Artistic Reflection Through

Co-Creative Support Tool Curation

by

Daniel Charles Ribaudo

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Computer Science

by

_____

May 2022

APPROVED:

_____

Professor Gillian Smith, Thesis Advisor

_____

Professor Lane T. Harrison, Thesis Reader

Abstract

Creativity Support Tools (CSTs) seek to aid their users in various stages of the creativity process. This thesis explores a Co-Creative Human in the Loop (HITL) CST in conjunction with / as a vehicle for the exploration of a set of image metrics. These image metrics seek to provide a comprehensive analysis of design images, specifically digital quilt designs, by placing them in meaningfully descriptive and distinct positions of the design space implied by these metrics. Through this placement within a comprehensive design space, images may be compared to one another directly through distance measures. The result is a system that allows artists and observers alike to identify meaningful subsets of a larger body of work, for the purposes of arranging, curating, understanding, or appreciating the body of work. The successes and failures of this system are examined through a user study with both qualitative and quantitative results.

Acknowledgements

I would first and foremost like to thank Professor Gillian Smith, the adviser of this thesis. Throughout this research, Gillian has provided excellent insight, ideas, and articles to jumpstart my research of myriad topics. I would also like to thank Professor Lane Harrison for being the reader of this thesis, as well as providing me with excellent research resources and feedback upon request. I would like to thank professor Charlie Roberts for his advice pertaining to the development of Web Applications, especially providing starting points to research further. I am deeply grateful for the support of the Project Code Crafters research group. Throughout the research period, I regularly attended meetings for this group, which both provided feedback to one another and a great deal of emotional support. Finally, I would like to give thanks to my loving partner, Anita Tomasa Jaimes, for celebrating every accomplishment along the way with me.

Table of Contents

List of Figures

I.   Introduction

Creativity Support Tools (CSTs) tools offer "the promise of making more people more creative more of the time" (Schneiderman, 2002). Creativity, for the purposes of this research, is defined as "the development of a novel product that has some value to the individual and to a social group" (Hewett et al., 2005). Thus, a creativity support tool could be software with a collection of functionalities which primarily facilitate some combination of: the brainstorming process as the prelude for a creative artifact, the development of creative artifacts, and the iteration or reflection upon an existing creative artifact or set of artifacts. For clarity, a creative artifact is a cohesive product or concept which embodies creativity, meaning that it is both novel and found valuable by a specific social group.

Within the field of CSTs, a broad range of different processes exist to support a user creatively. In a mixed-initiative co-creative (MI-CC) approach, the creative aspect(s) of the tool arise "via the [shared] interaction of a human initiative and a computational initiative" (Yannakakis et al., 2014). For any tool meant to aid a (human) user, there will be a human initiative. The user's interaction with and responses to the system with a set of goals / purpose in mind comprises this initiative. Rather than yielding static responses, or otherwise providing just the tools that a user would need to pursue their goals within a tool, a computational initiative contributes directly, or via suggestion, to the creative artifact being jointly developed within the context of a CST. This means that the CST itself may generate suggestions / a portion of the creative artifact being negotiated.

This MI-CC approach strongly resembles a Human in the Loop (HITL) approach for computational creativity. Whereas MI-CC is a model for a CST that involves a computational initiative, HITL is a model for machine creativity that involves a human initiative. Specifically, in generative tasks for machine creativity, HITL is "where our interactions and feedbacks with the model play a crucial role in training and/or generation" (Chung, 2021). There is clear symmetry between the models of MI-CC and HITL for their respective domains, and this research considers findings and formalizations of both.

Outlined within this research is the creation of a novel CST which uses an MI-CC approach. The CST is primarily concerned with the task of arranging design images in the context of a physical space, by utilizing a set of supporting, deterministic metrics which can be evaluated upon image data. The purpose of this research is therefore to document image metrics and related CST functionalities that could work in tandem as part of a system that allows artists and observers to better identify meaningful subsets of a larger body of work, for the purposes of arranging, curating, understanding, or appreciating said body of work.

The body of work that this tool is developed around is a collection of digital quilt designs. The decision upon this dataset was based around simplicity and convenience. It was simplistic to evaluate a set of quilt designs because quilt designs tend to be abstract images composed of a color palette and patterns (repeated blocks). This is evident in light of research-based quilt

generative software defining quilts in terms of palettes and blocks (Graves et al., 2021). This type of image stands in contrast to a natural image, which may necessitate object detection within its processing. The convenience of this decision arises from the fact that the adviser for this research already had a varied and diverse collection of digital quilt designs/plans on hand, including means of contacting the associated designers.

