimization in Display Advertising by Integrating Multimedia Metrics with Real-Time Bidding. In *Proceedings of MM'17, October 23-27, 2017, Mountain View, CA, USA.*, 5 pages. DOI: <https://doi.org/10.1145/3123266.3123966>

## 1 INTRODUCTION

The explosive growth of multimedia data on Internet has created huge opportunities for online advertising. According to the latest Interactive Advertising Bureau's annual report [24], the second-half revenue totaled \$39.8 billion in 2016, which is an increase of \$7.7 billion from the second-half revenue of 2015. Despite the increasing growth of revenue, the effectiveness of online advertising remains debatable. On the one hand, the user gets annoyed by the improper ads. Marketing data [10] shows that: the average click-through rate of display ads across all formats and placement is 0.06%; young adults tend to ignore online banner ads; and ad blocking has grown by 41% globally in the year 2015. The unpleasant ad experience further influences their site visits, and their perception of the displayed ad. On the other hand, the advertiser gets harmed by ineffective ad delivery. In terms of cost-per-impression pricing mechanism, the advertiser has to pay for every impression. Due to

user's low engagement towards the displayed ad, a large amount of budget that the advertiser spends on online advertising is wasted.

The majority of online displayed ads are served through RTB. Compared to the RTB system that focuses on the publisher's revenue (local maximization), we define the global optimization as the optimal trade-offs among all stakeholders. In this regard, the proper ad is not only the highest revenue from the economic point of view, but also the best fit with the context from the multimedia point of view. Fig. 1 illustrates how the multimedia metrics are incorporated with RTB. Our proposed framework consists of two stages: 1) the first stage uses the second-price (SP) auction model to deal with economic issues; 2) the second stage uses our proposed optimal re-ranking model to select the proper ad. Note that, the independent process of stage I ensures the equilibrium in ad auctions. The multimedia metrics take effect in stage II. Since they are exogenous, our proposed framework does not affect the advertiser's bidding behavior.

Due to the complex nature of display advertising, we are facing the following challenges:

- **What metrics to select?** User's engagement (e.g., view, click) towards the displayed ad varies in different context. Understanding the context and user's behavior becomes the key factor to improve the effectiveness of online advertising.
- **How to integrate the metrics?** Online advertising is a multidisciplinary topic, which involves economics, multimedia and psychology. The metrics represent the stakeholders' benefit from different domains. An easy but effective solution is the linear combination model. Thus, determining the optimal weights becomes the key issue.

The development of computer vision, multimedia analysis and machine learning techniques brings opportunities to solve these two challenges. Firstly, we select a set of multimedia metrics according to marketing data and consumer psychology literature. We then conduct a user-study experiment to demonstrate the effectiveness of the selected metrics. Secondly, we simulate a real-world advertising system based on an auction dataset and two multimedia datasets. By adding constrains on the changes of metrics when compared with existing system, we propose an optimization model to obtain the optimal weights. To the best of our knowledge, we are the first towards global optimization in display advertising by combining multimedia techniques and auction theory.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 describes the proposed optimization framework. Section 4 presents our preliminary results. Section 5 lists our works in progress. And section 6 concludes the paper.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MM'17, October 23-27, 2017, Mountain View, CA, USA.

© 2017 Copyright held by the owner/author(s). ISBN 978-1-4503-4906-2/17/10.

DOI: <https://doi.org/10.1145/3123266.3123966>

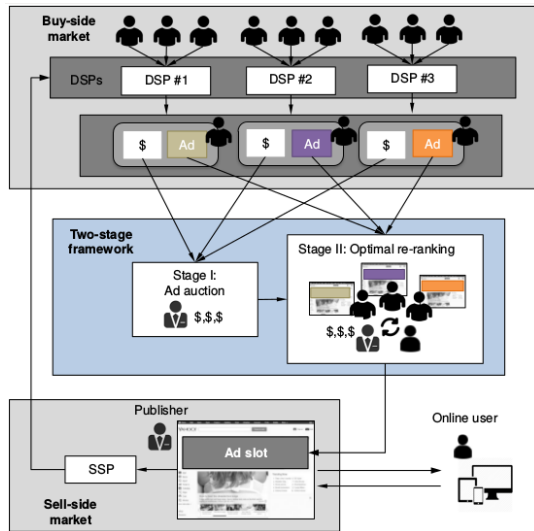


Figure 1: Overview of the proposed framework [6].

