

Human-in-the-Loop Bayesian Optimization Approach to Supporting Early-Stage Architectural Design

Masaya Tanaka¹, Kotaro Ono² and Satoshi Yamada³

^{1,2,3}*Ritsumeikan University.*

¹*ru0108kf@gmail.com, 0009-0000-5604-908X*

²*ru0043ev@ed.ritsumei.ac.jp, 0009-0004-0649-6120*

³*sy@fc.ritsumei.ac.jp, 0009-0009-3653-7196*

Abstract. Recent advancements in Computer Aided Design (CAD) applications have improved access to parametric design; however, their optimization methods have predominantly focused on quantitative metrics, such as environmental or structural performance. Consequently, subjective aspects are typically addressed by manual "slider" adjustment, which is often inefficient, susceptible to cognitive biases, and prone to entrapment in local optima. Herein we propose a human-centric computational design framework that integrates human-in-the-loop (HITL) and Bayesian optimization (BO) to incorporate subjective evaluation into the design process. We demonstrated this framework using a pavilion design task defined by six parameters and evaluated it via a user study (N = 40) comparing the proposed "Bayesian method" with the conventional "slider-based" method. The results demonstrated that the Bayesian method outperformed the slider-based method in final design quality, overall satisfaction, and perceived design diversity, while helping users discover their latent preferences by exploring parameter regions that might have been overlooked manually. Furthermore, "transparency" of the BO process was essential for fostering user trust and preserving a sense of agency. This approach shows promise as an interactive tool for facilitating exploratory design in the early stages, particularly for users with uncertain preferences.

Keywords. Bayesian optimization, human-in-the-loop, human sensibility, parametric design, computational design

1. Introduction

1.1. CHALLENGES OF PARAMETRIC DESIGN

Recently, the integration of Computer Aided Design (CAD) with procedural modelling, as exemplified by Rhinoceros/Grasshopper, has significantly expanded parametric design capabilities, allowing for the engineering of complex geometries and systems. However, prevailing optimization techniques face challenges from a human-centric viewpoint, because they emphasize theoretical and quantitative metrics, such as

environmental or structural performance, as objective functions. Although effective for engineering optimization, these methods overlook human sensibility, which is a crucial aspect of the design process. This oversight creates challenges in human-centric design processes. Consequently, in early design stages that require subjective evaluation, designers often depend on manual adjustments and inspections of numerous parameters. While manual exploration may be effective in low-dimensional parameter spaces, it becomes increasingly complex in higher dimensions, risking the neglect of preferred design areas, often resulting in confinement to local optima.

1.2. PRIOR RESEARCH AND THIS STUDY'S POSITIONING

Integration of such "sensibility" into the design process has been attempted in prior studies. For instance, interactive evolutionary computation (IEC), such as interactive genetic algorithms, utilizes human subjective selection as a fitness function (Takagi, 1998; Muley et al., 2003). However, these methods required feedback on numerous samples, incurring significant user evaluation costs that limited the exploration scope.

Bayesian Optimization (BO) has attracted attention as a promising solution to this evaluation cost problem. By modelling unknown functions using Gaussian processes and prioritizing sample efficiency, BO optimizes functions with fewer evaluations.

Approaches involving the application BO to subjective evaluation can be classified into two categories. The first utilizes crowdsourcing to derive generalized optimal parameters across a large user base, such as in parameter space analysis for visual design (Koyama et al., 2014) and interface design optimization (Dudley et al., 2019). The second aims to identify parameters personalized for an individual, with applications, including maximizing game engagement (Khajah et al., 2016) and optimizing font settings for reading speed (Kadner et al., 2021). In this study, we adopted the latter "individual-centric" approach.

For individual adaptation, Preferential Bayesian Optimization (PBO) using pairwise comparisons is a representative alternative (Brochu et al., 2010). However, we did not adopt this approach owing to issues, such as local optima observed in preliminary tests, coupled with the lack of absolute score information. To maximize the information obtained from a single evaluation and achieve efficient exploration, a BO approach using direct numerical evaluation was adopted in this study.

1.3. RESEARCH OBJECTIVES

This study has three main objectives.

- To build a human-centric computational design framework that integrates human-in-the-loop (HITL) and BO, and then places human sensibility in parametric design within a Rhinoceros/Grasshopper environment.
- To quantitatively verify, through a user study, the effectiveness of the proposed method in helping users discover their latent preferences and contribute to finding higher-quality design solutions.
- To identify the types of users, based on design experience and exploratory preferences, for whom the proposed method is beneficial in the early design stages.

