Applying the Gaussian Mixture Model to Generate Large Synthetic Data from a Small Data Set

Chanachok Chokwitthaya[1]; Yimin Zhu[2]; Supratik Mukhopadhyay[3]; and Amirhosein Jafari[4]

[1]College of Engineering, Louisiana State Univ., Baton Rouge, LA. E-mail: cchokw1@lsu.edu
[2]College of Engineering, Louisiana State Univ., Baton Rouge, LA. E-mail: yiminzhu@lsu.edu
[3]College of Computer Science and Engineering, Louisiana State Univ., Baton Rouge, LA. E-mail: supratik@csc.lsu.edu
[4]College of Engineering, Louisiana State Univ., Baton Rouge, LA. E-mail: ajafari1@lsu.edu

ABSTRACT

Immersive virtual environments (IVEs) have been widely used as multipurpose tools in many research areas. As design tools, IVEs have shown the potential to observe human-building interactions for buildings under design. IVEs allow researchers or designers to understand human-building interactions and refine building functions to optimally fulfill design goals. However, due to limitations of the technology, practical use of IVE often result in small sample sizes, which may negatively impact those applications (e.g., machine learnings) that require a large amount of data. This paper demonstrates the application of a Gaussian mixture model (GMM) as a method for generating independent and identically distributed (IID) samples of data originally obtained from IVE applications to solve the issue of the small sample size associated with IVE experiments. In this study, GMM is tested using an application that involves light switch uses in a single office. First, an IVE of the office is created to simulate key artificial light use events during design. Then, light switch locations, tasks, work area illuminance, and factors potentially influencing light switch uses are modeled in the IVE experiment. Finally, thirty people participated in the experiment to collect data. The results of this study show that the IVE data and the IID samples are not significantly different, affirming that the GMM can be a good candidate for tackling a small sample size issue associated with IVE experiments.

INTRODUCTION

Many studies have proven that building designs significantly influence building performance in several aspects, such as energy consumption, occupancy comfort, and sustainability (Granadeiro et al. 2013, and Wang et al. 2005). Human-building interactions are one of the major contributing factors to large uncertainties in building performance (Zhu et al. 2018). A large body of research has been dedicated to study the impact of human-building interactions on building performances. For instance, Clevenger and Haymaker (2006) applied building energy simulation to study the impact of human-building interactions. They showed that the predicted energy consumption changed at least 150% if the maximum and minimum values of inputs related to human-building interactions were applied to the simulations. Santin et al. (2009) applied the historical survey data from 15,000 houses along with three years of energy usage to conduct a statistical analysis related to human-building interactions affecting energy consumptions. They reported that human-building interactions contributed to 7.2% of energy consumption variation. Similar results were found by Kavousian et al. (2013) and D’Oca et al. (2014). Therefore, human-building interactions are one of the most significant factors that cause variations in building energy consumption, even though building envelopes, functions, and environments may be identical. Human-building interactions are key factors that should be considered during building designs to satisfy the goals and objectives of buildings.

Several studies have dedicated to investigate human-building interactions, as it is mentioned previously. They mainly used traditional methods, e.g., questionnaires (Attia et al. 2012, and Feng et al. 2016), filed studies (Rijal et al. 2007, and Amir Seyed et al. 2016), and laboratories (Wienold and Christoffersen 2006, and Schweiker and Wagner 2016) to observe and collect human-building interaction data. However, the traditional methods typically rely on observations of human-building interactions in existing buildings. Therefore, applying such information to new designs may lead to significant failures of accompishing design goals and objectives.

