Advances in Cognitive Sciences
Vol 24, No. 3, Autumn 2022
Research Article
Pages 88-104

Using computational visual attention models in evaluating audience attention in educational multimedia

Majid Shabani[1], Alireza Bosaghzadeh[2*], Reza Ebrahimpour[3], Seyed Hamid Amiri[2], Keyhan Latifzadeh[1]

1. MS, Artificial Intelligence, Department of Artificial Intelligence, Faculty of Computer Engineering, Shahid Rejaee Teacher Training University, Tehran, Iran
2. Assistant Professor of Artificial Intelligence, Department of Artificial Intelligence, Faculty of Computer Engineering, Shahid Rejaee Teacher Training University, Tehran, Iran
3. Professor of Artificial Intelligence, Department of Artificial Intelligence, Faculty of Computer Engineering, Shahid Rejaee Teacher Training University, Tehran, Iran

Recieved: 16 Dec. 2021
Revised: 21 Jul. 2022
Accepted: 21 Jul. 2022

Keywords
Visual attention
Educational multimedia
Saliency map

Corresponding author
Alireza Bosaghzadeh, Assistant Professor of Artificial Intelligence, Department of Artificial Intelligence, Faculty of Computer Engineering, Shahid Rejaee Teacher Training University, Tehran, Iran

Email: A.bosaghzadeh@sru.ac.ir

doi.org/10.30514/icss.24.3.88

Abstract

Introduction: With the advancement of knowledge and technology, multimedia service for education has been developed. One of the most overlooked issues is the public's attention to adapting the produced multimedia to their original design goals. Computational visual attention models of human brain-inspired can fill this gap and serve as a powerful tool for evaluating educational multimedia audience points of interest.

Methods: This multimedia educational study was designed with the help of the computational visual attention models to evaluate and determine what parts the human user will pay attention to when watching the video. The present study aimed to identify the locations of audience's attentionusing computational models of visual attention and compare them with the data obtained from an eye-tracking system to increase its impact on the audience and improve the quality of the produced multimedia.

Results: Experiments show that by evaluating visual attention locations based on eye data and locations determined by the computational visual attention models; in addition, to qualitatively evaluating the effects of the applied principles, the amount of audience's attention produced can be adjusted according to the model of the attention produced in the video. As a result, by reducing cognitive load, the multimedia effect of education will be increased.

Conclusion: The simulation results revealed that existing computational models of acceptable accuracy could determine the location of visual attention, which facilitates evaluating educational multimedia produced.

Citation: Shabani M, Bosaghzadeh AR, Ebrahimpour R, Amiri H, Latifzadeh K. Using computational visual attention models in evaluating audience attention in educational multimedia. Advances in Cognitive Sciences. 2022;24(3):88-104.

Extended Abstract

Introduction

With the advancement of knowledge and technology, multimedia service for education has developed. The cognitive theory of multimedia learning proposes the principle that helps multimedia designers and e-learning in produce optimal textual, graphic, visual, and auditory presentations. Each principle is based on comparingthe results of multimedia learning research in different situations and determines how much each is effective in bet-

tering students' learning.

The current research used five principles of principles mentioned in the Mayer multimedia learning book. In recent decades, many scientific studies have been aimed at modeling computational mechanisms in concentration orientation. One of the most overlooked issues is the audience's attention to adaptingthe produced multimedia to their original design goals. In this research, using visual attention models, this study try to predict the audience's attention to increase the educational impact of these videos by providing scientific solutions to improve the quality of educational multimedia.

Computational visual attention models of human brain-inspired can fill this gap and assist as a powerful tool for evaluating educational multimedia audience points of interest. The models used in this study are the four types of bottom-up visual attention with the best performance (6, 10, 20, 23), effectively improving and enhancing the educational quality in multimedia production.

Methods

In this study, the tests using the bottom-up visual attention models on educational multimedia based on five principles of the 12 Mayer principles (Five principles of Mayer are: Coherence principle to remove sub-elements and unnecessary elements, the Signaling Principle for signs to make the essential elements more salient, Redundancy principle through graphics and narrative to learn better, Spatial resolution providing words and images closely together on one page, and Temporal resolution providing words and images close together at the same time). Besides, the results were examined based on each model and type of experiment. Multimedia was evaluated to determine the locations of the audience's attention by the eye-tracking, and its results were used to evaluate the accuracy of the forecast results by computational models of visual attention.

This multimedia educational study designed with the help of the computational visual attention models is evaluated to determine what parts the human user will pay attention to when watching the video. Using computational models of visual attention, the current study attempt to identify the locations of the audience's attentionand compare them with the data obtained from an eye-tracking system to increase its impact on the audience and improve the quality of the produced multimedia.

Visual attention computational models are typically confirmed compared to the eye movements of human observers. Eye movements convey essential information about cognitive processes, such as reading, visual search, and scene perception. Accordingly, this study assumes that there is a model that produces a saliency map S, then compares and evaluates it with eye movements G (or fixation by the human eye). The evaluation of criteria used (including AUC, LCC, NSS, and SSIM) to determine the accuracy of predicting the locations determined by the model with the audience's locations.

Each of the computational models of visual attention is implemented and evaluated based on the visual saliency obtained from the multimedia produced based on the first five principles of Mayer. The values obtained are based on the above criteria in the work method for each shot (the sum of consecutive frames to create a scene).

Notably, the multimedia dimensions are 640*480 pixels and 30 frames per second.

Results

The ground-truth saliency map (GSM) for each frame was obtained by the three participants (visual health), who viewed a five-minute video (30 frames per second) in free observation mode. The video contains 48 shots and 5069 frames. This way, wherever the audience saccades on the point, it is the point of fixation in GSM. Then, for a more significant adaptation to the human vi-

sion region, a Gaussian filter is applied to it, and the final GSM is produced for each frame.

Some features were used, including color, intensity, orientation, and motion (obtained from the optical flow method). Furthermore, in some scenes, around some areas of the image, no visual attention is attracted and all the user's attention to the central part is for two reasons; The first reason is the lack of information in some areas of the image (center-surround difference), and the second reason is that individuals focus on the center of the image (center bias).

To examine the signaling principle in which the designer aimed at attracting the viewer's attention to the expression, the user's attention is attracted to the desired mark. The reasons for the audience's attention are the motion of the mark and the center-surround difference. These demonstrated that by better marking, the audience can be better attracted to the desired location and increase the learning effect.

The two consecutive frames in which the letters appear in time and place attracts visual attention, which is the effect of the object's motion in the two consecutive frames and follows the audience's desired content based on the spatial and temporal resolution principle.

When in the principles of signaling, coherence, and redundancy, the purpose of the multimedia designer is to attract the audience's attention to parts of the text and image. However, due to the lack of difference between the background and the thinness of the text and the sign, the audience's visual attention is not well received, and using visual attention models, changes such as creating a difference with the surroundings, proximity to the center, and motion the frames are made that attract better attention.

Overall, evaluated frames with the saliency maps of each model used for each shot, the highest adaptation to human eye data. In addition, the best performance in frames with a center-surround difference on the Itti et al. model (10), in the center bias frames of the GBVS model (23), and the frames with motion, and the same superpixels are the Wang model (20).

Conclusion

This study used models available in bottom-up visual attention to improve educational content. According to the obtained results, the computational models of visual attention can be used as a criterion for predicting the attractive vision of the audience. Although the models do not have the same results, they have a good performance and are close together in predicting areas of human attention, it is essential to increase the quality of educational multimedia using the possibility of these models.

Using visual or combination models of vision or combination of the audience can be anticipated, and by putting the necessary content in the pre-explained location by the model of visual attention, quality, and multimedia impact can be increased.

As can be seen in the experiments, a model alone does not have high efficiency in all cases, and the models have different results depending on the type of algorithm and the features used. By combining mentiond models, they can help improve the ultimate performance.

In the following, the current research strives to improve the values of these criteria by improving the above-mentioned model or creating a new computing model to adapt to actual data. Correspondingly, examining brain signals and the effect of attracting attention methods on cognitive reduction, as well as the provision of software that can be used to evaluate the multimedia education produced can be considered for future tasks.

Ethical Considerations

Compliance with ethical guidelines

This article is taken from the Master's Thesis of the first author. The present study was conducted in accordance with ethics such as the confidentiality of the participants

and the participants were free to leave the study. In this study, sufficient information on the research is provided and the results of honest, accurate and complete research have been published.

Authors' contributions

Majid Shabani: The first author that was responsible for researching, implementing, and analyzing the results. Alireza Bosaghzadeh: Corresponding author guiding implementing the research and reform of the article. Reza Ebrahimpour: Provided guidance in the method of working and data analysis. Seyed Hamid Amiri: Guiding implementing the research and reform of the article. Keyhan Latifzadeh: Responsible for researching and collecting samples.