2. Methodology

2.1. TWO EXPLORATION METHODS

To achieve optimization based on a subjective evaluation of the parametric design, we compared two exploration methods.

- Micro-exploration method (slider-based), which allows users to directly manipulate sliders for intuitive design creation.
- Subjective evaluation optimization method (Bayesian), which combines HITL and BO approaches for optimal design exploration.

2.2. SENSIBILITY EVALUATION OPTIMIZATION METHOD (BAYES- IAN)

2.2.1. System Overview and Evaluation Method

Figure 1 schematically shows the proposed system combining HITL and BO.

To incorporate human input into the optimization process, participants were asked to evaluate the presented objects on an integer scale from 0 to 100, considering multiple criteria, such as beauty, functionality, and novelty.

A comprehensive score was chosen over individual evaluation axes (e.g., beauty, functionality, and novelty) to avoid reduction in the information content of the evaluation emerging from using a finite set of axes and reduce the evaluation cost associated with multiple criteria.

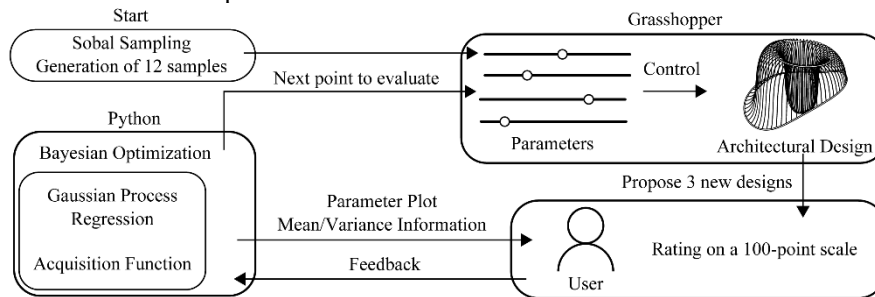


Figure 1. System Overview of the Bayesian Optimization for Architectural Design.

2.2.2. Formulation of Bayesian Optimization

BO is well-suited for optimizing the evaluation functions of black-box phenomena, particularly in multi-dimensional parameter spaces involving high-cost computations.

The BO approach employs Gaussian process regression (GPR) to estimate the unknown evaluation function $y = f(X)$ from input variables X to an output variable y . GPR assumes that function $f(X)$ follows the Gaussian process

$$f(X) \sim GP(\mu(X), k(X, X')), \quad (1)$$

where $\mu(X)$ is the mean function, and $k(X, X')$ is the kernel function. We used a

common radial basis function kernel expressed as

$$k(X, X') = \exp\left(-\frac{\|X - X'\|^2}{2l^2}\right), \quad (2)$$

where l is the characteristic length scale.

GPR can provide predictions for unknown points as a mean value, along with uncertainty (variance). The acquisition function employs this uncertainty to determine the next point(s) for evaluation by balancing exploitation and exploration.

We used qLogNEI (3) to simultaneously present multiple candidate solutions (a batch) and accommodate noisy human subjective evaluations.

$$qLogNEI(X) = E \left[\max \left(0, \max_{j=1, \dots, q} \tilde{g}(x_j) - g_{best} \right) \right], \quad (3)$$

where g_{best} is the maximum observed evaluation value, and $\tilde{g}(x_j)$ is the predictive distribution of the evaluation value at a candidate point x_j , as predicted by the GPR model (1). The batch size was set to $q = 3$. This value is practically approximated using Monte Carlo sampling.

2.2.3. Process Disclosure (Transparency)

During the experiment, Figure 2 was presented to the participants to ensure process transparency. The plot displays parameter locations of the current and previously evaluated designs. The colour of each point reflects the participant's evaluation score. Information regarding the dispersion of the evaluation values estimated from the GPR and the directionality of the subsequent suggestions based on this information, indicated by a text label ("Exploitation-oriented" or "Exploration-oriented"), were provided. For simplicity, the dispersion was presented as a "range."