Because quilt designs are prominent in this research, this paper also falls into the domain of computational craft. Computational craft is "a hybrid of tangible art and virtual computation" (DuMont & Fields, 2013). A majority of the quilt designs associated with the experiment have been designed with the ability to transfer to a corresponding tangible design. Thus, this dataset could be said to have a tangible craft aspect, which may influence any future work done with this research or the corresponding dataset.

This research explores two primary research questions, involving the individual aspects of, and interplay between, the developed Mixed-Initiative Co-Creativity Creativity Support Tool and accompanying set of image metrics:

1) What constitutes **members of a set of metrics**, which can be calculated **deterministically**, which can adequately describe **aesthetic differences** between images in a way that supports direct **comparison**?
2) What systems can **Co-Creative CSTs** utilize that support the **functionality** of potential members of the aforementioned set of **image/design-based metrics**?

II.   Related Work

Creativity Support Tools

There have been several workshops and discussions surrounding best practices and uses for CSTs. This research drew heavy inspiration from the topics discussed in a 2005 workshop (Shneiderman, 2005). Important within the workshop are the following (all related to CSTs): a listing of the dimensions of a CST taxonomy, research methodologies, focuses, design principles, and case studies.

A Creativity Support Tool Index also has been created, with the intent of standardizing ways in which these tools may be evaluated (Caroll & Latulipe, 2009). The implementation of this CST Index derives from the NASA TLX, which is a more generalized index for evaluating tools. The CST Index, made to specifically evaluate creativity-support-focused tools, hones in on the following categories: exploration, collaboration, engagement, effort/reward, transparency, and expressiveness.

Within the descriptions of the dimensions of the CST index, a large amount of focus is put into ideas around the tool's ability to have its user(s) enter into a flow state. The ideas of a flow state are part of a respected theory by the psychologist Mihaly Csikszentmihayli. Within this theory, flow comes from a deep, but effortless involvement (Csikszentmihalyi, 1997). The research around the CST index argues both that the flow state is tied to creativity of a user and that it corresponds to engagement with a given [software] tool (Caroll & Latulipe, 2009).

In the field of tangible CSTs, a research paper titled "It Is Your Turn: Collaborative Ideation with A Co-Creative Robot through Sketch" (Lin et al., 2020) explores a co-creative sketching process which has users alternate taking turns sketching components of a design. The study examined this process both through an application and through a robotic implementation that could draw on paper. This study found that both tools, though especially the robotic implementation, encouraged users in their exploration of designs and ideation. Though the research contained within this paper may translate to physical implementation (spatial layouts of designs in a physical space / digital designs reflective of quilts that could be created physically), this topic is not explored further within this research.

HITL (Human in the Loop) approaches for machine learning based methods of computational creativity are very similar to that of the MI-CC approaches for CSTs, as both involve partial creative initiative from a human and the computational system they are interacting with. There have been many advances in HITL in recent years, including short films such as Sunspring, jointly authored by humans and computers (Sharp, 2016). This artistic artifact was produced via LSTM-generated scripts, and performed by a human director and actors. Projects like this flip the traditional role of humans and computers (computers serving as a tool for humans), by having humans interpret and act out a set of instructions produced by a machine. This research

remains in a traditional approach, however, having the CST within the study provide novel recommendations to the user, who holds the final say.

Metrics

Design Space Metrics

There has been a lot of research in procedural content generation (PCG) comparing the differences in the expressive range of different generators which create the same class of artifacts. To understand the expressive range of generators, various metrics first must be introduced to define the space to evaluate generators against themselves and one another. In a level generator study done by Smith and Whitehead, level generators are examined via a set of 2D heatmaps. Each axis represents a different metric modeling the design space: linearity and leniency respectively. Parameters of generators are changed, and then multiple levels generated and scored post-generation. The heatmap resulting from binning the metrics plotted against each other is used to represent the tendencies and range of the generator (Smith & Whitehead, 2010).

Keeping with the research in the field of video game-centric PCG, another study was done as a follow up to the aforementioned study. In this more recent study, a far larger range of metrics are proposed, including aesthetic-specific metrics: palette type/variation, color psychology, and textures present (Canossa & Smith, 2015). The act of defining the space to evaluate the expressive range of the content available in terms of level aesthetics mirrors the work in this paper done with the aesthetic components of image data.