Funding

This article was supported by the Cognitive Sciences & Technologies Council with Research Code 6880, approved 08/10/1397, and the research project of Shahid Rajaee Teacher Training University with contract number 39110.

Acknowledgments

The authors thank all the participants in this study, who had ongoing participation, and respectable professors who provided guidance and advice.

Conflict of interest

The authors declare no conflicts of interest.

مجله تازه‌های علوم شناختی
دوره ۲۴، شماره ۳، پاییز ۱۴۰۱
مقاله پژوهشی
صفحات ۱۰۴-۸۸

استفاده از مدل‌های محاسباتی توجه بینایی در ارزیابی توجه مخاطب در چندرسانه‌ای‌های آموزشی

مجید شعبانی[۱]، علیرضا بساق‌زاده[۲*]، رضا ابراهیم‌پور[۳]، سید حمید امیری[۲]، کیهان لطیف‌زاده[۱]

۱. دانشجوی کارشناسی ارشد هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، تهران، ایران
۲. استادیار هوش مصنوعی، گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، تهران، ایران
۳. استاد هوش مصنوعی، گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، تهران، ایران

دریافت: ۱۴۰۰/۰۹/۲۵
اصلاح نهایی: ۱۴۰۱/۰۴/۳۰
پذیرش: ۱۴۰۱/۰۴/۳۰

واژه‌های کلیدی
توجه بینایی
چندرسانه‌ای آموزشی
نقشه برجستگی

نویسنده مسئول
علیرضا بساق‌زاده، استادیار هوش مصنوعی، گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، تهران، ایران
ایمیل: A.bosaghzadeh@sru.ac.ir

doi.org/10.30514/icss.24.3.88

چکیده

مقدمه: با پیشرفت دانش و فن‌آوری، استفاده از چندرسانه‌ای‌ها برای آموزش گسترش زیادی یافته است. یکی از مهمترین مسائلی که مورد غفلت واقع شده است مکان توجه مخاطب در راستای انطباق چندرسانه‌ای‌های تولید شده با اهداف اولیه طراحی آنها است. مدل‌های محاسباتی توجه بینایی الهام گرفته از عملکرد مغز انسان می‌توانند این خلاء موجود را پر کنند و به عنوان ابزاری قدرتمند جهت ارزیابی نقاط توجه مخاطب چندرسانه‌ای آموزشی به کار گرفته شوند.

روش کار: در این مقاله چندرسانه‌ای آموزشی طراحی شده، توسط مدل محاسباتی توجه بینایی مورد ارزیابی قرار می‌گیرد تا معلوم شود که کاربر انسانی در هنگام مشاهده فیلم به چه بخش‌هایی توجه خواهد کرد. با بهره‌گیری از مدل‌های محاسباتی توجه بینایی سعی بر کشف مکان‌های توجه مخاطب و مقایسه آن با داده‌های ردیابی چشمی گردیده تا در جهت افزایش میزان تاثیر آن بر مخاطب و بهبود کیفیت چندرسانه‌ای تولید شده استفاده گردد.

یافته‌ها: در آزمایش‌های صورت گرفته مشخص می‌گردد که با ارزیابی مکان‌های توجه بینایی با داده‌های چشمی و مکان‌های تعیین شده توسط مدل محاسباتی توجه بینایی علاوه بر ارزیابی کیفی تاثیرات اصول به کار برده شده می‌توان با تغییراتی متناسب با مدل توجه در ویدئو تولید شده میزان توجه مخاطب را بالا برد و در نتیجه، با کاهش بار شناختی، اثربخشی چندرسانه‌ای آموزشی را افزایش داد.

نتیجه‌گیری: نتایج شبیه‌سازی نشان می‌دهد که مدل‌های محاسباتی موجود بادقت قابل قبولی می‌توانند محل توجه بینایی را تعیین کنند که این موضوع باعث سهولت ارزیابی چندرسانه‌ای‌های آموزشی تولید شده می‌شود.

مقدمه

امروزه با ظهور اینترنت و پیشرفت‌های رایانه‌ای، روش‌های آموزش و یادگیری دچار تغییرات اساسی شده‌اند و از طرفی استفاده از آموزش رایانه‌ای، روش‌های یادگیری الکترونیکی، یادگیرنده‌ها و هوشمندسازی اثر مضاعف داشته‌اند (۱، ۲). یکی از این تحولات استفاده از نرم‌افزارها و ارائه‌های چندرسانه‌ای در امر آموزش است و به خصوص در آموزش زبان‌های خارجی مانند زبان پرکاربرد انگلیسی می‌تواند کارآمد باشد. طراحی و ساخت ارائه‌های چندرسانه‌ای اگر بدون استفاده از اصول خاصی باشد نه تنها ممکن است که کارآمد نباشد بلکه امکان اختلال در امر آموزش را نیز موجب می‌شود. برای طراحی ارائه‌های چندرسانه‌ای اصول دوازده‌گانه‌ای که از نظریه یادگیری شناختی چندرسانه‌ای برآمده‌اند (۳)، وجود دارد، رعایت هر یک موجب افزایش کیفیت یادگیری چندرسانه‌ای خواهد شد.

نظریه شناختی یادگیری چندرسانه‌ای اصولی را مطرح می‌کند که طراحان چندرسانه‌ای و آموزش الکترونیک را در ساخت ارائه‌های بهینه متنی، گرافیکی، تصویری و شنیداری کمک می‌کند. هر اصل، برآمده از مقایسه نتایج پژوهش‌های یادگیری چندرسانه‌ای در شرایط گوناگون است و مشخص می‌کند که هر یک به چه میزان در یادگیری بهتر دانش‌آموزان مؤثر است. دوازده اصلی که در کتاب یادگیری چندرسانه‌ای Mayer (۳) آمده عبارت است از: اصل انسجام، اصل نشانه‌گذاری، اصل افزونگی، اصل مجاورت فضایی، اصل مجاورت زمانی، اصل طبقه‌بندی، اصل پیش‌آموزش، اصل کیفیت، اصل چندرسانه‌ای، اصل شخصی‌سازی، اصل صدا، اصل تصویر.

علاوه بر این، از جمله عوامل مؤثر در طراحی چندرسانه‌ای به کارگیری صحنه‌هایی مناسب جهت افزایش توجه بینایی انسان می‌باشد. جریانی از داده‌های بینایی (10^{9}-10^{8} بیت) در هر ثانیه به چشم ما می‌رسد (۴، ۵)، توانایی سیستم بینایی انسان برای تشخیص برجستگی بینایی (Visual saliency) فوق‌العاده سریع و قابل اطمینان است. با این حال، مدل‌سازی محاسباتی این رفتار هوشمند اساسی همچنان یک چالش است (۳۱). از طرفی پردازش این داده‌ها در زمان واقعی بدون کمک سازوکارهای هوشمند برای کاهش میزان داده‌های بینایی نادرست کار بسیار دشواری است. فرآیندهای شناختی و پیچیده سطح بالا مانند تشخیص شی یا تفسیر صحنه، می‌توانند به گونه‌ای تبدیل شوند که قابل تشخیص باشند. این سازوکارهای تبدیل تحت عنوان توجه بینایی (Visual attention) شناخته می‌شوند (۶). در انسان‌ها، توجه توسط شبکیه چشم تسهیل شده است که شامل حفره مرکزی (Central fovea) با وضوح بالا و اطراف آن با وضوح پایین می‌باشد. در واقع توجه بینایی ساختاری را فراهم می‌کند که به وسیله آن به بخش‌های مهم صحنه که حاوی اطلاعات مهم‌تری می‌باشد بیشتر تمرکز شده و اطلاعات دقیق‌تری جمع‌آوری شود (۷).

یکی از مسائل مهم در ساخت چندرسانه‌ای‌های آموزشی انتقال بهتر مفاهیم به مخاطب با تاثیر بیشتر در زمان کمتر و با صرف انرژی کمتر می‌باشد، در این مقاله با استفاده از مدل‌های توجه بینایی سعی بر پیش‌بینی نواحی مورد توجه مخاطب می‌شود تا با ارائه راه‌کارهای علمی در جهت بهبود کیفیت چندرسانه‌ای‌های آموزشی میزان تاثیر آموزشی این نوع ویدئوها افزایش یابد.