## 2 RELATED WORK

From the economics perspective, researchers mainly focus on investigating the strategy of selling ad impressions. The generalized second-price (GSP) auction [8] and the Vickrey-Clark-Groves (VCG) auction [23] have been widely used on different platforms. As revenue is always the primary concern, the squashing parameter and reserve price have been discussed to increase the revenue of GSP auctions [14, 28]. GSP auction model has been proved to be able to improve social welfare as well [13].

From the multimedia perspective, researchers mainly focus on increasing user’s engagement towards displayed ads and meanwhile maintaining user’s online experience. Literature in marketing and consumer psychology has already shown that the contextual relevance between the content of the hosting webpage and the ads makes a large difference in their clickability [5], and it also has a leading effect on user’s online experience [18]. According to the utilized information for contextual similarity matching, we can classify most existing contextual advertising systems into three categories: text-based advertising [15], visual-based advertising [27] and targeted advertising [29]. To achieve better contextual relevance, the multimodal approach has been proposed which considers both textual and visual information for video [9, 19] and image advertising [21].

The works discussed so far only focus on the benefit of a specific stakeholder, regardless of the other stakeholders’ benefits. Trade-offs among multiple objectives or stakeholders have been discussed in several works. Likhodedov et al. [17] proposed a framework that linearly combines revenue and social welfare in a single-item auction. Bachrach et al. [2] proposed a framework that linearly combines relevant, welfare and clicks in sponsored search. Liao et al. [16] combined revenue and relevance for video advertising.

Our goal is to foster a vigorous and healthy online advertising ecosystem. Different from most existing works that focus on local maximum value, we focus on the global optimization of the whole advertising ecosystem. Display advertising is an interplay among all stakeholders and each stakeholder’s ad experience is

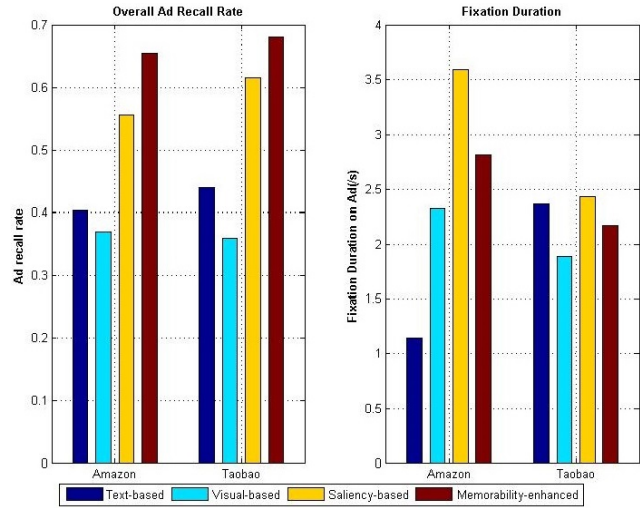


Figure 2: User’s engagement on displayed ad under different combination of multimedia metrics.

an important factor for advertising effectiveness. In the proposed framework, the publisher’s revenue will decrease in a short term in trade of improvements on the benefits of the advertiser and user. The increased ad experience will increase the demand of the advertiser’s advertising needs and the supply of the user’s webpage visits, which will boost the publisher’s revenue in a long run.

## 3 PROPOSED FRAMEWORK

We propose to incorporate multimedia metrics into RTB to improve the effectiveness of display advertising. Our framework consists of two stages (see Fig. 1). The first stage is based on the second-price auction model. And the second stage is based on the optimal re-ranking model to select the most appropriate ad. Since the auction model is simple but effective, we focus on the optimal re-ranking model. When adopting the weighted linear combination model, the key issues are: what metrics should be considered and how to determine the optimal weights.

### 3.1 Select Multimedia Metrics

As the user is the potential customer to the advertiser, any ad that can deliver the advertiser’s information to the user can be viewed as an effective ad. It has been recognized that the user’s acceptance of a message is sequence of psychology process: attention, comprehension, yielding, retention and action [1]. Motivated by this finding, we suggest the following three multimedia metrics that can help improve the effectiveness of online advertising.