Immersive virtual environment (IVE) is a new tool that has been used in several research fields such as emergency evacuation (Kinateder et al. 2014 and Kobes et al. 2010), building designs (Tutt and Harty 2013, and Hong and Michalatos 2016), and occupant behavior predictions (Kwon et al. 2013 and Chokwitthaya et al. 2017). A growing number of research studies have recently highlighted the potential of IVE in studying human-building interactions. Heydarian et al. (2015) studied occupant lighting preferences in a single office using IVE. They concluded that IVE helped users to understand human-building interactions and satisfaction towards different design options. Saeidi et al. (2015) validated human-building interaction in light switching by using IVEs and showed that IVEs were capable of replicating experiences in a real environment. Niu et al. (2015) developed a framework to integrate building designs with IVEs to help building designers capture human-building interactions and identify context patterns. Recently, Saeidi et al. (2018) proposed a method to design experiments and collect human-building interaction data in IVEs, called spatial-temporal event-driven modeling (STED). They validated human-building interaction data obtained from an IVE experiment and statistically compare the data with human-building interaction obtained from a physical environment. They confirmed that IVEs had the capability of capturing human-building interactions since human-building interaction data obtained from the IVE experiment were not significantly different from those obtained from the real environment.

Potentially, IVEs provide several advantages for researchers and designers to capture human-building interactions such as replicating the context of buildings under design, allowing designers or researchers to control experimental conditions, and including interested variables, which may be difficult to do using traditional methods. Consequently, IVEs have strong potentials to support designers or researchers in observing and capturing human-building interactions in simulated building context during designs. However, IVEs have various limitations, including short experiment sessions, small data samples, and negative impacts on participants (Chokwitthaya et al. 2019a). Such limitations make it impossible to continuously collect human-building interactions data in IVEs for a long period of time. Therefore, most IVE experiments can only provide a small sample size that may not be sufficient to support sophisticated data-driven applications such as machine-learning algorithms (Chokwitthaya et al. 2018).

Several research studies have illustrated that data generated using machine learning techniques could be highly realistic, robust, and meaningful to replace real data (Jaderberg et al. 2014) , and to be used for training and learning in machine-learning algorithms, especially when real data are scarce, and/or cannot be unveiled (Le et al. 2017, Tremblay et al. 2018).

A Gaussian Mixture Model (GMM) (Bishop 2006) is one of the outstanding alternative approaches for generating a synthetic dataset based on data from IVE experiments (i.e., increasing the number of IID samples). A GMM is an unsupervised classification method, defined as a convex combination of multiple Gaussian (normal) distribution with individual means and variances (Singh et al. 2010). The GMM has been proven to have better performance than many other clustering methods such as k-means (Guestrin 2007), k-nearest neighbor (Lanjewar et al. 2015), and multivariate kernel density (MVKD) (Morrison 2011). The GMM has been applied for data clustering in several research areas such as speech recognition (Muthusamy et al. 2015 and Lanjewar et al. 2015), false detection (Bo Zong et al. 2018), and image segmentation (Ragothaman et al. 2016). In a field of generating IID samples, the GMM has widely been used. Awwad et al. (2005) adapted the sequential EM algorithm in a GMM to generate synthetic data for adaption failure detection, where the data slowly change over time. In an IVEs-related study, Liu et al. (2019) used a GMM to augment the raw dataset obtained from the driving experiment in IVE and produce synthetic driving record. With the proven potentials of the GMM, it is a good candidate to generate IID samples of human-building interaction data obtained from an IVE experiment that can reduce the impact of a small sample size and enhance the effectiveness of IVE data.

This paper presents the application of a GMM as a method to learn samples of IVE experiments and generate large IID dataset. In the following, the authors first define the research hypothesis, and then introduce the methodology along with the case study. Then, the results, conclusions, and future work are discussed based on the application.

HYPOTHESIS

The objective of this study is to prove that IID samples generated by GMM can represent human-building interaction data obtained from an IVE experiment. Therefore, the hypothesis tested in this study is that there is no statistically significant difference in means between IVE data and IID samples predicted using a GMM. A two-tailed t-test (α = 0.05, equal variance) is applied to investigate the statistically significant difference between the mean of the IVE data and the mean of the IID samples. The following equations represent the test of the hypothesis in this study:

$$H_0: \textit{mean of the IVE data} - \textit{mean of the IID samples} = 0$$
$$H_1: \textit{mean of the IVE data} - \textit{mean of the IID samples} \neq 0$$

METHODOLOGY AND APPLICATION

Human-Building Interactions in IVE Experiment

The experimental configuration (Figure 1) is designed using the recommended space of a large private office with a dimension of 5.5 x 4.2 x 3.2 meters and a net area of 22 square meters (Voss 2000). The model is designed and drawn by using Autodesk AutoCAD 3D. Autodesk 3DSmax is used to render the environmental brightness of the IVE experiment. Then, it is exported to Unreal Engine 4 (UE4) for constructing the IVE experiment, as shown in Figure 2.