در دهه‌های اخیر، بسیاری از پژوهش‌های علمی با هدف مدل‌سازی سازوکارهای محاسباتی در جهت‌دهی تمرکز صورت گرفته است. نوروبیولوژیک‌ها نشان داده‌اند که چگونه نورون‌ها خود را سازگارتر می‌کنند تا اشیاء مورد نظر را بهتر نمایش دهند (۸). دانشمندان علوم اعصاب محاسباتی مدل‌های شبکه عصبی را برای شبیه سازی و توضیح رفتارهای توجه ساخته‌اند (۹). با الهام از این مطالعات، دانشمندان حوزه روباتیک و بینایی رایانه‌ای تلاش کرده‌اند تا مسائل اساسی پیچیدگی محاسباتی را برای ساخت سامانه‌هایی که بتوانند توجه بینایی را در زمان واقعی تشخیص دهند (۱۰، ۱۱) حل کنند. برای بررسی مدل‌های توجه از دیدگاه‌های روان‌شناختی، نوروبیولوژیکی و محاسباتی می‌توان مطالعه نمود (۱۲، ۱۳، ۳۱، ۳۲).

موضوع اصلی در مدل‌سازی توجه بینایی، چگونگی، زمان و چرایی انتخاب مناطقی از تصویر به عنوان برجستگی می‌باشد. با توجه به این عوامل، چندین تعریف و دیدگاه محاسباتی در کاربردهای مختلف ارائه شده است که بارزترین آنها، الهام گرفتن از آناتومی و عملکرد سیستم بینایی اولیه انسان می‌باشد (۱۰، ۱۱). علاوه بر این، توجه یک مفهوم کلی، شامل تمام عواملی است که بر سازوکارهای انتخاب تاثیر می‌گذارند، چه این که از پایین به بالا به نظر بیایند یا از بالا به پایین مورد انتظار باشند (۷، ۳۱، ۳۲).

در ادامه ابتدا برخی از مدل‌های توجه بینایی و معیارهای ارزیابی تشریح می‌شود و سپس این مدل‌ها را بر روی چندرسانه‌ای طراحی شده اجرا کرده و نتایج آن در راستای اصول طراحی چندرسانه‌ای ارزیابی و مورد بررسی قرار می‌گیرد.

تقریبا تمام مدل‌های توجه به طور مستقیم یا غیرمستقیم از مفاهیم شناختی الهام گرفته شده‌اند. به طور معمول الگوریتم‌های توجه بینایی می‌توانند به سه مرحله تقسیم شوند: استخراج: استخراج بردارهای ویژگی در مکان‌هایی بر روی صفحه تصویر یا فریم که معمولا با استفاده از فیلترهای الهام گرفته از زیست‌شناسی انجام می‌شود. به طور عمده شامل ویژگی‌های سطح پائین مانند: شدت روشنایی، تغییر جهت، رنگ، حرکت و ویژگی‌های سطح بالا مانند: رنگ پوست، چهره (۱۰، ۱۴، ۱۵، ۳۳). فعال‌سازی: از طریق نقشه فعال‌سازی با استفاده از بردارهای ویژگی انجام می‌شود که معمولا با تفریق نقشه‌های ویژگی در مقیاس‌های مختلف (به عنوان مثال «Center-Surround» برای «مرکز_محور») همراه می‌باشد. نرمال‌سازی/ترکیب: نرمال‌سازی نقشه فعال‌سازی (به دنبال ترکیب نقشه‌ها در یک نقشه) که به طور معمول به وسیله یکی از سه روش ذیل انجام می‌شود: ۱. الگوریتم نرمال‌سازی بر اساس حداکثر محلی (۱۰) ("max-ave")، ۲. الگوریتم تکراری بر مبنای کانولوشن با یک فیلتر DoG و ۳. تعامل غیرخطی که مقادیر محلی را با میانگین وزنی از مقادیر اطراف تقسیم می‌کند به شیوه‌ای که به شکل داده‌های روان‌شناختی مطابقت دارد (۱۶). برای دسته‌بندی و تفکیک مدل‌های توجه و شناخت بهتر آن فاکتورهای دیگر نیز وجود

دارد که دامنه کاربرد مدل‌های مختلف را تعیین می‌کند (۷، ۳۴).

از جمله عوامل موثر در فرآیند توجه بینایی، تفاوت مرکز_محور است که الهام گرفته از پاسخ‌های عصبی در (Lateral Geniculate Nucleus) و قشر V1 است. رنگ‌ها مانند کانال‌های رنگی قرمز_سبز و آبی_زرد الهام گرفته از نورون‌های رنگ متضاد در قشر V1 پردازش می‌شود و هر چه میزان این اختلاف بیشتر باشد توجه انسان بهتر جلب می‌گردد (۱۷). یکی دیگر از عوامل مؤثر در توجه بینایی تمایل به مرکز می‌باشد که عناصر نزدیک به مرکز توجه بینایی را بیشتر جلب می‌کنند (۱۸). همچنین میدان دید که میدان پاسخ‌گویی نورون‌ها (Receipt field) است در توجه انسانی مؤثر می‌باشد.

در این مقاله برای استخراج برجستگی‌های بینایی و پیش‌بینی مکان‌های توجه بینایی از چند مدل طراحی شده در سال‌های اخیر و همچنین ترکیبی برخی از آنها با مدل طراحی شده توسط آقای Wang و همکارانش (۱۹، ۲۰) استفاده می‌شود. در ادامه مدل‌های محاسباتی توجه بینایی مورد استفاده در این مقاله به طور خلاصه تشریح می‌گردد.

مدل پایه Itti و همکاران از سه کانال ویژگی، رنگ، شدت و جهت‌گیری استفاده می‌کند (۱۰). این مدل، مبنایی برای مدل‌های بعد و معیار استانداردی برای مقایسه شده و نشان داده شده است که با حرکات چشم انسان در کارهای مشاهده آزاد مرتبط می‌باشد (۲۱). ابتدا یک تصویر یا فریم ورودی به یک اهرام گاوسی وارد می‌شود و هر سطح هرم با σ مناسب به کانال‌های رنگی (شامل قرمز (R)، سبز (B)، آبی (B)، زرد (Y))، شدت (I) و جهت‌گیری‌های محلی (O_e) تجزیه می‌شود. از این کانال‌ها، "نقشه‌های ویژگی" مرکز_محور (مرکز C و محور S) برای ویژگی‌های رنگ، شدت و فیلتر گابور (در چهار زاویه) برای جهت‌گیری ساخته و نرمال می‌شوند. در هر کانال، نقشه‌ها در سراسر مقیاس جمع می‌شوند و مجددا نرمال می‌شوند. سپس این نقشه‌ها به صورت خطی به صورت مجزا در هر کانال جمع و نرمال شده و یک «نقشه‌های وضوح» برای هرکانال ایجاد می‌کنند. سپس نقشه‌های وضوح به صورت خطی با هم ترکیب و تشکیل یک نقشه برجستگی را می‌دهند. در انتها، بر اساس مدل شبکه Winner-Take-All به ترتیب نقاط برجسته‌تر انتخاب و سپس مهار می‌شود و به نقطه برجسته دیگر انتقال می‌یابد تا بدین ترتیب تمام نقاط برجسته انتخاب شوند. به طور کلی این مدل در پیش‌بینی توجه بینایی با سازوکار بینایی انسان بسیار منطبق است و به خوبی تفاوت شدت روشنایی، رنگ و مرکز_محور را نمایش می‌دهد و برای شناسایی شیء مفید است (۲۲).

Harel و همکارانش مدل برجستگی بینایی مبتنی بر گراف (GBVS) را معرفی کرده‌اند (۲۳). در این مدل ابتدا نقشه‌های فعال‌سازی در کانال‌های خاصی از ویژگی ایجاد می‌شود و سپس به طریقی که وضوح بیشتر گردد نرمال‌سازی می‌شوند و با دیگر نقشه‌ها ترکیب می‌شوند. این مدل به لحاظ زیست شناختی قابل قبول است و به طور موازی عمل می‌کند. روش کار به این صورت است که ابتدا نقشه‌های ویژگی (M نقشه ویژگی، n ابعاد، R موقعیت برجسته) را در مقیاس‌های مکانی چندگانه استخراج می‌کنند. هرم مقیاس مکانی ابتدا از ویژگی‌های تصویر یا فریم استخراج می‌شود: شدت، رنگ و جهت‌گیری (شبیه به Itti و همکاران (۱۰)). سپس، یک گراف کامل متصل بر روی تمام نقاط شبکه هر نقشه ویژگی ساخته می‌شود (هر پیکسل به عنوان یک گره باهم تشکیل یک شبکه می‌دهند). وزن بین دو گره متناسب با شباهت مقادیر ویژگی و فاصله مکانی آنها است. اختلاف بین دو مکان (i, j) و (p, q) در نقشه ویژگی، با ارزش‌های مربوطه $M(i,j)$ و $M(p,q)$ به صورت $d((i,j)||(p,q)) = \left|log\frac{M(i,j)}{M(p,q)}\right|$ تعریف می‌شود.