**Contextual relevance** has been widely discussed in computational multimedia advertisement [20]. Contextual relevance refers to the similarity between the content of the host webpage and the ad description provided by the advertiser. It has been proved to be effective in increasing the ad CTR and meanwhile maintaining user’s online experience.

**Visual saliency** measures whether the ad image can be easily spotted out within the host webpage. Through a series of eye-tracking experiments, recent research found that: users intentionally avoid looking at the banner ads [25], and they tend to ignore

**Table 1: Specifications of metric variables for Stage II.**

Variable	Computation			Input for Stage II
	Method	Input	Output	
Publisher's revenue	Stage I	Bid	Payment	$x_{1,i,z}$ = normalized $p_{i,z}$
Advertiser's utility	Stage I	Bid, payment	Utility	$x_{2,i,z}$ = normalized $(v_{i,z} - p_{i,z})$
Ad memorability	MemNet [12]	Ad image	Memorability score	$x_{3,i,z}$ = normalized score
Ad CTR	Given by data	CTR	CTR	$x_{4,i,z}$ = normalized CTR
Contextual relevance	TakeLab [26]	Title, keywords, description	Semantic similarity score	$x_{5,i,z}$ = normalized score
Ad saliency	MBS [30]	Web page snapshot, ad image	Saliency map and ratio	$x_{6,i,z}$ = normalized ratio

ads located on the bottom and right area [22]. By introducing visual saliency, we can increase the probability that the user will notice the displayed ad intentionally or unintentionally.

**Image memorability** measures how likely the user will recall the ad correctly. It is an important metric for ad branding conception. It has been shown that image memorability is a stable and intrinsic property of images that is shared across different viewers. By introducing image memorability, the advertiser's benefit will be improved.

To demonstrate the effectiveness of our introduced multimedia metrics, we proposed a novel multimodal framework for contextual video advertising [7]. Given an online video and a set of ad candidates, we first roughly identify a set of relevant ads based on textual information matching. We then carefully select a set of candidates based on the visual content matching. In this regards, our selected ads are contextually relevant to the video in terms of both textual information and visual content. We finally select the most salient ad as the most appropriate one. We further extend the proposed framework by considering the image memorability.

In the user-study experiment, we selected 20 popular YouTube videos as our test video set, and we collected the products information from Amazon and Taobao to construct the ad dataset. We invited 80 participants to join our experiment, and we used an eye-tracker device to record their eye gaze information. Since user's behavior is sensitive to the purpose of the experiment, they did not have prior knowledge about the experiment. We also built a browser interface which is similar to YouTube to simulate the real-world online video viewing environment. The experimental result is shown in Fig. 2. By comparing the saliency-based advertising strategy to the text-based and visual-based strategy, we can find that: the average fixation duration on the displayed ad increases with introducing visual saliency. By comparing the memorability enhanced advertising to the text-based and visual-based strategy, we can find that: the average ad recall rate increases with considering image memorability. By comparing the memorability-enhanced advertising to saliency-based advertising, we observe an interesting finding: higher fixation duration does not ensure higher ad recall rate. This finding indicates that the memorability is an intrinsic property of images, which is consistent with the results in [12].

### 3.2 Optimal Re-ranking Model

Besides the multimedia metrics discussed in the previous section, we further consider three metric variables that is commonly discussed in online advertising: the publisher's revenue, the advertiser's utility and the ad CTR. Revenue is always the primary concern in display advertising. If advertiser is selected as the winning

bidder, the revenue is her paid price, which equals to the bidding price of the next bidder. The utility shows the advertiser's cost saving. It is defined as the difference between her value and payment. The ad CTR is the more direct variable to show the effectiveness of online advertising. It can be obtained from the historical data. Table 1 specifies all the metric variables as the inputs of the second stage.

All advertisers are re-ranked in the second stage. Given a set of advertisers in auction  $z$ , the rank score of advertiser  $i$ ,  $i \in z$ , is defined as follows:

$$rs_{i,z} = \sum_{k \in K} \omega_k x_{k,i,z}, \quad (1)$$

where  $x_{k,i,z}$  is the input value of metric variable  $k$ ,  $\omega_k$  is the corresponding weight.