The human-building interaction on light switching is selected in the case study since the visual simulation is the most matured IVE capability (Chokwitthaya et al. 2019b). Variables considered in IVE are selected based on previous literature that shows influences of them on light switch uses, such as light switch locations (Heydarian et al. 2015), office tasks (Boyce 2004), and work area illuminance (Hunt 1980). Table 1 shows variables considered in the IVE experiment, which involve switch locations, office tasks, and work area illuminance. Audio cues remind participants about tasks that they are supposed to do during the experiment as well as the locations of the switch. The actual switch is presented in the experiment.

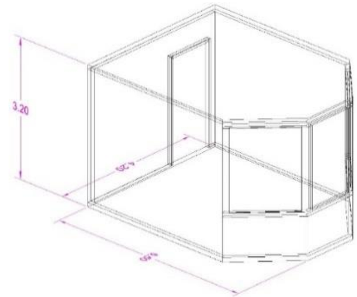

Figure 1. The configuration in Autodesk AutoCAD 3D

Figure 2. The configuration in the IVE

Table 1. Conditions of variables considered in the IVE experiment

Variables considered in IVE experiment		
Switch location	***Office task***	***Work area illuminance (lux)***
On the wall by an entrance door	Reading or writing which will take an hour	50
		100
On the desk	Relaxing (e.g., lunch break) which will take an hour	150
		200
	A meeting which will take an hour	350
	Precision drawing which will take an hour	500
Total = 2	*Total = 4*	*Total = 6*

Thirty college students participated in the experiment. They used a head-mounted display (HMD) to experience the IVE. They were placed at the desk and assumed to be an owner of the office, who could operate the switch. The experiment comprises a training part and an experimenting part. The training part introduces and immerses the participants with the IVE, which takes about 10 minutes. Then, the experimenting part begins. The experimenting part involves 48 scenarios according to variables in Table 1 (2 switch locations x 4 office tasks x 6 work area illuminance), which takes about 40 minutes. The study has been approved by the local Institutional Review Board (IRB).

During the experiment, the participants are asked to answer the following question "*Please rate your need of turning the light on under the provided situation*." The participants answer the questions by selecting one of the choices shown in the experiment (i.e., very unlikely, not likely, neutral, likely, and very likely) designed based on Likert scale (Likert 1932). The interpretation of Likert scale for the IVE experiment is demonstrated in Table 2.

Table 2. Interpretation of Likert scale in the IVE experiment

Choice	*Very unlikely*	*Not likely*	*Neutral*	*Likely*	*Very likely*
Probability of switching on (%)	1	25	50	75	99

GMM to Generate IID Samples from the IVE Experimental Data

The GMM first learns the obtained IVE data and then produces IID samples. The IVE data are modeled in terms of a mixture of multiple Gaussian components (*1, …, z*), where each component has its own mean (μ) and variance (σ^2). The GMM assumes that each IVE data point (x) probabilistically belong to z, and it tries to infer the distribution for each component separately. Since each x probabilistically belongs to z, a mixing coefficient (π_z) is introduced as a probability that x is drawn from z. Therefore, the joint distribution of x given z (i.e., $p(x, z)$) can be defined regarding a marginal distribution (i.e., $p(z)$) and a conditional distribution (i.e., $p(x|z)$), as it is shown in Equation (1) (Awwad et al. 2005):

$$p(x,z)=\sum_{z} p(z)\,p(x\,|\,z)=\sum_{z=1}^{Z}\pi_z Gaussian\left(x\,|\,\mu_z,\sigma_z^2\right), where \sum_{z=1}^{Z}\pi_z=1 \qquad (1)$$