لبه‌های جهت‌دار از (i,j) گره به (p,q) گره با توجه به عدم تشابه و فاصله آنها در شبکه M به صورت $q)) = d((i,j)\|(p,q)).F(i-p,j-q)$ وزن‌دهی می‌شوند که F به صورت $w((i,j),(p,$ $exp(-\frac{a^2+b^2}{2\sigma^2})$ به دست می‌آید. $F(a,b) =$

گراف‌های حاصل به عنوان زنجیره مارکف با نرمال کردن وزن لبه‌های خروجی هر گره به ۱ و با تعریف یک رابطه هم‌ارز بین گره‌ها و حالت‌ها، و نیز بین وزن‌های لبه و احتمال انتقال، ایجاد می‌شوند. توزیع تعادل آنها به عنوان نقشه‌های فعال‌سازی ($A(p,q)$) و برجستگی پذیرفته می‌شود. در توزیع تعادل، گره‌های متفاوت نسبت به گره‌های اطراف خود، مقادیر زیادی می‌گیرند. نقشه‌های فعال‌سازی به صورت $w_2((i,j)),((p,q))=A(i-p,j-q)$ نرمال می‌شوند تا بر روی جزئیات مشخص (مکان‌های برجسته) تأکید کنند. سپس نقشه‌های نرمال شده باهم ترکیب و یک نقشه کلی تبدیل می‌شوند. یکی از ویژگی‌های GBVS، «تمایل به مرکز» می‌باشد، زیرا در زمان اجرای الگوریتم گره‌ها به سمت گره مرکزی نزدیک‌تر (که دارای مقدار فعال‌سازی بالاتر می‌باشد) متمایل می‌شوند و این موضوع با تمایل انسان‌ها به تمرکز خیرگی چشم به مرکز تصاویر چه در زمان تصویربرداری و چه مشاهده آزاد شباهت دارد که از نقاط قوت این مدل می‌باشد (۲۳).

Hou و Zhang یک روش ساده برای تشخیص برجستگی بینایی در دامنه فرکانس ارائه می‌دهند (۶). مدل آنها مستقل از ویژگی‌ها، دسته‌ها یا سایر اشکال دانش پیشین از اشیا است. با تجزیه و تحلیل لگاریتم طیف ورودی یک تصویر ورودی، مانده طیفی یک تصویر را در حوزه طیفی استخراج و یک روش سریع برای ساخت نقشه برجستگی مربوطه در

حوزه مکانی پیشنهاد می‌کنند. آنها دریافتند که طیف لگاریتم از تصاویر مختلف یک روند مشابه دارند، هر چند که هر کدام دارای خاصیت‌های آماری منحصر به فردی هستند، همچنین طیف‌های میانگین بر روی ۱، ۱۰ و ۱۰۰ تصویر، یک حالت خطی محلی را در میانگین طیف لگاریتم دارد. از طریق تبدیل فوریه تصویر ورودی، دامنه و فاز مشتق می‌شود. سپس لگاریتم طیف L(f) از تصویر نمونه‌برداری با روش کاهش‌یافته کیفیت (Down-sampled) محاسبه می‌شود. از L(f)، باقی‌مانده طیفی R(f) به صورت $R(f)=(L(f)-h_n(f)*L(f)$ به دست می‌آید، که $h_n(f)$ یک فیلتر متوسط محلی(نرمال ساز) n x n است. با استفاده از تبدیل فوریه معکوس، آنها نقشه برجستگی را در حوزه مکانی ایجاد می‌کنند. ارزش هر نقطه در نقشه برجستگی برای نشان دادن خطای تخمین، مربع می‌شود. در نهایت، آنها نقشه برجستگی را با یک فیلتر گاوسی برای اثر بینایی بهتر نرم می‌کنند. همچنین آنها دریافتند که تغییر مقیاس (برابر با انتخاب اندازه تصویر ورودی) منجر به نتیجه دیگری در نقشه برجستگی می‌شود، استفاده از این مدل به ما کمک می‌کند تا نواحی برجسته در تمام تصویر به صورت سراسری شناسایی گردد.

Wang و همکارانش چارچوبی ارائه دادند که به وسیله توجه بینایی ویدیو را تقسیم‌بندی کرده و اشیایی را از آن استخراج می‌کند (۱۹، ۲۰). فریم ورودی به سه تصویر: نقشه بخش‌بندی بر روی سوپر پیکسل‌ها، نقشه لبه فضایی زمانی به دست آمده از احتمال لبه‌های ایستا و نقشه اندازه شیب جریان نوری تبدیل می‌شود. برای هر سوپر پیکسل، احتمال اشیاء به دست می‌آید و تخمین برجستگی گراف درون هر فریم و بین دو فریم متوالی مشخص (جداسازی) می شود. یک روش چکیده نواحی اسکلت شیء (Skeleton Regions Abstraction Method) که برای به دست آوردن تخمین نهایی برجستگی از طریق جداسازی مناطق اسکلت مرکزی با مقادیر برجستگی بالاتر است، استفاده می‌شود. در نهایت، با ترکیب نقشه برجستگی فضایی مکانی، مدل‌های ظاهرسازی عمومی و مدل‌های موقعیت پویا که از اطلاعات حرکتی در میان چند فریم متوالی استفاده می‌کنند پیکسل‌های تقسیم‌بندی به طور صحیح تولید می‌شوند. در این مدل نقشه برجستگی نهایی بسیار تحت تاثیر حرکت نواحی برجسته می‌باشد از این رو این مدل با هدف استخراج اطلاعات حرکتی در هر دو فریم متوالی مورد استفاده قرار می‌گیرد.

مدل‌های مورد استفاده در این مطالعه از نوع توجه بینایی پایین به بالا (bottom-up) می‌باشد که برای بهبود و بالا بردن کیفیت آموزشی در تولید چندرسانه‌ای های آموزشی بسیار مؤثر می‌باشد که در ادامه نحوه به کارگیری و تاثیرات این مدل‌ها در بهبود چندرسانه‌ای آموزشی تشریح می‌گردد.

روش کار

در این پژوهش آزمایش‌های با استفاده از مدل‌های تشریح شده در مقدمه بر روی چند رسانه‌ای آموزشی ساخته شده بر اساس ۵ اصل از اصول ۱۲گانه Mayer (۳) انجام می‌شود و نتایج بر اساس هر مدل و نوع آزمایش مورد بررسی قرار می‌گیرد. چند رسانه‌ای ساخته شده با در نظر گرفتن اصول Mayer سعی در افزایش میزان یادگیری مخاطب دارد که توسط مدل‌های مذکور ارزیابی می‌گردد. تعریفی اجمالی از پنج اصل ابتدایی از اصول ۱۲گانه Mayer به شرح ذیل می‌باشد:

۱) اصل انسجام (Coherence principle): افراد زمانی که عناصر فرعی حذف شوند، نسبت به زمانی که این عناصر وجود داشته باشد، بهتر یاد می‌گیرند.

۲) اصل نشانه‌گذاری (Signaling principle): افراد بهتر یاد می‌گیرند اگر سرنخ‌هایی برای بارز کردن عناصر ضروری مطالب اضافه شوند.

۳) اصل افزونگی (Redundancy principle): مردم از طریق گرافیک و روایت، بهتر یاد می‌گیرند نسبت به مواقعی که اطلاعات را از طریق گرافیک، روایت و متن چاپ‌شده دریافت کنند.

۴) اصل مجاورت مکانی (Spatial resulotion): دانش‌آموزان بهتر یاد می‌گیرند اگر واژگان و تصاویر متناظر نزدیک به هم در یک صفحه یا نمایش‌گر ارائه شوند نسبت به زمانی که دور از هم قرار گیرند.

۵) اصل مجاورت زمانی (Temporal resolution): دانش‌آموزان بهتر یاد می‌گیرند زمانی که واژه‌ها و تصاویر متناظر به صورت همزمان نمایش داده شوند نسبت به زمانی که به صورت پیوسته نمایش داده می‌شوند.

چندرسانه‌ای ساخته شده جهت تعیین مکان‌های توجه مخاطبان توسط دستگاه ردیابی چشمی (Eye tracker) مورد ارزیابی قرار می‌گیرد و نتایج آن برای ارزیابی میزان صحت نتایج پیش‌بینی شده توسط مدل‌های محاسباتی توجه که در بخش ۲ تشریح گردید مورد استفاده قرار می‌گیرد (۳۵). معیارهای ارزیابی مورد استفاده که در ادامه تشریح می‌گردد تعیین‌کننده میزان دقت پیش‌بینی مکان‌های توجه تعیین شده توسط مدل با مکان‌های مورد توجه مخاطب است. لازم به ذکر است، ابعاد چند رسانه‌ای آموزشی برابر با ۴۸۰*۶۴۰ پیکسل و نرخ ۳۰ فریم در ثانیه می‌باشد.

