Visual Analytics Research at Universidade Federal do Ceará

Emanuele Santos*‡, José F. de Queiroz Neto‡, George A. M. Gomes†, Antonio José M. Leite Junior †,
Joaquim Bento Cavalcante Neto*, Creto Augusto Vidal*,
Ticiana L. Coelho da Silva‡ and Jose Antonio F. Macedo‡
*CRAB - Computer Graphics, Virtual Reality, Animation, and Visualization Group
Universidade Federal do Ceará, Fortaleza - CE - Brazil
Email: {emanuele, cvidal, joaquimb}@dc.ufc.br
https://crab.ufc.br/en/
† Instituto UFC Virtual
Universidade Federal do Ceará, Fortaleza - CE - Brazil
Email: {melojr, george}@virtual.ufc.br
‡Insight Data Science Lab
Universidade Federal do Ceará, Fortaleza - CE - Brazil
Email: {emanuele, florencio, ticianalc, jose.macedo}@insightlab.ufc.br
https://insightlab.ufc.br/

I. Introduction

In this paper, we describe the research in visual analytics that is a partnership between two research groups at the Federal University of Ceara, *CRAB – Computer Graphics, Virtual Reality, Animation, and Visualization Group* and *Insight Data Science Lab*. This fruitful collaboration joined CRAB's expertise in Visualization with Insight Lab's expertise in Data Science to bring about valuable research and development projects in the public and private sector in many areas, including crime analysis, mobility data visualization, education, and games. We briefly describe our main projects in these areas and their social, technological, and scientific contributions in the following sections.

II. Crime Analysis

Crime has become a central problem in many countries in the world. In Brazil, crime is a theme of growing interest. Violent crimes, in particular, have dramatically increased in some parts of Brazil in recent decades [1], challenging governments and the whole society [2]. We have been working in crime analysis since 2016, when we started a collaboration with Secretaria de Segurança Pública e Defesa Social do Estado do Ceará – SSPDS-CE. We first developed MSKDE (Marching Squares Kernel Density Estimation) [3], a technique to facilitate the generation of crime hotspot maps. The collaboration grew into a much larger project named SPI - Scientific and Technology Intelligence in Public Safety. SPI's goal was to develop scientific studies in data science, machine learning, and visualization to apply technological solutions that included individuals and vehicles identification, anomalous patterns discovery, and visual analytics. Among the products developed under SPI, we can highlight CrimeWatcher [4], a visual analytics solution that facilitates tracking crime over time and identifying its geographic patterns (see Figure 1). To help understanding how high-rate crime regions change over time, we developed a partition approach to interpolate 2D polygon sets that can be used to visualize temporal changes of the crime hotspots as an animation [5]. We have also approached the problem of approximating KDE hotspots to the road network [6], so the police patrols can be performed more efficiently [7].

SSPDS-CE is currently using CrimeWatcher (with the new name STATUS) to explore crime data, generate hotspot maps and plan patrols. Almost a year after the SPI project started, its ideas and solutions evolved into a new project with the Brazilian Ministry of Justice and Public Safety called Sinesp Big Data and Artificial Intelligence for Public Safety[1]. Under Sinesp Big Data, CrimeWatcher was refactored to become Sinesp Geointeligência and now is being used in many Brazilian States.

III. Mobility Data Visualization

Our group has also been working on visual analytics for mobility data. Inspired by animated vector field visualization techniques [9], we developed a technique [10] to visualize trajectory data. The result is a visualization such that the movement of an object leaves a trail that persists on the road, similarly to a path line in vector field visualization. However, in our visualization, this trail is slowly fading out, and its length represents the magnitude of the object's speed. To be used in real-time applications, we do not compute vector fields neither map matching, i.e., the trajectory data are processed to