Diverging from the domain of level generation, a study done by Mahajan uses metrics to provide feedback to students working to create interactive narratives with branching paths. The metrics within this study are focused on both providing feedback through categorizing the students' work in templatized form and comparing it to similar projects. The students reviewed projects found to be similar to their own project according to the system, for inspiration (Mahajan, 2019). This process of comparing artifacts by defining them in terms of a set of metrics, especially for the purposes of meaningful feedback, is what the study within this paper focuses upon in the context of a CST.

Image Aesthetic Metrics

Some research pertaining to the expressivity of creative artifacts in terms of metrics has directly examined aesthetic values of image data. The Bilvideo research project utilizes, in part, data extracted from MPEG-7 image descriptors as a means of providing image search functionality for videos. In this research, semantics of images and videos are defined and standardized through the MPEG-7 metrics that can be calculated upon 2D image data, allowing semantic queries by providing examples or related sketches (Baştan et al., 2010). There is a clear value

for image metrics that have been formulated in the context of search-based problems in determining similarity between two images. These metrics can be used to augment a metrics-defined aesthetic space.

In terms of aesthetic metrics, another study evaluated the first impressions that users had of web pages. This study calculated a set of metrics on images representing these web pages, to ascertain whether there existed a relationship between the calculated metrics and feedback after a limited viewing period. These metrics, such as symmetry, balance, and color, focused primarily on aesthetics to mirror the aesthetic judgments given within the user study. (Harrison et al., 2015). These metrics were evaluated in a stand-alone manner, whereas this research investigates its body of metrics in the context of a CST designed to support the information they convey about the similarity between images on specific axes.

Similar Methodologies

A CST concerned with creating quilts was created by Project Code Crafters, to study the relationships between quilters and computing technology (Graves et. al., 2021). The research in this paper draws many parallels with this project, especially in the focus topic. Both explore CSTs in the context of computational craft- more specifically in quilting. Though differences arise between generating quilt designs and comparing already existing quilt designs, many similar design principles are shared across these projects' respective implementations.

A study named, "All Through the Night: A Comparison of Two Dollhouses", investigates the differences between physical and corresponding virtual spaces in a craft context. The research, in part, outlines the importance of the context that observers view individual artistic artifacts within (Royer, 2019). This legitimizes the focus of this research being placed on how the individual artistic artifacts within a body of work relate to one another in meaningful ways.

Research that involves curating a subset of images and other creative artifacts also exists, but it differs from the research done in this paper by virtue of relying entirely on Natural Language Processing (NLP). Rather than using the data directly associated with the creative artifacts to arrange them, the algorithms involved use NLP on the metadata (artist, date, theme, topic, etc.) to construct a curation arrangement entirely programmatically. The arrangement must fit a set of pre-defined narrative structures permitted for the system to use. (Wolff et al., 2012)

III.   Creativity Support Tool

The CST developed during the research period focuses on the process of curation by providing an interface for users to place images relative to one another. This CST has an MI-CC approach, as it offers suggestions based on previous user selections. These suggestions are based upon what the system determines to be similar to those already adjacently placed. This similarity is calculated via the metrics discussed in the next section, and they are a vital part of how this CST functions. Conversely, the functionality that the CST implements seeks to elevate the utility of these metrics for a user.

For implementation, the CST takes the form of a Web Application. HTML, JavaScript, and CSS are used to build the Web Application, without the help of any third party website making tools. The JQuery library is also used for additional functionality, mostly for the ease in which it parses Comma Separated Values (CSV) files. The Web Application is hosted through GitHub Pages, though it can be downloaded and subsequently run through a local server to provide offline functionality.

Like the CST proposed by Project Code Crafters, which has its functionality split into distinct tabs (Graves et. al), this CST has its functionalities split between four distinct screens. These screens contain collections of related functionalities around specific paradigms. The decision to break up the tool in this way stems from a proposed set of CST guidelines titled "Casual Creators". One of its proposed design patterns is "Limiting actions to encourage exploration", made easier if "choices are appropriate, limited in number…"(Compton & Mateas, 2015). Thus, this split intends to limit user choices presented at once into appropriate and meaningful clusters of functionality. The four distinct screens are described in detail below: Main Menu, Image Search, Layout Editor, and Save Editor.
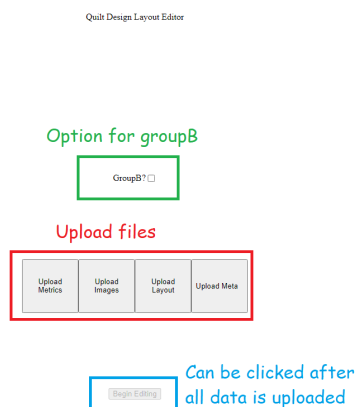
Main Menu

Figure 1 - An annotated screenshot of the "Main Menu"

The "Main Menu" screen (Figure 1) contains a simple set of functionalities. The primary functionality is uploading the relevant files for the CST to use. The files to be uploaded are: metrics, images, layout, and metadata. Pressing each of the corresponding buttons prompts a file system input, and providing a file (or files) in the expected format will turn the button green.