The GMM is trained by implementing an expectation-maximization (E-M) algorithm (Bishop 2006). It comprises of an expectation step and a maximization step (see Figure 1). The expectation step (E-step) computes expectations, or responsibilities (i.e., $\gamma(z)$) by using current parameter values (i.e., μ_z, σ_z^2, and π_z). The maximization step (M-step) re-estimates parameters using the expectations obtained from E-step. The maximum likelihood for Gaussian determines the convergence of the training process. In other words, the E-M algorithm tries to maximize Equarion (2) based on convergence criteria. For a best model selection, the values of the Bayesian Information Criterion (BIC) (Scott Shaobing and Gopalakrishnan 1998) and the Akaike Information Criterion (AIC) (Akaike 1998) are calculated to determine an optimal number of components (Z). The best model is the model that has minimum BIC or AIC.

$$\sum_{i=1}^{N}\gamma(z)\,log\left(x\,|\,z\right)-\sum_{i=1}^{N}\gamma(z)\log Gaussian\left(x\,|\,\mu_z,\sigma_z^2\right)<convergence\,criterion \qquad (2)$$

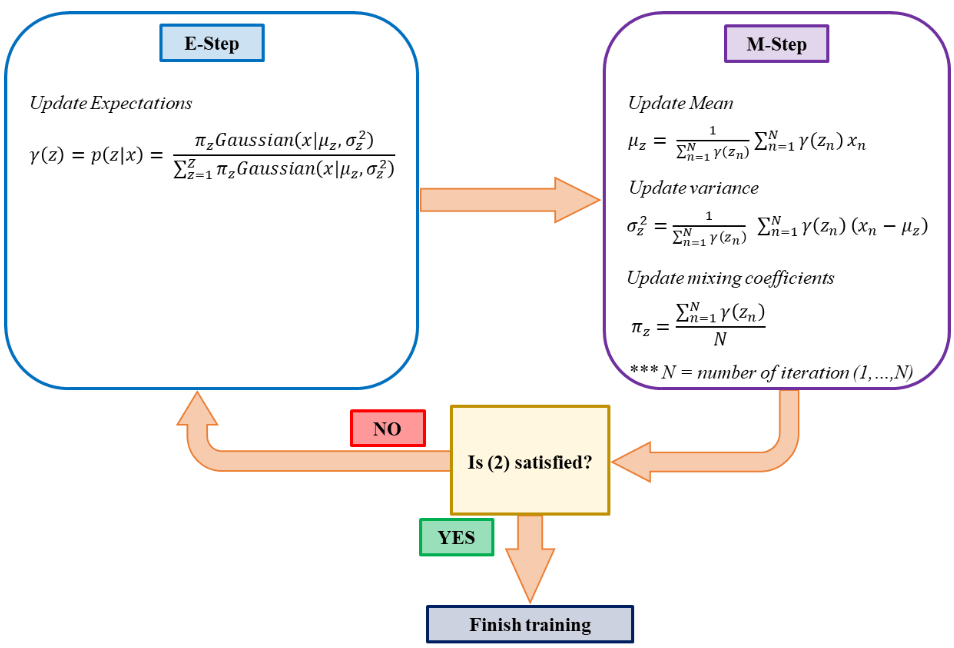

Figure 3. E-M algorithm

In this case study, the *K*-mean algorithm is used to initialize the GMM parameters (Lee 2004). The IVE data are classified into four groups regarding the office tasks to avoid having too many GMM components in this case study. The convergence criterion for training the GMM is 10^{-2}. After training, the GMM is executed to generate IID samples regarding IVE data.

RESULTS

Table 1 illustrates the overall results of the case study, including the number of components of the optimal GMM models, means and standard errors (S.E.) of the probabilities of switching on, and the hypothesis testing for each individual group. From the values of BIC and AIC, the number of components of the optimal GMM models is obtained based on the criterion mentioned previously. A huge number of scenarios in the IVE experiment (48 scenarios) result in a widespread of the IVE data and a high number of components. From the results of the hypothesis testing, P-value is greater than 0.05 for all cases, meaning H_0s are accepted for all cases. Therefore, the conclusion of the hypothesis testing can be drawn as there are no significant differences between the means of the IVE data and the IID samples predicted by the GMM. The results of hypothesis testing affirm that GMM is capable of being used as an alternative tool for generating IID samples of IVE data.