[1] https://www.justica.gov.br/news/collective-nitf-content-1566331890.72

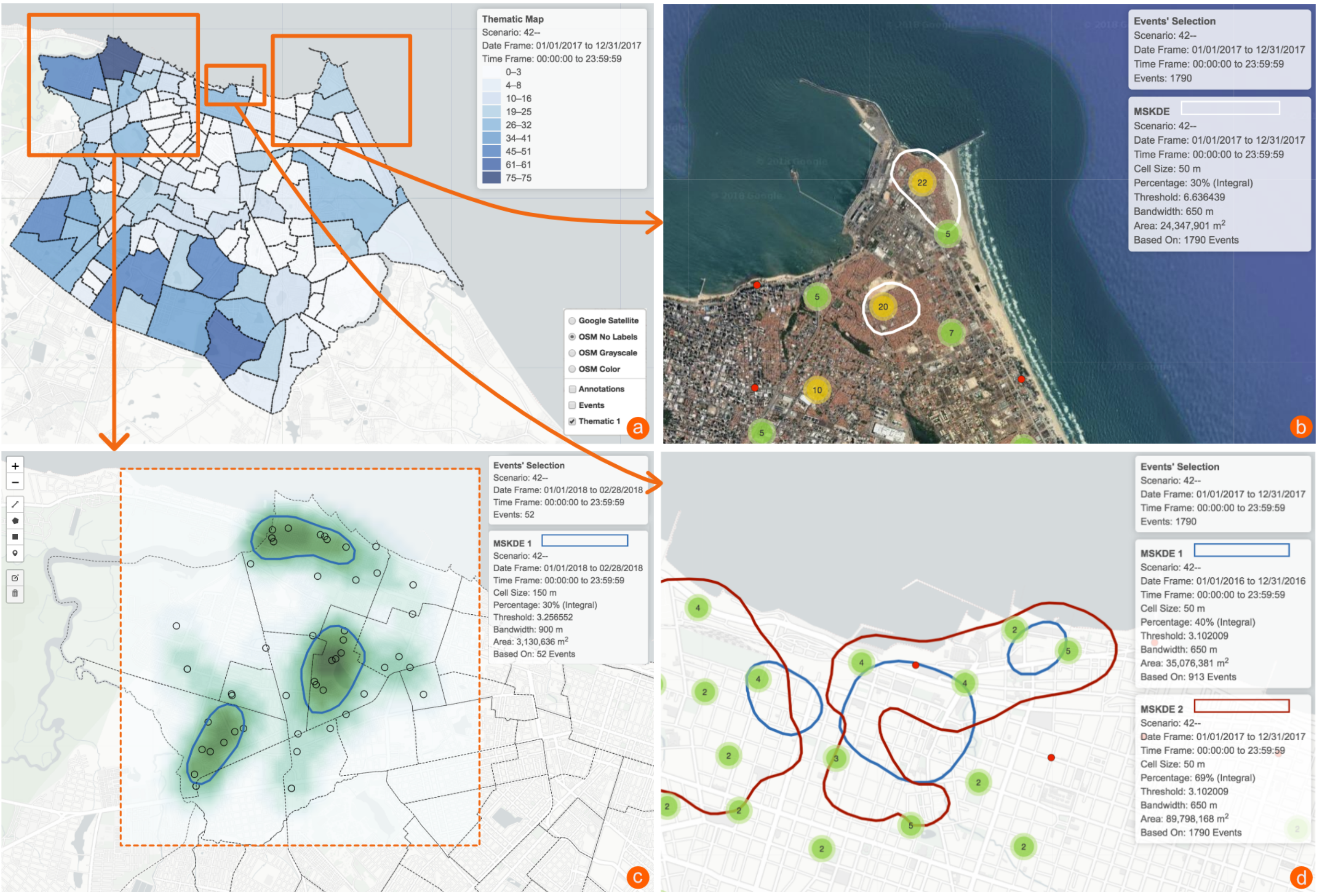

Fig. 1. CrimeWatcher's different visualizations. In (a), a combination of a geographic map with a choropleth map. In (b), a satellite geographic map is used for better understanding the spatial context of the region. MSKDE contours are displayed in white, improving the contrast with the background. Crimes are represented as red circles or circular clusters. In (c), a combination of a Weighted-KDE map below an MSKDE map with a blue border and no fill. In (d), a more detailed image with a gray-scaled background. In this case, we show two MSKDE maps regarding two different periods of time (2016 and 2017). Source: [8]

visualize the objects' movement directly. We also applied this technique to discover and keep track of hot routes (routes with large volumes of traffic in the recent past) from a stream of raw trajectory data in real-time [11]. Figure 2 illustrates how hot routes are discovered in our approach.