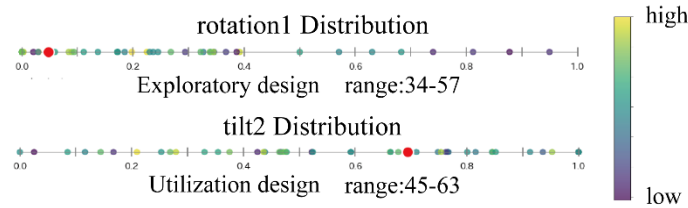


Figure 2. Part of the provided figure.

2.3. EVALUATION EXPERIMENT

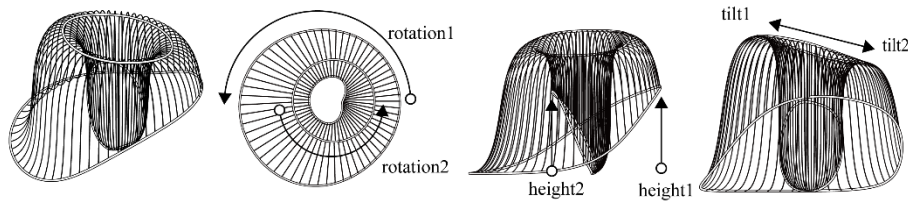


Figure 3. Definition of Design Parameters.

Based on our previous study (2025), we conducted an experiment using a pavilion with six design variables: rotation1, rotation2, tilt1, tilt2, height1, and height2, each normalized to a range from 0 to 1 (Figure 3).

2.3.1. Experimental Procedure

Figure 4 shows the operation interfaces used in this experiment for the "Bayesian method" and the "slider-based method." Each participant was asked to create a design for the same task using both the Bayesian and slider-based methods. The order of the methods was randomized to consider order and adaptation effects. Subsequently, the participants were asked to complete a questionnaire.

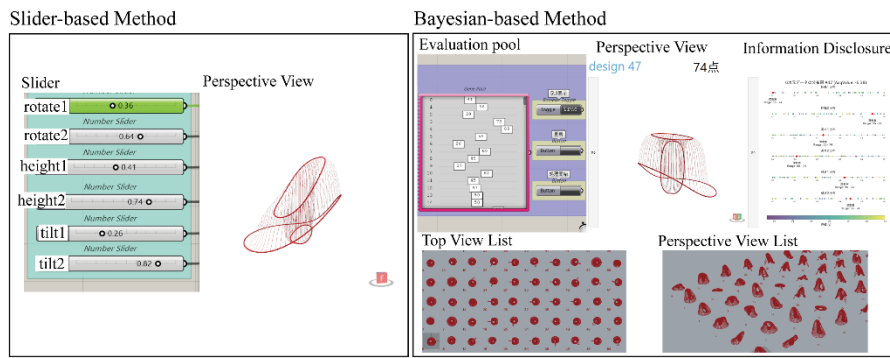


Figure 4. Left: Slider Method UI Right: Bayesian Method UI

2.3.2. Evaluation Items

In the questionnaire, the participants were asked to indicate which method ("Bayesian method," "Slider-based method," or "Neither") was superior across five evaluation criteria:

1. Final design selection
2. Design Diversity
3. Ease of Convergence
4. Understanding and Ease of Use
5. Overall Satisfaction

In addition, the participants were asked to rate on a five-point scale the extent to which the following supplementary information, presented during the design process, influenced their design actions and final evaluation, or how helpful it was.

1. Process Disclosure: includes 2–4
2. Range Disclosure
3. Labelling
4. Design Parameter Plot

Furthermore, we solicited free-text responses regarding participants' impressions of each method and the reasons for their evaluations. The response data were aggregated, and the evaluation trends for both methods across each criterion were analysed.

2.3.3. Statistical Analysis

We conducted an analysis based on both quantitative data from the final design and subjective evaluation data from the questionnaire. In the quantitative data analysis, we assessed the similarity of solutions from both methods based on correlation coefficients. In the subjective evaluation data analysis, we evaluated the significance of the selection proportion difference between the "Bayesian method" and "slider-based method" using a proportion test.

3. Results

3.1. EXPLORATION ANALYSIS OF PARAMETER SPACE

Figure 5 presents the distribution graphs and statistical analysis for each parameter of the final design solutions obtained by both methods ("slider-based" and "Bayesian").

The "slider-based method" exhibited a tendency for parameters to concentrate around intermediate values (near 0.5). However, in the "Bayesian method," many participants adopted extreme values, such as 0 or 1.