معیارهای ارزیابی

مدل‌های محاسباتی توجه بینایی به طور معمول در مقایسه با حرکات چشم ناظران انسانی تأیید می‌شوند. حرکات چشم اطلاعات مهمی در مورد فرآیندهای شناختی مانند خواندن، جستجوی بینایی و

ادراک صحنه را انتقال می‌دهند (۲۴). بنابراین فرض می‌کنیم مدلی داریم که یک نقشه برجستگی S را تولید می‌کند، سپس آن را با داده‌های حرکات چشم G (یا موقعیت‌های خیره شده توسط چشم انسان) مقایسه و ارزیابی می‌کنیم، برای این‌کار روش‌های گوناگونی پیشنهاد شده است. یک روش این است که S را به عنوان یک رده‌بندی دودویی در نظر بگیریم و با استفاده از تحلیل تئوری تشخیص سیگنال از محدوده زیر منحنی ROC (AUC) برای ارزیابی عملکرد این طبقه‌بندی استفاده کنیم. همچنین می‌توان S و G را به عنوان متغیرهای تصادفی در نظر گرفت و از ضریب همبستگی خطی (Linear Correlation Coefficient) یا نرمال‌سازی مسیر پیمایش برجستگی (Normalized Scanpath Saliency) برای اندازه‌گیری ارتباطات آماری استفاده کرد (۷). در ادامه به تشریح برخی از معیارهای ذکر شده می‌پردازیم:

الف) محدوده زیر منحنی (AUC): AUC محدوده زیر منحنی مشخصه عملکرد گیرنده (ROC) (۲۵) است که به عنوان محبوب‌ترین معیار عمومی، برای ارزیابی یک سیستم طبقه‌بندی باینری با یک آستانه متغیر (معمولا برای طبقه‌بندی بین دو روش استدلالی و تصادفی) استفاده می‌شود. با استفاده از این معیار، مدل ESM به عنوان یک طبقه‌بند باینری در هر پیکسل در تصویر می‌باشد؛ به عبارت دقیق‌تر، پیکسل‌هایی با مقادیر برجستگی بالاتر از آستانه، به عنوان خیرگی صحیح طبقه‌بندی می‌شوند، در حالی که بقیه پیکسل‌ها به عنوان خیرگی غیرصحیح طبقه‌بندی می‌شوند (۱۵، ۲۶). سپس از خیرگی در دستگاه ردیاب چشمی به عنوان مقدار درست استفاده می‌شود. با استفاده از متغیر آستانه، منحنی ROC به عنوان نرخ مثبت کاذب (False Positive Rate) در مقابل نرخ مثبت واقعی (True Positive Rate) ترسیم می‌شود، سطح زیر منحنی نشان می‌دهد که چه میزان نقشه برجستگی به خوبی خیرگی‌های واقعی چشم انسان را پیش‌بینی می‌کند. پیش‌بینی کامل و صحیح برابر با ۱ می‌شود.

ب) ضریب همبستگی خطی (LCC): این اندازه‌گیری به طور گسترده‌ای برای مقایسه رابطه بین دو تصویر برای کاربردهایی مانند: ثبت تصویر، تشخیص اشیاء و اندازه‌گیری تناقض (۲۷، ۲۸، ۳۰) استفاده می‌شود. ضریب همبستگی خطی که به صورت $LCC(G,S) = \frac{\sum_{x,y}(G(x,y)-\mu_G).(S(x,y)-\mu_S)}{\sqrt{\sigma_G^2.\sigma_S^2}}$ محاسبه می‌گردد، قدرت یک رابطه خطی بین دو متغیر را اندازه‌گیری می‌کند. G و S نشان‌دهنده GSM (نقشه خیرگی چشم انسان) و ESM (نقشه برجستگی مدل) می‌باشد. μ و σ میانگین و واریانس مقادیر در نقشه‌های برجستگی هستند. مزیت جالب LCC توانایی مقایسه دو متغیر با ارائه یک ارزش عددی بین ۱+ و ۱- است. وقتی همبستگی نزدیک به ۱+ / ۱- است، تقریبا یک رابطه کاملا خطی بین دو متغیر وجود دارد.

ج) نرمال‌سازی مسیر پیمایش برجستگی (NSS): نرمال‌سازی مسیر پیمایش برجستگی (۲۱، ۲۹) به عنوان مقدار پاسخ موقعیت چشم انسان تعریف شده است، $(x_k y_k)$ در مدل ESM که میانگین صفر و انحراف معیار واحد دارد نرمال شده است.

NSS $NSS = \frac{1}{\sigma_s}(S(x_h,y_h)-\mu_S)$ برای هر ساکاد چشم محاسبه و سپس میانگین و خطای استاندارد بر روی مجموع مقادیر NSS محاسبه می‌شود. $NSS=1$ نشان می‌دهد که موقعیت‌های چشم افراد در منطقه‌ای قرار دارد که تراکم پیش‌بینی شده، انحراف معیاری بالاتر از حد متوسط دارد) همچنین، $NSS\leq 0$ نشان می‌دهد که کارایی مدل بهتر از انتخاب یک موقعیت تصادفی بر روی نقشه نمی‌باشد. NSS به تغییر پارامترها (میانگین، موقعیت) مقاوم نیست.

د) معیار شباهت (SSIM): حاصل جمع مقادیر کمینه در هر نقطه از توزیع است. نحوه محاسبه شباهت برای دو نقشه برجستگی SM و FM به صورت $S = \sum_{x=1}^{X} min(SM(x), FM(x))$ می‌باشد. اگر مقدار شباهت برابر ۱ باشد بیانگر یکسان بودن توزیع دو نقشه برجستگی است و اگر مقدار صف بیانگر تفاوت کامل بین دو نقشه برجستگی است (۳۳).

یافته‌ها

در این بخش هر یک از مدل‌های محاسباتی توجه بینایی مبتنی بر برجستگی بینایی بر روی چندرسانه‌ای ساخته شده بر اساس پنج اصل اول از اصول ۱۲گانه ساخت چند رسانه‌ای، پیاده سازی و نتایج آن مورد ارزیابی قرار می‌گیرد. لازم به ذکر است مقادیر به دست آمده بر اساس معیارهای مذکور در بخش‌های ۲ و ۳ برای هر شات (مجموع فریم‌های متوالی برای خلق یک صحنه) می‌باشد.

نقشه برجستگی اصلی (GSM) توسط داده های چشمی ۳ شرکت کننده (دارای سلامت بینایی) که در حالت مشاهده آزاد، ویدئویی ۵ دقیقه‌ای (۳۰ فریم بر ثانیه) را مشاهده کرده اند، به دست آمده است. این ویدئو شامل ۴۸ شات و ۵۰۶۹ فریم می‌باشد. بدین ترتیب که هر جا چشم مخاطب بر نقطه‌ای خیره شود آن موقعیت به عنوان نقطه تثبیت خیرگی در GSM می باشد. سپس برای انطباق بیشتر با محدوده دید انسان، به ازای هر فریم جداگانه، یک فیلتر گاوسی بر روی آن اعمال و GSM نهایی تولید می‌شود.

مقایسه روش‌های مختلف

برخی از ویژگی‌های بررسی شده شامل رنگ، شدت روشنایی، تغییر

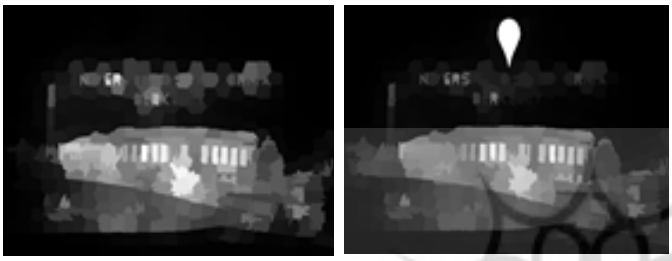
جهت و حرکت می‌باشد. نمونه‌ای از این ویژگی‌ها در برخی از فریم‌ها قابل مشاهده می‌باشد. همان‌طور که در شکل ۱-الف به وضوح دیده می‌شود، در حواشی تصویر هیچ توجه بینایی جلب نشده و تمامی توجه کاربر به بخش مرکزی می‌باشد، این مسئله به دو دلیل است؛ دلیل اول ناشی از نبود اطلاعات در اطراف تصویر (تفاوت مرکز محور) و دلیل دوم تمرکز انسانی بر روی مرکز صفحه (تمایل به مرکز) که در بخش ۲ بیان گردید.