The optimal weights can be obtained from the historical data. The weights of variables in the rank score represent the publisher's preferences towards the metrics. As described earlier, the publisher needs to sacrifice a certain amount of revenue in the short term in order to increase the benefits of other stakeholders. Let  $\theta_1$  be the maximum pre-determined loss rate of revenue and let  $\theta_k, \forall k \in K \setminus \{1\}$ , be the minimum increase rate of other variables. Given the maximum loss and the minimum increase targets, we can obtain the optimal weights from the training set  $\tilde{Z}$ .

Our algorithm can be expressed as follows:

$$\max_{\omega_1, \dots, \omega_6} \sum_{z \in \tilde{Z}} rs_{*,z}, \quad (2)$$

$$\text{subject to } 0 \leq \omega_k \leq 1, \forall k \in K, \quad (3)$$

$$\sum_{k \in K} \omega_k = 1, \quad (4)$$

$$|\xi_{1,\tilde{Z}}| \leq |\theta_1|, \theta_1 \leq 0, \quad (5)$$

$$\xi_{k,\tilde{Z}} \geq \theta_k, \theta_k \geq 0, \forall k \in K \setminus \{1\}, \quad (6)$$

where  $rs_{*,z}$  is the rank score of the selected advertiser in auction  $z$  by our proposed model,  $\xi_{k,\tilde{Z}}$  is the mean of changes of variable  $k$  between our proposed model and the ground truth, defined by

$$\xi_{k,\tilde{Z}} = \frac{\sum_{z \in \tilde{Z}} (x_{k,*,z} - x_{k,-,z})}{\sum_{z \in \tilde{Z}} x_{k,-,z}}, \quad (7)$$

where  $x_{k,*,z}$  is the input value of metric variable  $k$  for the selected advertiser in auction  $z$  by our proposed model,  $x_{k,-,z}$  is the input value of metric variable  $k$  for the selected advertiser in auction  $z$  in the ground truth.

The optimal weights maximize the sum of rank scores of the select advertisers from all auctions in the training set. Eqs. (3)-(4)

ensure each variable has an impact in the re-ranking but its impact has an upper bound. Eqs. (5)-(6) further specify the lower bounds of trade-offs where the maximum decrease of the publisher's revenue and the minimum increases for other variables.

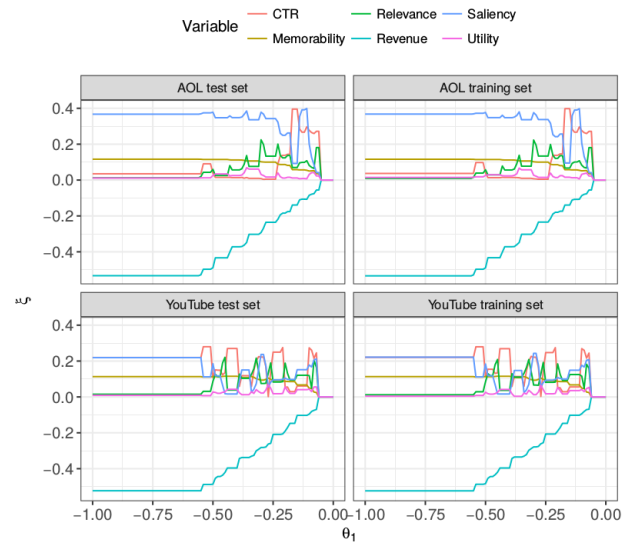
## 4 EXPERIMENT

To validate the proposed framework, both bidding information (e.g., bid price, ad CTR) and multimedia information (e.g., text for contextual relevance matching, ad image for visual saliency and image memorability) for each ad are needed. We use two independent datasets to simulate the real-world advertising system: the multimedia dataset (data collected from AOL and YouTube) represents the interaction between the publisher and the user, and the auction dataset shows the interplay between the publisher and the advertiser. We compare the proposed framework with the GSP auction model, which has been widely used in existing RTB system. Since revenue is always the major concern in online advertising, it will be interesting to investigate how the loss of revenue will affect the other metrics. In this experiment, we set  $\theta_1$  to a non-positive value and  $\theta_k = 0, k = 2 \dots 6$ . We conduct 10-fold cross validation on the both AOL and YouTube dataset and the results are shown in Fig 3.