Table 3. Results of the case study

Group	**GMM component**	**Dataset**	**Mean ± S.E. of the probability of switching on**	**Hypothesis Testing**	
Drawing	47	*IVE*	0.59 ± 0.018	*P-value*	0.753
		IID sample	0.59 ± 0.006	*Accept H_0?*	Yes
Meeting	48	*IVE*	0.62 ± 0.018	*P-value*	0.931
		IID sample	0.62 ± 0.006	*Accept H_0?*	Yes
Reading	48	*IVE*	0.58 ± 0.018	*P-value*	0.800
		IID sample	0.59 ± 0.006	*Accept H_0?*	Yes
Relaxing	44	*IVE*	0.30 ± 0.015	*P-value*	0.157
		IID sample	0.27 ± 0.005	*Accept H_0?*	Yes

CONCLUSION AND FUTURE WORK

In this paper, an application of GMM for generating IID samples of IVE data was illustrated. The presented case study used an IVE application to capture human-building interactions on the use of a light switch. Then, a GMM was trained by using the human-interactions data obtained from the IVE experiment. The IID samples were drawn for the trained GMM model. The IVE data and the IID samples were statistically compared by using the two-tailed t-test. The result of hypothesis testing demonstrated the potential of the GMM for generating IID samples of IVE data.

The main contribution of this study is to solve the problem of small sample sizes associated with IVE experiments by drawing a large pool of IID samples from the GMM model. The main advantage of having a large pool of IID samples of IVE data is to support further analyses that require large datasets, such as machine-learning algorithms.

It should be noted that a major limitation of a GMM is overfitting, as it is mentioned in literature (Valente 2003, and Bone et al. 2011). Besides, only the mean values are used in this study to assess the quality of the synthetic generated data. As future work, additive white Gaussian noises (AWGN) will be added to the IVE data to prevent overfitting in the training process. Moreover, the diversity of participants may need to be considered to reduce the variation of IVE data and fit certain designs. It will be done using questionnaires to investigate the demographics and energy behaviors of participants. In addition, the satisfactory level of sense of presence needs to be analyzed to evaluate the quality of the IVE data in future works.

ACKNOWLEDGMENT

This paper is supported by the U.S. National Science Foundation Award #1640818. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Akaike, H. (1998). “Information theory and an extension of the maximum likelihood principle.” *Selected papers of hirotugu akaike*, Springer, New York, NY, 199–213.

Amir Seyed, Karava Panagiota, Konstantzos, and Iason. (2016). “Occupant interactions with shading and lighting systems using different control interfaces : A pilot fi eld study.” *Building and Environment*, 97, 177–195.

Attia, S., Evrard, A., and Gratia, E. (2012). “Development of benchmark models for the Egyptian residential buildings sector.” *Applied Energy*, Elsevier Ltd, 94(2012), 270–284.

Awwad, B., Hasan, S., and Gan, J. Q. (2005). “Machine Learning and Data Mining in Pattern Recognition.” *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer Berlin Heidelberg, 96–106.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, Berlin, Heidelberg.

Bo Zong, Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., and Chen, H. (2018). “Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection.” *Iclr2018*, (2016), 1–13.

Bone, D., Black, M. P., Li, M., Metallinou, A., Lee, S., and Narayanan, S. S. (2011). “Intoxicated speech detection by fusion of speaker normalized hierarchical features and GMM supervectors.” *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (August), 3217–3220.

Boyce, P. (2004). “Lighting research for interiors: the beginning of the end or the end of the beginning.” *Lighting Research & Technology*, 36(4), 283–293.

Chokwitthaya, C., Collier, E., Zhu, Y., and Mukhopadhyay, S. (2019a). “Improving Prediction Accuracy in Building Performance Models Using Generative Adversarial Networks (GANs).” *arXiv preprint arXiv:190605767 2019 June 13*, arXiv preprint arXiv:1906.05767. 2019 June 13.