فریم (۱): فریم اصلی

فریم (۱، ۲): فریم اصلی

فریم (۲): برجستگی بینایی

الف



فریم (۳، ۴): برجستگی بینایی

ب

شکل ۱- الف. تصویری از یک فریم ویدئو به همراه برجستگی بینایی به دست آمده از ویژگی‌های سطح پائین، ب) دو فریم ابتدایی و انتهایی از چند فریم متوالی و توجه بینایی مبتنی بر برجستگی‌های بینایی جهت نمایش افزایش توجه به موقعیتی خاص بر اساس اصل نشانه‌گذاری

در شکل ۱-ب اصل نشانه‌گذاری بررسی شده است که در آن هدف طراح جلب توجه بیننده به عبارت نام دانشگاه است. همان‌طور که در شکل ۱-ب فریم ۴ دیده می‌شود، توجه کاربر به سمت علامت مکان جلب می‌شود و دلیل آن حرکت و تباین بالای علامت اضافه شده با پس‌زمینه سفید رنگ اطراف آن است، این موضوع نشان می‌دهد که با نشانه‌گذاری صحیح می‌توان توجه مخاطب را به مکان مورد نظر بهتر جلب نمود و اثر یادگیری را در او افزایش داد.

فریم (۱): فریم اصلی

فریم (۲): برجستگی بینایی

الف

فریم (۱، ۲): فریم اصلی

فریم (۳، ۴): برجستگی بینایی

ب

شکل ۲- الف. نمونه‌ای از دو فریم متوالی و توجه بینایی مبتنی بر برجستگی‌های بینایی به دست آمده از ویژگی حرکت، ب) دو فریم ابتدایی و انتهایی از چند فریم متوالی و توجه بینایی مبتنی بر برجستگی‌های بینایی جهت نمایش اصل مجاورت زمانی و مکانی در توالی حروف

در شکل ۲-الف ویژگی حرکت (Motion) که عمدتا بر روی تشخیص اختلاف مکانی اشیاء بین فریم‌های متوالی می‌باشد، قابل مشاهده است که با استفاده از روش جریان نوری (Optical Flow) (۱۹، ۲۰) به دست می‌آید. همان‌طور که در شکل ۲-الف مشاهده می‌شود قسمت دهان شمایل انسانی که در تصویر است بین دو فریم دارای تغییر می‌باشد و این تغییر در تصویر توجه بینایی در شکل ۲-الف فریم ۴ قابل مشاهده می‌باشد. در دو فریم متوالی شکل ۲-الف مقادیر به دست آمده بر اساس معیارهای ذکر شده برای هر یک از مدل‌های مذکور در بخش ۲ مطابق جدول ۱ است، همان‌طور که مشاهده می‌شود مدل‌های مختلف دارای مقادیر مناسب و نزدیک به هم هستند که در مجموع معیارها، مدل Wang و همکاران (۲۰) دارای عملکرد بهتری می‌باشد.

در شکل ۲-ب فریم ۱ و ۲ دو فریم از توالی چند فریم متوالی مشاهده می‌شود که در آن حروف به ترتیب زمانی و مکانی نمایش داده می‌شود و متناسب با آن در شکل ۲-ب فریم ۳ و ۴ توجه بینایی را جلب می‌کند که این امر اثر حرکت شیء در تصویر و دنبال کردن مطلب مورد نظر توسط مخاطب را در توالی زمانی و مجاورت مکانی نمایش می‌دهد. در فریم‌های شکل ۲-ب مقادیر به دست آمده بر اساس معیارهای ذکر شده برای هر یک از مدل‌های مذکور مطابق جدول ۲ می‌باشد، همان‌طور که در این جدول مشاهده می‌شود مدل‌ها دارای مقادیر مناسب و نزدیک به هم می‌باشند که در مجموع معیارها، مدل GBVS (۲۳، ۳۳) دارای عملکرد بهتری می‌باشد.

جدول ۱. مقادیر معیارهای ارزیابی برای دو فریم متوالی در شکل ۲- الف

مدل	ROC	NSS	LCC	SSIM
(۲۳)	۰/۷۱۶۶	۲/۴۹۳۵	۰/۳۷۶۶	۰/۲۱۸۹
(۱۰)	۰/۸۱۶۴	۲/۲۹۱۱	۰/۴۰۲۶	۰/۲۲۵۲
(۶)	۰/۸۰۷۸	۲/۰۰۱۴	۰/۴۰۰۲	۰/۲۰۵۴
(۲۰)	۰/۶۹۷۳	۲/۴۰۶۵	۰/۶۱۲۴	۰/۳۷۰۷

جدول ۲. مقادیر معیارهای ارزیابی برای فریم‌های متوالی در شکل ۲- ب

مدل	ROC	NSS	LCC	SSIM
(۲۳)	۰/۶۲۱۰	۱/۷۳۰۵	۰/۵۱۷۹	۰/۳۵۶۹
(۱۰)	۰/۶۵۰۴	۱/۷۱۰۲	۰/۵۱۲۰	۰/۳۴۴۶
(۶)	۰/۵۲۵۹	۱/۴۱۳۶	۰/۴۰۶۹	۰/۳۱۶۷
(۲۰)	۰/۵۲۸۰	۱/۲۰۳۸	۰/۳۸۱۲	۰/۲۶۹۳

فریم (۵، ۶، ۷، ۸): برجستگی بینایی

شکل ۳. چند فریم متوالی و توجه بینایی مبتنی بر برجستگی‌های بینایی جهت ارزیابی جلب توجه در تغییرات ناگهانی در فریم‌های متوالی

در شکل ۳ اصل مجاورت فضایی قابل مشاهده است و بیان می‌کند که حرکت نداشتن اشیاء می‌تواند باعث حذف توجه انسان از یک شیء شود. در این چند فریم متوالی، همان‌طور که مشاهده می‌شود طبق (اصل مجاورت مکانی) در ابتدا که پرچم آمریکا نمایش داده می‌شود، توجه بینایی به دلیل وجود مؤلفه حرکت جلب می‌شود اما پس از گذشت چند فریم دیگر توجه به آن محل کاهش می‌یابد.

برای فریم‌های شکل ۳ مقادیر به دست آمده بر اساس معیارهای ارزیابی برای هر یک از مدل‌های مذکور در بخش ۲ مطابق جدول ۳ است.

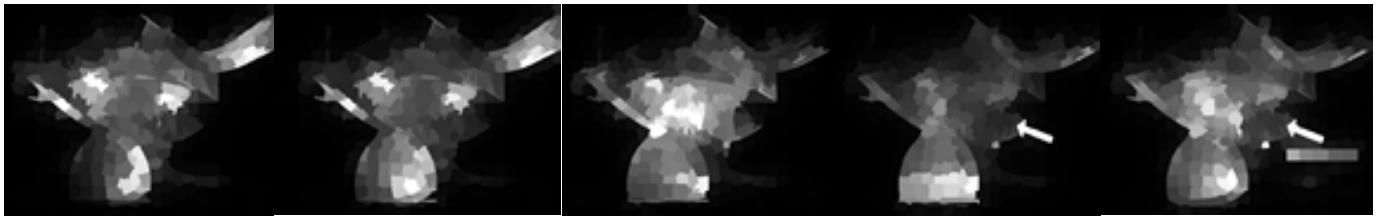
همان‌طور که در جدول ۳ مشاهده می‌شود مدل‌ها دارای مقادیر مناسب و نزدیک به هم می‌باشند که در مجموع معیارها، مدل GBVS دارای عملکرد بهتری می‌باشد (۲۳).

جدول ۳. مقادیر معیارهای ارزیابی برای فریم‌های متوالی در شکل ۳

مدل	ROC	NSS	LCC	SSIM
(۲۳)	۰/۵۹۳۳	۰/۹۶۰۷	۰/۲۳۷۲	۰/۱۴۷۰
(۱۰)	۰/۵۶۱۰	۰/۵۶۹۶	۰/۱۵۱۰	۰/۱۲۱۵
(۶)	۰/۴۹۳۷	۰/۰۶۳۳	۰/۰۱۶۸	۰/۰۹۰۴
(۲۰)	۰/۵۲۷۷	۰/۳۰۵۵	۰/۰۷۶۰	۰/۱۰۸۰

در شکل ۴ هدف طراح چندرسانه‌ای بر اساس اصل نشانه‌گذاری، انسجام و افزونگی جلب توجه بیننده به سمت متن قسمت‌هایی از تصویر می‌باشد که این کار را در دو مرحله انجام می‌دهد، مرحله اول در شکل ۴ فریم ۲ و ۳ قابل مشاهده می‌باشد اما به دلیل اختلاف کم با پس‌زمینه و نازک بودن متن و نشانه، جلب توجه به خوبی صورت نمی‌گیرد که در شکل ۴ فریم ۷ و ۸ قابل مشاهده می‌باشد. در مرحله بعد در شکل ۴ فریم ۴ و ۵ به علت اختلاف نشانه با اطراف، نزدیکی با مرکز و حرکت مناسب، توجه را بهتر جلب می‌کند که در شکل ۴ فریم ۹ و ۱۰ قابل مشاهده می‌باشد. برای فریم‌های شکل ۴ مقادیر به دست آمده بر اساس معیارهای ارزیابی در جدول ۴، مشاهده می‌شود که مدل‌ها دارای مقادیر مناسب و نزدیک به هم هستند و مدل Wang و همکاران دارای عملکرد بهتری است (۲۰).