From this figure, we can conclude that there exist trade-offs among stakeholders in the proposed framework, and the trade-offs change with the publisher's specific needs (the threshold values  $\theta_k, k = 1 \dots 6$ ). We have the following observations: 1) the changes of variables in the test set is similar to that of the training set on both AOL and YouTube multimedia datasets. The consistent results confirm our cross-validation; 2) when  $\theta_1$  is close to 0, all variables remain unchanged. This is because our weight determination model (see section 3.2) is not able to find a solution to the constraints. In other words, the publisher is too stingy to lose revenue and too greedy to achieve improvement on the other variables. In this scenario, we will use the traditional RTB system to select the ad with highest bid price; 3) when  $\theta_1$  further decreases, the revenue follows a monotone decreasing pattern and the other variables increase to various degrees. According to our optimal weights determination model, the solution space will increase with the decrease of  $\theta_1$ . We will select the optimal weights with the highest accumulated rank score; 4) when  $\theta_1$  decreases below a threshold value, the changes of metrics remains stable. Note that, the optimal weights changes with the given threshold values. Through analyzing the historical data and the changes of variables will help the publisher decide a threshold of revenue loss.

## 5 WORK IN PROGRESS

The framework described above is only applicable to the situation where each webpage has a single ad slot. We are currently working on a generalized framework where a webpage has multiple ad slots. Recently, we have constructed another multimedia dataset from Yahoo and MSN. We find that: the percentage of single slot webpage is about 18.3% (3,390 out of 18,447) in Yahoo, and 13.8% (1,484 out of 10,717) in MSN, respectively. The multi-slot webpage makes the computational framework more complex. In the current RTB setting, these ad slots are sold separately through individual auctions [3]. The following scenarios are likely to happen: 1) competitive ads are displayed on the same webpage (e.g., Apple iPhone



**Figure 3: The effect of  $\theta_1$  on the changes of variables, where X-axis represents the decrease threshold of revenue ( $\theta_1$ ) and Y-axis represents the changes of variables when compared with existing RTB system.**

7 and Samsung Galaxy S8 are displayed in the same webpage about mobile techniques). The competitive ads have a negative effect on user's brand conception [4]; 2) an ad with a higher bid price will be placed at a less salient slot. The eye-tracking experiment has shown that user's web surfing eye-gaze follows the 'F' shape [11]. This finding indicates the importances of different ad slot locations within the webpage. We will propose a computational framework that considers both the webpage-ad context and the ad-ad pair context. The limitations above can be acted as the hard/soft constraints.

## 6 CONCLUSION

In this paper, we propose a computational framework to improve the effectiveness of display advertising by incorporating multimedia metrics. The metrics are: the contextual relevance to ensure user's online experience as well as increase the ad CTR, the visual saliency to draw user's attention and the image memorability to enhance user's brand conception. We first conduct an eye-tracking experiment to demonstrate the effectiveness of the introduced metrics. After that, we discuss how to integrate the multimedia metrics into RTB. We also propose an optimization model to determine the optimal weights of the metrics. Our experimental results show that the proposed framework is able to increase the benefits of the advertiser and the user with just a slight decrease in the publisher's revenue. However, better engagements of advertisers and users will increase the demand of advertising and supply of webpage visits, which will boost the publisher's revenue in the long run.

## 7 ACKNOWLEDGEMENT

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centre in Singapore Funding Initiative.