Chokwitthaya, C., Dibiano, R., Saeidi, S., Mukhopadhyay, S., and Zhu, Y. (2018). “Enhancing the Prediction of Artificial Lighting Control Behavior using Virtual Reality (VR): A Pilot Study.” *In Construction Research Congress 2018*, 216–233.

Chokwitthaya, C., Saeidi, S., Zhu, Y., and Kooima, R. (2017). “The Impact of Lighting Simulation Discrepancies on Human Visual Perception and Energy Behavior Simulations in

Immersive Virtual Environment." *Congress on Computing in Civil Engineering, Proceedings*.

Chokwitthaya, C., Zhu, Y., Dibiano, R., and Mukhopadhyay, S. (2019b). "Combining context-aware design-specific data and building performance models to improve building performance predictions during design." *Automation in construction*, Elsevier, 107(July), 102917.

Clevenger, C. M., and Haymaker, J. (2006). "The Impact Of The Building Occupant On Energy Modeling Simulations." *In Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, 1–10.

D'Oca, S., Fabi, V., Corgnati, S. P., and Andersen, R. K. (2014). "Effect of thermostat and window opening occupant behavior models on energy use in homes." *Building Simulation*, 7(6), 683–694.

Feng, X., Yan, D., Wang, C., and Sun, H. (2016). "A preliminary research on the derivation of typical occupant behavior based on large-scale questionnaire surveys." *Energy and Buildings*, Elsevier B.V., 117, 332–340.

Granadeiro, V., Duarte, J. P., Correia, J. R., and Leal, V. M. S. (2013). "Building envelope shape design in early stages of the design process: Integrating architectural design systems and energy simulation." *Automation in Construction*, Elsevier B.V., 32(September), 196–209.

Guestrin, C. (2007). "Unsupervised learning or Clustering – Gaussian mixture models." *Machine Learning*, 10701.

Heydarian, A., Carneiro, J. P., Gerber, D., and Becerik-Gerber, B. (2015). "Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance." *Building and Environment*, Elsevier Ltd, 89, 217–228.

Hong, Y., and Michalatos, P. (2016). "LumiSpace: A VR Architectural Daylighting Design System." *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, 1–2.

Hunt, D. R. G. (1980). "Predicting Artificial Lighting Use- A Method Based Upon Obseved Patterns of Behavior." *Lighting Research and Technology*, 12(1), 7–14.

Jaderberg, M., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). "Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition." 1–10.

Kavousian, A., Rajagopal, R., and Fischer, M. (2013). "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior." *Energy*, Elsevier Ltd, 55, 184–194.

Kinateder, M., Müller, M., Jost, M., Mühlberger, A., and Pauli, P. (2014). "Social influence in a virtual tunnel fire - Influence of conflicting information on evacuation behavior." *Applied Ergonomics*, 45(6), 1649–1659.

Kobes, M., Helsloot, I., De Vries, B., and Post, J. (2010). "Exit choice, (pre-)movement time and (pre-)evacuation behaviour in hotel fire evacuation - Behavioural analysis and validation of the use of serious gaming in experimental research." *Procedia Engineering*, Elsevier, 3(December 2010), 37–51.

Kwon, J. H., Powell, J., and Chalmers, A. (2013). "How level of realism influences anxiety in virtual reality environments for a job interview." *International Journal of Human Computer Studies*, Elsevier, 71(10), 978–987.

Lanjewar, R. B., Mathurkar, S., and Patel, N. (2015). "Implementation and comparison of speech emotion recognition system using Gaussian Mixture Model (GMM) and K-Nearest Neighbor