فریم (۱، ۲، ۳، ۴، ۵): فریم اصلی

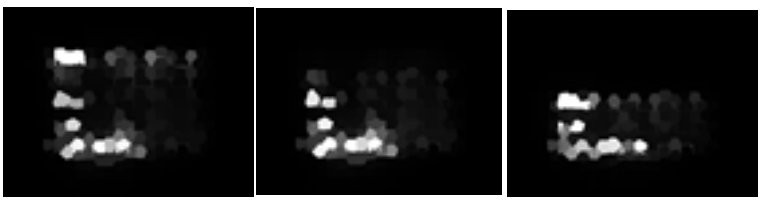


فریم (۶، ۷، ۸، ۹، ۱۰): برجستگی بینایی

شکل ۴. چند فریم متوالی و توجه بینایی مبتنی بر برجستگی‌های بینایی بیانگر توجه بینایی بسیار ضعیف به نشانه اول در فریم‌های ابتدایی و در مقابل توجه بیشتر به نشانه‌گذاری دوم به علت اندازه و موقعیت نزدیک به مرکز نشانه

جدول ۴. مقادیر معیارهای ارزیابی برای فریم متوالی در شکل ۴

مدل	ROC	NSS	LCC	SSIM
(۲۳)	۰/۶۳۶۷	۰/۷۷۴۹	۰/۱۸۴۴	۰/۱۴۲۴
(۱۰)	۰/۶۳۷۶	۱/۰۳۳۷	۰/۲۶۱۷	۰/۱۵۷۸
(۶)	۰/۵۵۶۳	۰/۶۰۱۴	۰/۱۳۳۶	۰/۱۲۷۰
(۲۰)	۰/۵۹۵۱	۰/۸۸۹۴	۰/۲۲۵۰	۰/۱۸۵۸



فریم (۱، ۲، ۳): فریم اصلی

فریم (۴، ۵، ۶): برجستگی بینایی

شکل ۵. نمونه چند فریم متوالی و تصویر توجه بینایی مبتنی بر برجستگی‌های بینایی بیان‌گر توجه بیشتر به سوی کلمات به علت داشتن اندازه مناسب و اختلاف رنگ با پس‌زمینه در فریم‌های متوالی

در شکل ۵ با رعایت اصل مجاورت فضایی و زمانی و به علت اندازه مناسب متن‌ها، اختلاف رنگ با پس‌زمینه و نیز دارای حرکت مناسب توجه به طور مناسبی به محل مورد نظر جلب شده است که در شکل ۵ فریم ۴، ۵ و ۶ قابل مشاهده می‌باشد. برای فریم‌های شکل ۵ مقادیر به دست آمده بر اساس معیارهای ارزیابی در جدول ۵، مشاهده می‌شود که مدل‌ها دارای مقادیر مناسب و نزدیک به هم می‌باشند که مدل Itti و همکاران دارای عملکرد بهتری است (۱۰).

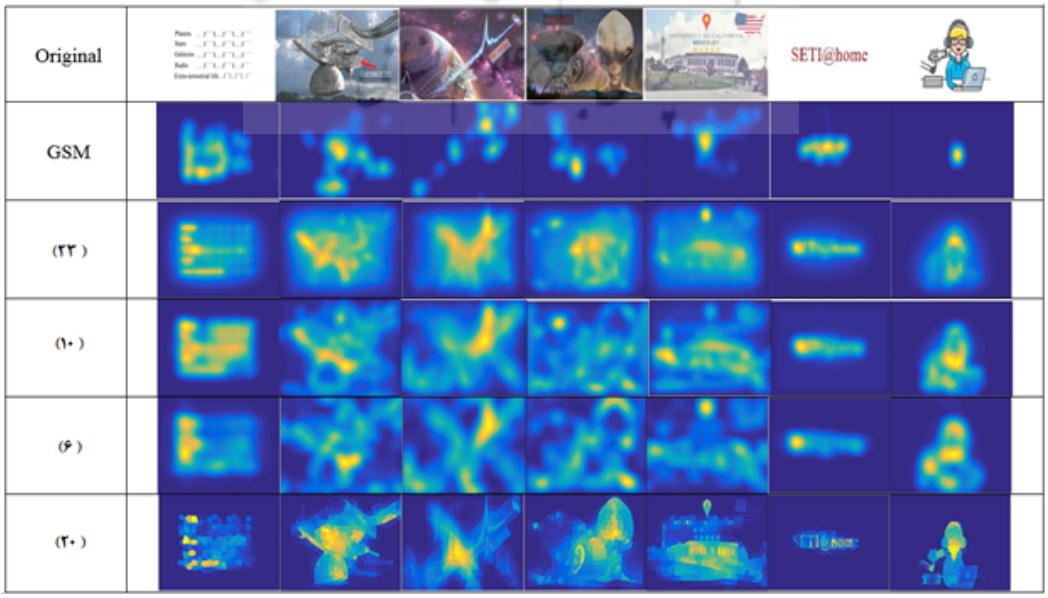
جدول ۵. مقادیر معیارهای ارزیابی برای فریم متوالی در شکل ۵

مدل	ROC	NSS	LCC	SSIM
(۲۳)	۰/۶۱۰۳	۱/۴۵۰۶	۰/۴۱۳۵	۰/۲۷۳۴
(۱۰)	۰/۶۵۷۷	۱/۳۹۹۲	۰/۳۹۳۰	۰/۲۸۱۵
(۶)	۰/۵۶۹۰	۱/۲۲۰۰	۰/۳۲۱۰	۰/۲۴۳۳
(۲۰)	۰/۵۶۶۵	۱/۱۱۰۲	۰/۳۵۴۴	۰/۳۰۱۲

همان‌طور که در جدول ۱ تا جدول ۵ مشاهده می‌شود روش‌های GBVS (۲۳)، Itti و همکاران (۱۰)، Wang و همکاران (۲۰) و Hou و Zhang (۶) به ترتیب دارای عملکرد بهتر در شرایط متناسب به خصوصیات ذکر شده در فریم‌ها (مانند حرکت، اختلاف رنگ مرکز و حاشیه اشیاء و..) هستند (جدول ۶).

جدول ۶. رتبه‌بندی مدل‌ها

رتبه	مدل
۱	(۲۳)
۲	(۱۰)
۳	(۶)
۴	(۲۰)



شکل ۶. نقشه‌های برجستگی بینایی به دست آمده توسط ۴ مدل توجه بینایی برای مجموعه فریم‌های ارزیابی شده

در شکل ۶ مجموع فریم‌های ارزیابی شده به همراه نقشه‌های برجستگی هر یک از مدل‌های مذکور برای هر شات (مجموع توجه بینایی برای هر شات به صورت مجزا) به نمایش درآمده است. همان‌طور که مشاهده می‌شود بیشترین انطباق با داده‌های چشم انسان در فریم‌های دارای تفاوت مرکز محور مربوط به مدل Itti و همکاران (۱۰) و در فریم‌های داری تمایل به مرکز مربوط به مدل GBVS و در فریم‌های دارای حرکت و نیز سوپر پیکسل‌های یکسان مربوط به مدل Wang و همکاران است (۲۰).

بحث

در واقع، این مطالعه از مدل‌های موجود در توجه بینایی پایین به بالا برای بهبود محتوای آموزشی استفاده می‌کند. همان‌طور که در نتایج به دست آمده مشاهده گردید مدل‌های محاسباتی توجه بینایی می‌توانند به عنوان معیاری جهت پیش‌بینی مکان توجه مخاطب مورد استفاده قرار گیرد. هر چند مدل‌ها دارای نتایج یکسانی نیستند اما دارای عملکردی مناسب و نزدیک به هم در پیش‌بینی نواحی مورد توجه انسان‌ها توجه دارند که با استفاده از امکانی که این مدل‌ها در اختیار ما قرار می‌دهند می‌توان کیفیت ساخت چندرسانه‌ای‌های آموزشی را افزایش داد.

نتیجه‌گیری

با اجرای مدل‌های محاسباتی توجه بینایی بروی چند رسانه‌ای آموزشی ایجاد شده بر اساس اصول مذکور و بررسی نتایج آن بر اساس معیارهای مورد نظر شاهد عملکرد خوب و انطباق مناسب این مدل‌ها با داده‌های چشم انسان بودیم، می‌توان با استفاده از مدل‌های توجه بینایی و یا ترکیبی از آنها مکان‌های مورد توجه مخاطب پیش‌بینی شده و با قرار دادن مطالب مهم مورد نظر در مکان پیش‌بینی شده توسط مدل توجه بینایی، کیفیت و تاثیرگذاری چندرسانه‌ای آموزشی را افزایش داد، این امر سبب آن می‌گردد که علاوه بر کاهش میزان بار شناختی ذهن مخاطب، موارد با اهمیت در محلی مناسب از تصویر قرار گیرند که بیشتر مورد توجه بینایی است و مخاطب در زمان بسیار کم‌تر و کیفیت بیشتر مطالب آموزشی را درک و فهم نماید.