## REFERENCES

- [1] L. Alwitt and A. Mitchell. *Psychological processes and advertising effects*. Lawrence Erlbaum, 1985.
- [2] Y. Bachrach, S. Ceppi, I. Kash, P. Key, and D. Kurokawa. Optimising trade-offs among stakeholders in ad auctions. In *Proceedings of the 15th ACM Conference on Economics and Computation*, pages 75–92, 2014.
- [3] O. Ben-Zwi, M. Henzinger, and V. Loitzenbauer. Ad exchange: Envy-free auctions with mediators. In *The 11th International Conference on Web and Internet Economics*, pages 104–117, 2015.
- [4] R. R. Burke and T. K. Srull. Competitive interference and consumer memory for advertising. *Journal of consumer research*, 15(1):55–68, 1988.
- [5] P. Chatterjee, D. Hoffman, and T. Novak. Modeling the clickstream: implications for web-based advertising efforts. *Marketing Science*, 22(4):520–541, 2003.
- [6] X. Chen, B. Chen, and K. Mohan. Optimizing trade-offs among stakeholders in real-time bidding by incorporating multimedia metrics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017.
- [7] X. Chen, T. V. Nguyen, and M. Kankanhalli. Salad: A multimodal approach for contextual video advertising. In *2015 IEEE International Symposium on Multimedia*, pages 211–216, 2015.
- [8] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, 2007.
- [9] J. Guo, T. Mei, F. Liu, and X.-S. Hua. Adon: an intelligent overlay video advertising system. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 628–629. ACM, 2009.
- [10] Hubspot. Horrifying display advertising. <https://blog.hubspot.com/marketing/horrifying-display-advertising-stats>, 2015.
- [11] N. Jakob. F-shaped pattern for reading web content. *Alertbox: Current Issues in Web Usability*, 2006.
- [12] A. Khosla, A. Raju, A. Torralba, and A. Oliva. Understanding and predicting image memorability at a large scale. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2390–2398, 2015.
- [13] S. Lahaie and P. McAfee. Efficient ranking in sponsored search. In *Proceedings of the 7th International Conference on Internet and Network Economics*, pages 254–265, 2011.
- [14] S. Lahaie and D. Pennock. Revenue analysis of a family of ranking rules for keyword auctions. In *Proceedings of the 8th ACM Conference on Electronic Commerce*, pages 50–60, 2007.
- [15] L. Li, T. Mei, X. Niu, and C. Ngo. PageSense: style-wise web page advertising. In *Proceedings of the 19th International Conference on World Wide Web*, pages 1273–1276, 2010.
- [16] W.-S. Liao, K.-T. Chen, and W. 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. ACM, 2008.
- [17] A. Likhodedov and T. Sandholm. Auction mechanism for optimally trading off revenue and efficiency. In *Proceedings of the 4th ACM conference on Electronic commerce*, pages 212–213, 2003.
- [18] S. McCoy, A. Everard, P. Polak, and D. Galletta. The effects of online advertising. *Communications of the ACM*, 50(3):84–88, 2007.
- [19] T. Mei, X. Hua, L. Yang, and S. Li. VideoSense: towards effective online video advertising. In *Proceedings of the 15th ACM International Conference on Multimedia*, pages 1075–1084, 2007.
- [20] T. Mei and X.-S. Hua. Contextual internet multimedia advertising. *Proceedings of the IEEE*, 98(8):1416–1433, 2010.
- [21] T. Mei, L. Li, X. Hua, and S. Li. ImageSense: towards contextual image advertising. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 8(1):6, 2012.
- [22] J. Owens, B. Chaparro, and E. Palmer. Text advertising blindness: the new banner blindness? *Journal of Usability Studies*, 6(3):172–197, 2011.
- [23] D. Parkes. *Algorithmic Game Theory*, chapter Online Mechanisms. Cambridge University Press, 2007.
- [24] PWC. IAB Internet advertising revenue report: 2016 full year results. Technical report, IAB, 2016.
- [25] P. Sajjacholapunt and L. Ball. The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness. *Frontiers in Psychology*, 5, 2014.
- [26] F. Šarić, G. Glavaš, M. Karan, J. Šnajder, and B. Bašić. Takelab: systems for measuring semantic text similarity. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics*, pages 441–448, 2012.
- [27] S. H. Sengamedu, N. Sawant, and S. Wadhwa. vadeo: video advertising system. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 455–456. ACM, 2007.
- [28] D. Thompson and K. Leyton-Brown. Revenue optimization in the generalized second-price auction. In *Proceedings of the 14th ACM conference on Electronic Commerce*, pages 837–852, 2013.
- [29] X.-J. Wang, M. Yu, L. Zhang, R. Cai, and W.-Y. Ma. Argo: intelligent advertising by mining a user's interest from his photo collections. In *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*, pages 18–26. ACM, 2009.
- [30] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech. Minimum barrier salient object detection at 80 fps. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1404–1412, 2015.