The correlation analysis results indicated that although a correlation was observed between the two methods for parameters related to "height," there was no correlation for many of the other parameters.

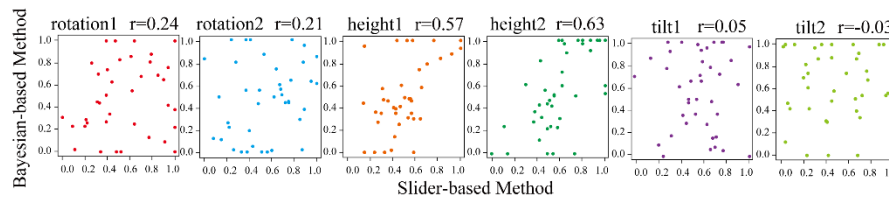


Figure 5. Comparison of Final Parameter Values: Bayesian (Y-axis) vs. Slider (X-axis). 'r' is the correlation coefficient.

3.2. SUBJECTIVE EVALUATION OF DESIGN METHODS

Figure 6 shows the results of the participants' subjective evaluations of the two design methods and their post-experiment evaluations of the information (labelling, range disclosure, and design parameter plot) presented during the design process.

The "Bayesian-based" method significantly outperformed the "Slider-based" method on key indicators measuring the effectiveness of the research. For the final design selection, 62.5% of the participants chose the "Bayesian-based" proposal, demonstrating a statistically significant difference (proportion test, $p = 0.025 < 0.05$). Similarly, for overall satisfaction, the "Bayesian-based" method received 62.5% support and was rated significantly higher ($p = 0.006 < 0.05$). Furthermore, 85% preferred the "Bayesian-based" method for design diversity, representing a highly statistically significant difference ($p < 0.001$). However, for ease of convergence, the "Slider-based" method gained 52.5% support, but no statistically significant difference was observed ($p = 0.36 > 0.05$). Regarding understanding and ease of use, support for the "Bayesian-based" method was limited to 40% and evaluations were divided, but no significant difference was observed ($p =$

0.25 > 0.05).

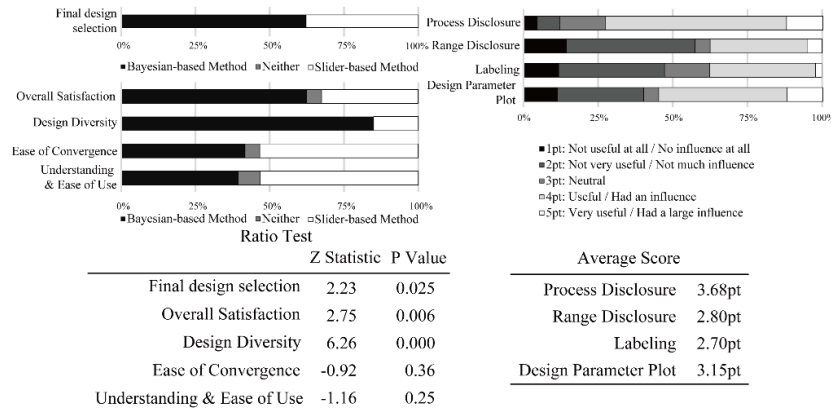


Figure 6. Results of Subjective Evaluations and Assessment of Disclosed Information.

Among individual pieces of disclosed information, the "design parameter plot" rated highest (3.15pt/5pt). Conversely, "labelling" and "range disclosure" rated relatively lower. However, "process disclosure," which integrated these elements, achieved a slightly positive evaluation (3.68pt). Sessions averaged 85 trials. Assuming 30s/trial, total interaction time averaged approximately 40 minutes, maintaining near real-time responsiveness. Participants favoring the Slider-based method completed slightly more trials on average than those preferring the Bayesian method.

4. Discussion

4.1. EFFECTIVENESS OF THE BAYESIAN METHOD

The most significant result of this study is that 63% of participants judged Bayesian method-generated designs to be preferable over designs created manually using the slider-based method.

This can be inferred from the exploration of the parameter space (Figure 5). When using the slider-based method, users unconsciously avoided extreme values (0 or 1) and tended to finalize their designs using intermediate values, which were easier to adjust. This indicates the potential influence of cognitive biases in manual operations.