(K-NN) techniques." *Procedia Computer Science*, 49(1), 50–57.
Le, T. A., Baydin, A. G., Zinkov, R., and Wood, F. (2017). "Using synthetic data to train neural networks is model-based reasoning." *Proceedings of the International Joint Conference on Neural Networks*, 2017-May, 3514–3521.
Lee, K. Y. (2004). "Local fuzzy PCA based GMM with dimension reduction on speaker identification." *Pattern Recognition Letters*, 25(16), 1811–1817.
Likert, R. (1932). "A technique for the measurement of attitudes." *Archives of psychology*.
Liu, Q., Mukhopadhyay, S., Zhu, Y., and Gudishala, R. (2019). "Improving Route Choice Models by Incorporating Contextual Factors via Knowledge Distillation."
Morrison, G. S. (2011). "A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model-universal background model (GMM-UBM)." *Speech Communication*, Elsevier B.V., 53(2), 242–256.
Muthusamy, H., Polat, K., and Yaacob, S. (2015). "Improved emotion recognition using gaussian mixture model and extreme learning machine in speech and glottal signals." *Mathematical Problems in Engineering*, 2015.
Niu, S., Pan, W., and Zhao, Y. (2015). "A virtual reality integrated design approach to improving occupancy information integrity for closing the building energy performance gap." *Sustainable Cities and Society*, Elsevier B.V., 27, 275–286.
Ragothaman, S., Narasimhan, S., Basavaraj, M. G., and Dewar, R. (2016). "Unsupervised Segmentation of Cervical Cell Images Using Gaussian Mixture Model." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, 1374–1379.
Rijal, H. B., Tuohy, P., Humphreys, M. A., Nicol, J. F., Samuel, A., and Clarke, J. (2007). "Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings." *Energy and Buildings*, 39(7), 823–836.
Saeidi, S., Chokwitthaya, C., Zhu, Y., and Sun, M. (2018). "Spatial-temporal event-driven modeling for occupant behavior studies using immersive virtual environments." *Automation in Construction*, Elsevier, 94(May), 371–382.
Saeidi, S., Rizzuto, T., Zhu, Y., and Kooima, R. (2015). "Measuring the Effectiveness of an Immersive Virtual Environment for the Modeling and Prediction of Occupant Behavior." *Sustainable Human-Building Ecosystems - Selected Papers from the 1st International Symposium on Sustainable Human-Building Ecosystems*, (2006), 159–167.
Santin, O. G., Itard, L., and Visscher, H. (2009). "The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock." *Energy and Buildings*, 41(11), 1223–1232.
Schweiker, M., and Wagner, A. (2016). "The effect of occupancy on perceived control, neutral temperature, and behavioral patterns." *Energy and Buildings*, Elsevier B.V., 117, 246–259.
Scott Shaobing, C., and Gopalakrishnan, P. S. (1998). "Clustering Via the Bayesian Information Criterion With." *International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat No 98CH36181)*, IEEE, Seattle, Washington, 645–648.
Singh, R., Pal, B. C., and Jabr, R. A. (2010). "Statistical Representation of Distribution System Loads Using Gaussian Mixture Model." *Power Systems IEEE trans on*, 25(1), 29–37.
Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., and Birchfield, S. (2018). "Training deep networks with synthetic data: Bridging the reality gap by domain randomization." *IEEE Computer Society Conference on*

Computer Vision and Pattern Recognition Workshops, IEEE, 2018-June, 1082–1090.
Tutt, D., and Harty, C. (2013). "JJourneys through the CAVE: The use of 3D immersive environments for client engagement practices in hospital design." *Proceedings 29th Annual Association of Researchers in Construction Management Conference, ARCOM 2013*, (September), 111–121.
Valente, F. (2003). "Variational Bayesian GMM for speech recognition." *European Conference on Speech*, 441–444.
Voss, J. (2000). "Revisiting Office Space Standards." *Grand Rapids, MI: Haworth*, 1–6.
Wang, W., Zmeureanu, R., and Rivard, H. (2005). "Applying multi-objective genetic algorithms in green building design optimization." *Building and Environment*, 40(11), 1512–1525.
Wienold, J., and Christoffersen, J. (2006). "Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras." *Energy and Buildings*, 38(7), 743–757.
Zhu, Y., Saeidi, S., Rizzuto, T., Roetzel, A., and Kooima, R. (2018). "Potential and challenges of immersive virtual environments for occupant energy behavior modeling and validation: A literature review." *Journal of Building Engineering*, 19.