همان‌طور که در آزمایش‌های صورت گرفته مشاهده می‌شود یک مدل به تنهایی در همه موارد از کارآمدی بالا برخوردار نیست و مدل‌ها باتوجه به نوع الگوریتم و ویژگی‌های مورد استفاده دارای نتایج متفاوتی می‌باشند که با ترکیب بهره‌برداری از آنها می‌توان به بهبود عملکرد نهایی کمک شایانی نمود.

در ادامه کار تلاش می‌کنیم با بهبود مدل مذکور و یا ایجاد مدل محاسباتی نوین مقادیر معیارهای مذکور را افزایش داده تا با داده‌های واقعی انطباق بیشتری داشته باشد. همچنین بررسی سیگنال‌های مغزی و اثر روش‌های جلب توجه در کاهش بارشناختی و نیز تهیه نرم‌افزاری که بتواند جهت ارزیابی چندرسانه‌ای آموزشی تولید شده مورد استفاده قرار گیرد، می‌تواند برای کارهای آینده مورد توجه قرار گیرد.

ملاحظات اخلاقی

پیروی از اصول اخلاق در پژوهش

این مقاله برگرفته از پایاننامه کارشناسی ارشد نویسنده اول می‌باشد. پژوهش حاضر با رعایت اصول اخلاقی مانند محرمانه بودن مشخصات شرکتکنندگان انجام شد و آنان در ترک مطالعه آزاد بودند. همچنین اطلاعات کافی در مورد انجام مطالعه ارائه گردیده و نتایج پژوهش‌صادقانه، دقیق و کامل منتشر شده است.

مشارکت نویسندگان

مجید شعبانی: وظیفه مطالعه، ایجاد، پیاده‌سازی و تحلیل و بررسی نتایج را بر عهده داشت. علیرضا بساق‌زاده: نویسنده مسئول و راهنمای مراحل اجرایی پژوهش و اصلاح مقاله بود. رضا ابراهیم‌پور: راهنمایی در زمینه روش کار و تجزیه تحلیل داده‌ها را بر عهده داشت. سید حمید امیری: راهنمای مراحل اجرایی پژوهش و اصلاح مقاله بودند. کیهان لطیف‌زاده: وظیفه تولید و جمع‌آوری نمونه را بر عهده داشتند.

منابع مالی

این مقاله مورد حمایت ستاد توسعه علوم و فناوری‌های شناختی با کد پژوهشی ۶۸۸۰ مصوبه ۱۳۹۷/۱۰/۰۸ و طرح پژوهشی دانشگاه تربیت دبیر شهید رجایی با شماره قرارداد ۳۹۱۱۰ بوده است.

تشکر و قدردانی

با تشکر از کلیه افراد شرکت‌کننده در این مطالعه، که مشارکت منظم داشتند و اساتید محترمی که در این کار به راهنمایی و مشاوره پرداختند.

تعارض منافع

نویسندگان مقاله حاضر هیچ‌گونه تعارض منافعی را گزارش نکرده‌اند.

References

1. Shafiei H, Zare H. Effectiveness of attention bias modification by computerized attention training on reducing social anxiety of adolescents. *Advances in Cognitive Sciences*. 2019;21(2):108-120. (Persian)

2. Shamsfard M, Abd Elahzade Barfroush A. Extracting conceptual knowledge from text: Using linguistic and semantic templates. *Advances in Cognitive Sciences*. 2002;4(1):48-66. (Persian)

3. Mayer RE. Multimedia learning. In: Brain HS, editor. Psychology of learning and motivation. Vol. 41. New York:Academic Press;2002. pp. 85-139.

4. Koch K, McLean J, Segev R, Freed MA, Berry II MJ, Balasubramanian V, et al. How much the eye tells the brain. *Current Biology*. 2006;16(14):1428-1434.

5. Itti L. Models of bottom-up and top-down visual attention. Pasadena, California:California Institute of Technology;2000.

6. Hou X, Zhang L. Saliency detection: A spectral residual approach. In 2007 IEEE Conference on Computer Vision and Pattern Recognition; 2007 Jun 17-22; Minneapolis, MN, USA IEEE; 2007. pp. 1-8.

7. Borji A, Itti L. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012;35(1):185-207.

8. Treue S. Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*. 2001;24(5):295-300.

9. Rolls ET, Deco G. Attention in natural scenes: Neurophysiological and computational bases. *Neural Networks*. 2006;19(9):1383-1394.

10. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 1998;20(11):1254-1259.

11. Tsotsos JK, Culhane SM, Wai WY, Lai Y, Davis N, Nuflo F. Modeling visual attention via selective tuning. *Artificial Intelligence*. 1995;78(1-2):507-545.

12. Itti L, Koch C. Computational modelling of visual attention. *Nature Reviews Neuroscience*. 2001;2(3):194-203.

13. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*. 1995;18(1):193-222.

14. Rae R. Gesture-based human-machine communication based on visual attention and adaptivity. [PhD Dissertation]. Bielefeld:Bielefeld University;2000.

15. Cerf M, Harel J, Einhäuser W, Koch C. Predicting human gaze using low-level saliency combined with face detection. In: Platt JC, Koller D, Singer Y, Roweis S, editors. Advances in neural information processing systems. Vol 20. Cambridge:MIT Press;2008. pp. 241-248.

16. Itti L, Braun J, Lee D, Koch C. Attentional modulation of human pattern discrimination psychophysics reproduced by a quantitative model. Advances in neural information processing systems. Vol 2. Cambridge:MIT Press;1998. pp. 789-795.

17. Frintrop S. VOCUS: A visual attention system for object detection and goal-directed search. Berlin, Heidelberg:Springer;2006.

18. Tatler BW. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*. 2007;7(14):4.

19. Wang W, Shen J, Porikli F. Saliency-aware geodesic video object segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015 June 7-12; Boston, MA, USA; IEEE;2015. pp. 3395-3402.

20. Wang W, Shen J, Yang R, Porikli F. Saliency-aware video object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;40(1):20-33.

21. Parkhurst D, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*. 2002;42(1):107-123.

22. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW. SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*. 2008;8(7):32.

23. Harel J, Koch C, Perona P. Graph-based visual saliency. Ad-

vances in neural information processing systems 19. Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems; 2006 December 4-7; Vancouver, British Columbia, Canada; 2006. pp. 545-552

24. Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*. 1998;124(3):372-422.

25. Green DM, Swets JA. Signal detection theory and psychophysics. New York:Wiley;1966.

26. Bruce N, Tsotsos J. Saliency based on information maximization. In: Weiss Y, Scholkopf B, Platt J, editors. Advances in neural information processing systems. Vol 18. Cambridge:MIT Press;2005. pp. 155-162

27. Jost T, Ouerhani N, Von Wartburg R, Muri R, Hugli H. Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*. 2005;100(1-2):107-123.

28. Rajashekar U, Bovik AC, Cormack LK. Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*. 2006;6(4):7.

29. Peters RJ, Iyer A, Itti L, Koch C. Components of bottom-up gaze allocation in natural images. *Vision Research*. 2005;45(18):2397-2416.

30. Privitera CM, Stark LW. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2000(9):970-982.

31. Wen H, Zhou X, Sun Y, Zhang J, Yan C. Deep fusion based video saliency detection. *Journal of Visual Communication and Image Representation*. 2019;62:279-285.

32. Tavakoli HR, Borji A, Rahtu E, Kannala J. DAVE: A deep audio-visual embedding for dynamic saliency prediction. *arXiv preprint arXiv*:1905.10693. 2019 May 25.

33. Bosaghzadeh A, Shabani M, Ebrahimpour R. A computational-cognitive model of visual attention in dynamic environments. *Journal of Electrical and Computer Engineering Innovations (JECEI)*. 2021;10(1):163-174.

34. Riche N, Duvinage M, Mancas M, Gosselin B, Dutoit T. Saliency and human fixations: State-of-the-art and study of comparison metrics. Proceedings of the IEEE international conference on computer vision; 2013 December 1-8; Sydney, NSW, Australia; IEEE; 2013. pp. 1153-1160.

35. Sarailoo R, Latifzadeh K, Amiri SH, Bosaghzadeh A, Ebrahimpour R. Assessment of instantaneous cognitive load imposed by educational multimedia using electroencephalography signals. *Frontiers in Neuroscience*. 2022;16:744737.