It is also hard to investigate in a large trajectory dataset where and when the major occurrences are and which trajectories need to be observed in more detail [12]. We also explored other forms of visualization of trajectory data [13] based on a special data structure named Probabilistic Suffix Tree (PST) [14]. In our proposed analytics solution, users can solve complex problems such as investigating where and when a person probably lives, works or studies.

IV. Education

One of the most difficult challenges that educators and leaders in higher education face today is reducing the high student dropout rates in their institutions. In Brazil, the student dropout causes a loss of public resources in federal universities and produces a lack of skilled workers, which impacts negatively on the country's development [15]. To tackle this problem, we proposed a student dropout prediction strategy [16] based on the classification with reject option paradigm. In such a strategy, our method classifies students into dropout prone or non-dropout prone classes and may also reject classifying students when the algorithm does not provide a reliable prediction. The rejected students are the ones that could be classified into either class and so are probably the ones with more chances of success when subjected to personalized intervention activities. In the proposed method, the reject zone can be adjusted so that the number of rejected students can meet the available workforce of the educational institution. Our method was tested on a dataset collected from 892 undergraduate students from 2005 to 2016. One factor that also plays a significant role in the development of students' knowledge and their performance is the structure of the curriculum [17]. We also proposed a data mining technique [18] that evaluates a curriculum's structure based on academic data collected from the undergraduate students. Our approach is based on the Synthetic Control Method (SCM), which has

Fig. 2. Discovering hot routes with our approach, using a real-world raw dataset. (a) A traffic jam propagating along the road in the left dotted rectangle. (b) The evolution of the hot routes in the right dotted rectangle in four snapshots taken every half hour. A video showing animated visualizations is available on https://www.youtube.com/watch?v=mDUy956-vSA.

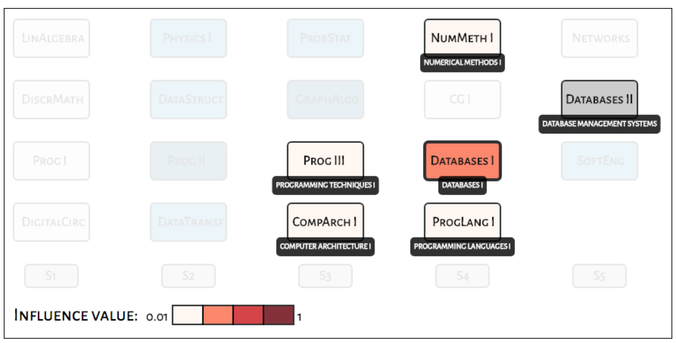

Fig. 3. This view shows the influence value view for *Databases II* captured by the model, in which the saturation increases according to the influence rate. In the model, *Databases II* is more influenced by *Databases I*, being consistent with the official curriculum. Source: [18]

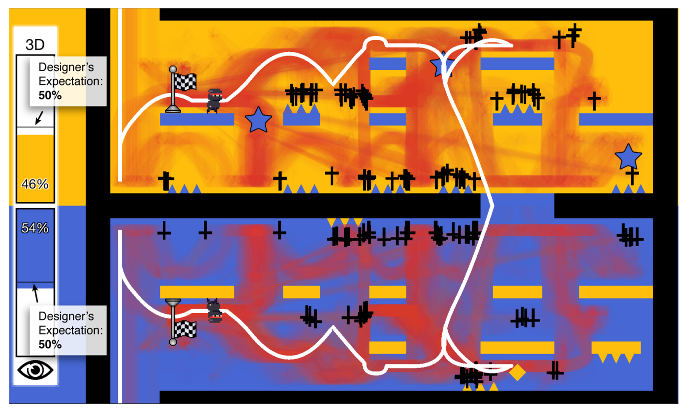

Fig. 4. Visualization of a challenge. Crosses around trigger-controlled spikes suggest that these deaths happened because players did not see them. The white path represents the optimal solution. Red paths represent actual players' movements (154 paths in this segment). Source: [22]

been successfully applied in previous studies in social science and economics [19], [20]. The SCM builds a linear model describing the relation between courses based on a dataset of student performance information. In the resulting linear model, each course is described as a convex combination of previous courses, thus improving the interpretability of the model. In addition to providing the relation between courses, the proposed method can also be used to predict students' grades in a specific course based on their previous grades. The results can be visualized inside a user-friendly tool (see Figure 3), which allows for contrast and comparison between the official structure and the structure found based on the data. On another education front, we also tried visual analysis to teach abstract concepts such as the PST data structure used in mobility data visualization. We performed a practical experiment [21] that tries to make the teaching of PSTs more applied, based on the development of a specific application. The application was tested and analyzed quantitatively and qualitatively, demonstrating that the proposed solution can be a viable educational alternative.