By contrast, the Bayesian method, guided by its acquisition function, explores and presents parameter combinations that humans may psychologically avoid testing. This enabled the discovery of latent, high-quality design solutions that were unattainable using the slider-based method, yielding a high level of support for the final design using the Bayesian method. For example, Participant C (Figure 7) chose a Bayesian-generated design, which differed significantly in shape from a slider-based design.

Conversely, in cases where the slider-based design was chosen (e.g., Participants A and B), the designs from both methods tended to be similar in shape, probably because the operability of the slider-based method was evaluated during the final fine-tuning phase

after a specific design intent had been established.

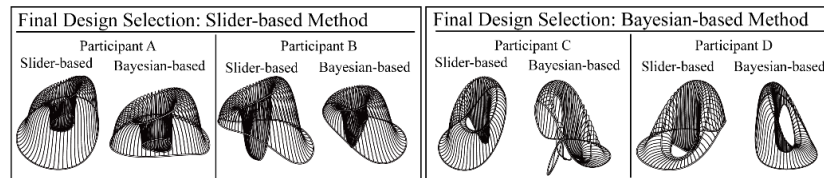


Figure 7. Examples of Final Design Selections by Participants.

4.2. EXPLORATION AND CONVERGENCE

In the subjective evaluation of design diversity, the Bayesian method received 85% support. This is consistent with Dedeler et al. (2024), who compared an IEC to a slider-based method (conventional parametric method) for pattern colorization. They reported conventional slider manipulation can cause "confusion or boredom" in users, whereas 77.7% of participants agreed the IEC was "exploratory."

Our average of 85 trials in our method suggests superior sample efficiency compared to typical IEC processes, which require larger population sizes and multiple generations to converge. While the 40-minute duration is non-negligible, the BO approach achieves high-quality exploration with a relatively lower evaluation count.

The slider-based method tended to converge on ordinary solutions, whereas the Bayesian method was superior in "diversity," indicating that computational methods support the user's exploratory design process.

Thus, the Bayesian method functioned as a strong facilitator of divergent exploration across the design space. By contrast, "ease of convergence" received comparable support for both the "slider-based" and "Bayesian" methods, possibly reflecting individual differences in user characteristics, as discussed below.

4.3. TRANSPARENCY

Chan et al. (2022) highlighted the risk of BO diminishing designers' subjective evaluations regarding "agency" and "creativity," and identified transparency as a key challenge. Therefore, to ensure transparency, our system explicitly presented participants with the parameter values proposed by the Bayesian method and the predicted range of evaluation scores (reflecting the expected value and variance).

The "Bayesian method" outperformed the "slider-based method" in final design selection and overall satisfaction. We infer that this transparency mitigated the black-box nature of BO, enabling users to preserve a sense of agency in their design process and engage in trial and error by trusting the artificial intelligence (AI) suggestions rather than feeling controlled by it. Notably, the positive evaluation for "the fact that information was disclosed" (3.68 pt) was higher than the evaluations for the "usefulness of the individual pieces of information" (ranging from 2.7 pt to 3.15 pt).

Therefore, regardless of content utility, the system's simple stance of disclosing its process and intent is crucial in fostering a collaborative relationship and building user trust with AI. In particular, nominal transparency is important.

4.4. USER CHARACTERISTICS

Regarding "divergence" in the design process, participants responded positively to both methods. However, differences in user characteristics emerged due to "convergence" and "evaluation," reflecting distinct orientations toward early-stage design.

The slider-based method is suitable for those with a "Design Refinement" orientation, who consider a "trial-and-error" approach to set a clear goal early and converge detailed toward it. Specifically, participants reported that this method offered a sense of conscious creation and allowed them to approach their preferences directly.

By contrast, the Bayesian method is suitable for those with a "Design Cognitive Extension" orientation, who consider an "intuitive" approach because their initial preferences are unclear. In particular, participants reported that this method facilitated the "extension" of design possibilities and the consideration of unexpectedly diverse designs, allowing them to finalize the design efficiently without starting from scratch.

This aligns with trial counts and behavioural observations. The slider-based method (Refinement type) had slightly more trials and longer experimental times than the Bayesian method (Extension type). This is because "refinement" requires deliberation, whereas "extension" enables rapid and intuitive decisions to discover new possibilities.

These results suggest that, depending on user characteristics, it is possible to efficiently support both "detailed examination" for refining designs and "discovery of possibilities" for extending design cognition.