V. Games

Modern game developers apply Games User Research (GUR) methods to make informed game design decisions based on their target audience. We combined traditional methods, including observation, interview, and questionnaires, with two affordable complementary methods, namely webcam-based eye-tracking and telemetry, along with data visualization in a playtesting routine [22]. By developing three versions of a hardcore 2D platform game that demands multitasking abilities using different GUR methods, we were able to find that the chosen complementary methods cover a significant amount of gameplay issues. The metrics and eye-tracking data visualization provided insights about multitasking and level design (see Figure 4). Furthermore, we discuss the challenges of evaluating prototypes regarding a more enjoyable experience when frustration is a desirable gameplay element.

VI. Conclusion

In this paper, we described the research in visual analytics that has been a partnership between two research groups at the Federal University of Ceara, *CRAB – Computer Graphics, Virtual Reality, Animation, and Visualization Group* and *Insight Data Science Lab*. This fruitful collaboration joined CRAB's expertise in Visualization with Insight Lab's expertise in Data Science to bring about valuable research and development projects in the public and private sector in many areas, including crime analysis, mobility data visualization, education, and games. We briefly described our main projects in these areas and their social, technological, and scientific contributions. As future work, we expect to continue our partnership with SSPDS-CE to further develop our research in Crime Analysis. We are also planning on extending our collaboration with other research areas. We are particularly interested in User-Centered Machine Learning and Data Science.

Acknowledgment

The authors would like to thank all the current and former students and collaborators who helped them create all the research work produced in the past few years. Their work has been funded by CNPq, CAPES, FUNCAP, and the Brazilian Ministry of Justice and Public Safety.

References

[1] R. S. Lima, S. Bueno, B. Rodrigues, A. C. Pekny, L. Figueiredo, P. N. Pröglhöf, I. Sobral, V. Moneo, C. A. P. Aparício, T. Kahn, C. Ricardo, D. Cerqueira, F. L. Oliveira, F. S. Silva, L. Pires, L. O. Ramos, L. G. Cunha, M. F. Baird, M. F. T. Peres, N. Pollachi, R. Alcadipani, R. Custódio, R. Miki, R. Muggah, and U. Peres, "Anuário brasileiro de segurança pública 2014," São Paulo, SP, Brazil, Tech. Rep., 2014. [Online]. Available: http://www.forumseguranca.org.br/storage/8_anuario_2014_20150309.pdf

[2] C. Izique, "Crescimento da violência no país surpreende pesquisadores," São Paulo, SP, Brazil, 2013. [Online]. Available: https://exame.com/brasil/violencia-democracia-e-direitos-humanos/

[3] J. F. d. Q. Neto, E. Santos, and C. A. Vidal, "Mskde - using marching squares to quickly make high quality crime hotspot maps," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2016, pp. 305–312.

[4] J. F. de Queiroz Neto, E. Santos, C. A. Vidal, and D. S. Ebert, "A visual analytics approach to facilitate crime hotspot analysis," *Computer Graphics Forum*, vol. 39, no. 3, pp. 139–151, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13969

[5] A. R. C. Ramos, E. Santos, and J. B. Cavalcante-Neto, "A partition approach to interpolate polygon sets for animation," in *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2019, pp. 139–146.

[6] F. C. F. N. Junior, T. L. C. d. Silva, J. F. d. Q. Neto, J. A. F. d. Macêdo, and W. C. Porcino, "A novel approach to approximate crime hotspots to the road network," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility*, ser. PredictGIS'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 53–61. [Online]. Available: https://doi.org/10.1145/3356995.3364538

[7] S. P. Chainey, J. A. S. Matias, F. C. F. Nunes Junior, T. L. Coelho da Silva, J. A. F. de MacÃªdo, R. P. MagalhÃ£es, J. F. de Queiroz Neto, and W. C. P. Silva, "Improving the creation of hot spot policing patrol routes: Comparing cognitive heuristic performance to an automated spatial computation approach," *ISPRS International Journal of Geo-Information*, vol. 10, no. 8, 2021. [Online]. Available: https://www.mdpi.com/2220-9964/10/8/560

[8] J. F. de Queiroz Neto, "A visual analytics approach for geocoded crime data," Ph.D. dissertation, Federal University of Ceara, Fortaleza, CE, Brazil, 4 2020.

[9] HINT.FM. (2012) Wind map. [Online]. Available: hint.fm/wind

[10] G. A. M. Gomes, E. Santos, and C. A. Vidal, "Interactive visualization of traffic dynamics based on trajectory data," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2017, pp. 111–118.

[11] G. A. Gomes, E. Santos, C. A. Vidal, T. L. C. da Silva, and J. A. F. Macedo, "Real-time discovery of hot routes on trajectory data streams using interactive visualization based on gpu," *Computers & Graphics*, vol. 76, pp. 129 – 141, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097849318301420

[12] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski, "Space, time and visual analytics," *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1577–1600, 2010.

[13] A. J. M. Leite, E. Santos, C. A. Vidal, and J. A. F. D. Macêdo, "Visual analysis of predictive suffix trees for discovering movement patterns and behaviors," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2017, pp. 103–110.

[14] C. L. Rocha, I. R. Brilhante, F. Lettich, J. A. F. De Macedo, A. Raffaetà, R. Andrade, and S. Orlando, "TPRED: a Spatio-Temporal Location Predictor Framework," in *Proceedings of the 20th International Database Engineering & Applications Symposium*. ACM, 2016, pp. 34–42.

[15] R. L. L. e. Silva Filho, P. R. Motejunas, O. Hipólito, and M. B. D. C. M. Lobo, "A evasão no ensino superior brasileiro," *Cadernos de Pesquisa*, vol. 37, no. 132, pp. 641–659, 2007. [Online]. Available: http://www.scielo.br/pdf/cp/v37n132/a0737132http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-15742007000300007&lng=pt&nrm=iso&tlng=pt

[16] A. M. Barbosa, E. Santos, and J. P. P. Gomes, "A machine learning approach to identify and prioritize college students at risk of dropping out," in *XXVIII Simpósio Brasileiro de Informática na Educação SBIE (Brazilian Symposium on Computers in Education)*, Recife, Nov 2017, pp. 1497–1506.

[17] C. Anuradha and T. Velmurugan, "A data mining based survey on student performance evaluation system," *5th IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2014*, no. December 2014, pp. 43–47, 2015. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84944399574&partnerID=40&md5=473cdf444b694afecbf5d0fc36e917fd

[18] A. M. Barbosa, A. N. de Araujo Neto, E. Santos, and J. P. P. Gomes, "Using learning analytics and visualization techniques to evaluate the structure of higher education curricula," in *XXVIII Simpósio Brasileiro de Informática na Educação SBIE (Brazilian Symposium on Computers in Education)*, Recife, Nov 2017, pp. 1297–1306.

[19] P. Hinrichs, "The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities," *The Review of Economics and Statistics*, vol. 94, no. 3, pp. 712–722, 2012. [Online]. Available: http://dx.doi.org/10.1162/REST_a_00170

[20] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program," *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 493–505, 2010. [Online]. Available: http://dx.doi.org/10.1198/jasa.2009.ap08746

[21] A. J. M. L. Junior, E. Santos, C. A. Vidal, and C. L. Rocha, "Empregando Análise Visual e Sensemaking no Ensino de Predictive Suffix Trees," in *XXIX Simpósio Brasileiro de Informática na Educação SBIE (Brazilian Symposium on Computers in Education)*, Fortaleza, Nov 2018, pp. 1043–1052.

[22] A. S. Bastos, E. Santos, G. A. M. Gomes, and M. A. Mourão, "Evaluating the Use of Affordable User Testing and Visualization Techniques in Level Design of a Hardcore 2D Platform Game," in *XVIII SBGames - Art & Design Track*, Rio de Janeiro, Oct 2019, pp. 145–154.