5. Conclusion and Limitation

In this study, we proposed a method combining HITL and BO to support human-centric subjective evaluation in architectural design and validated its effectiveness on a six-parameter pavilion design task. The experimental results show that the proposed method (Bayesian method) helps users discover latent high-quality design solutions by promoting a broader exploration of the parameter space, thereby avoiding the cognitive bias seen in manual exploration (slider-based method) of selecting intermediate values.

A suitable method may differ depending on human-centric user characteristics. We observed that the "Refinement-oriented" users (aiming to fine-tune) preferred the "slider-based method," whereas the "Exploration-oriented" users (seeking discovery) preferred the "Bayesian method." This distinction may reflect the characteristics of designers and clients in practice, respectively.

Furthermore, the disclosure of the prediction process of BO demonstrated that even if it is "nominal transparency"—where the user does not reach a detailed understanding of the information's content—the system's act of disclosing its intent contributes to building trust in the AI and maintaining a human-centric, agentive design process.

This approach is expected to be utilized dually as an interactive tool for design idea creation and for visualizing clients' latent needs.

However, this study has limitations. The participants were limited to 40 design students, excluding experienced architects or non-expert clients. The task was restricted to a single six-dimensional pavilion problem; generalization to higher-dimensional or different design tasks requires caution. Additionally, this study relied on subjective evaluations. Future work should incorporate quantitative metrics to objectively validate

the exploration efficiency. Finally, validating Preference Bayesian Optimization (PBO) as an alternative to numerical evaluation remains important for future research.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP23K04201 (Co-Creation Between Deep Learning-Based Generative AI and Humans).

References

- Brochu, E., de Freitas, N., & Ghosh, A. (2007). Active preference learning with discrete choice data. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), *Advances in Neural Information Processing Systems 20* (pp. 409-416). Curran Associates, Inc.
- Chan, L., Liao, Y.-C., Mo, G. B., Dudley, J. J., Cheng, C.-L., Kristensson, P. O., & Oulasvirta, A. (2022). Investigating positive and negative qualities of human-in-the-loop optimization for designing interaction techniques. In *CHI Conference on Human Factors in Computing Systems (CHI '22)* (pp. 1-14). ACM.
<https://doi.org/10.1145/3491102.3501850>
- Dedeler, E. G., & Bittermann, M. S. (2024). A pattern coloring method using rank-based interactive evolutionary algorithm. In *ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia CAADRIA 2024* (pp. 19-28). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).
- Dudley, J. J., Jacques, J. T., & Kristensson, P. O. (2019). Crowdsourcing interface feature design with Bayesian optimization. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)* (pp. 1-12). ACM.
<https://doi.org/10.1145/3290605.3300482>
- Fernando, R. A. (2014). Space planning and preliminary design using artificial life. In *Proceedings of the 19th International Conference on Computer-Aided Architectural Design Research in Asia CAADRIA 2014* (pp. 657-666). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).
- Kadner, F., Keller, Y., & Rothkopf, C. A. (2021). AdaptiFont: Increasing individuals' reading speed with a generative font model and Bayesian optimization. In *CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1-11). ACM.
 doi:10.1145/3411764.3445140
- Khajah, M. M., Roads, B. D., Lindsey, R. V., Liu, Y.-E., & Mozer, M. C. (2016). Designing engaging games using Bayesian optimization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 5571-5582). ACM.
<https://doi.org/10.1145/2858036.2858253>
- Koyama, Y., Sakamoto, D., & Igarashi, T. (2014). Crowd-powered parameter analysis for visual design exploration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14)* (pp. 65-74). ACM.
<https://doi.org/10.1145/2642918.2647386>
- Takagi, H. (1998). Interactive evolutionary computation: System optimization based on human subjective evaluation. In *IEEE International Conference on Intelligent Engineering Systems* (pp. 17-19). <https://hdl.handle.net/2324/1670060>
- Tanaka, M., & Yamada, S. (2025). Implementation of a Design Exploration Support System in Parametric Design: Optimization through Sensibility Evaluation Using Human-in-the-Loop. In *Summaries of technical papers of Annual Meeting Architectural Institute of Japan (Kyushu)* (pp. 191-192). Architectural Institute of Japan.
- Muley, H. B., Saravana Kumar, G., & Mukerjee, A. (2003). Interactive shape optimization for aesthetics using genetic algorithm. In *EDS PLM User Conference 2003*.