The metrics upload expects a CSV file containing the preprocessed metrics data for each image. The CST does not calculate the metrics itself, instead relying on a pipeline to preprocess the image data before the user interacts with the CST. There are several strengths associated with this approach, the first of these being a quick ability to continue working upon loading the tool. Because the metrics are saved outside the CST as a file, which is imported into the CST, the CST does not need to recalculate the metrics every time it is loaded / accessed. Moreover, because the CST imports the metrics, different metrics may be utilized; neither the specific set of metrics, nor its implementation, is directly tied to the functionality of the CST. Finally, this proprocess and upload approach allowed a custom pipeline involving several different executables and languages to be utilized in the process of calculating the determined set of metrics on the image data.

Both the images and the layout expect image data to be displayed within the application. The images are loaded by the CST dynamically, but the application needs permissions to access the files from the file system. Thus, the CST prompts its user to pick these files to upload, and thus grant permission to use.

Finally, the metadata uploaded is the set of locations that images can be placed upon the layout. This is also in CSV format, containing data about the positions of each location that an image can be placed (referred to as a design node) and about that design node's relationship with nearby nodes. This data is preprocessed for similar reasons to the metrics data, discussed above.

After uploading all of the requested data, all of the upload buttons will appear green. Additionally, the "Begin Editing" button at the bottom of the screen will become enabled (it starts disabled). Upon clicking this button, the user is transitioned to the "Layout Editor" screen and all of the remaining functionalities of the CST are initialized with the data contained in the file selections.
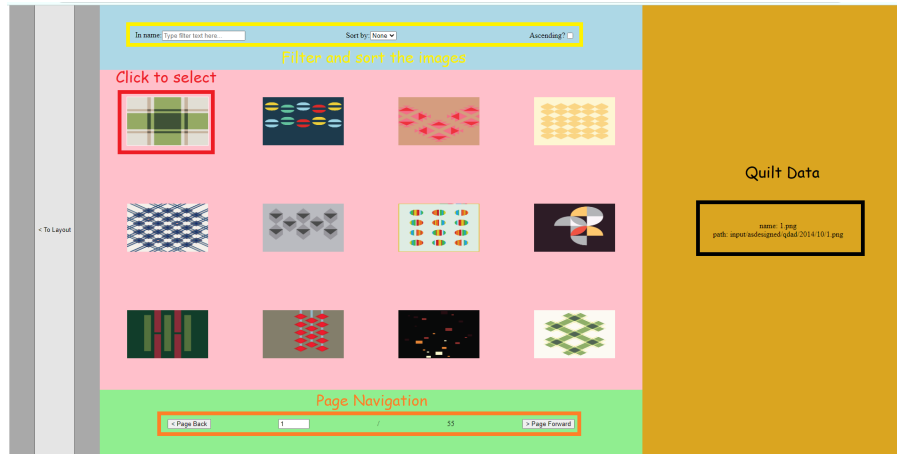
Image Search

Figure 2 - An annotated screenshot of the "Image Search"

Users can browse the images they have uploaded to the CST in the "Image Search" screen (Figure 2). This screen contains a viewable listing of images within its center. To view details of an image, such as name and associated metrics, the user may click the image they would like to investigate. The set of images is broken into pages, which can be navigated through using the user interface (UI) at the bottom of the screen.

The details of a selected image will appear at the right side of the screen. These details will include the filename and path of the image, as well as all of the metric values associated with the image (metrics in alphabetical order).

To better organize the set of images according to the uploaded metrics, and to enable users to find an image more quickly, a set of search tools is at the top of this screen. One tool filters images by whether their name contains an input string, allowing users to potentially find a specific topic or quickly find an image they know the name of. The next tool sorts images by a specific metric, if selected. The sort order, ascending or descending, can also be changed by the user. This functionality serves two purposes: to demonstrate a relationship between the images on a single metric axis, and to provide meaningful examples of each metric to the user. The user can view both extremes of each metric because of the ascending checkbox.
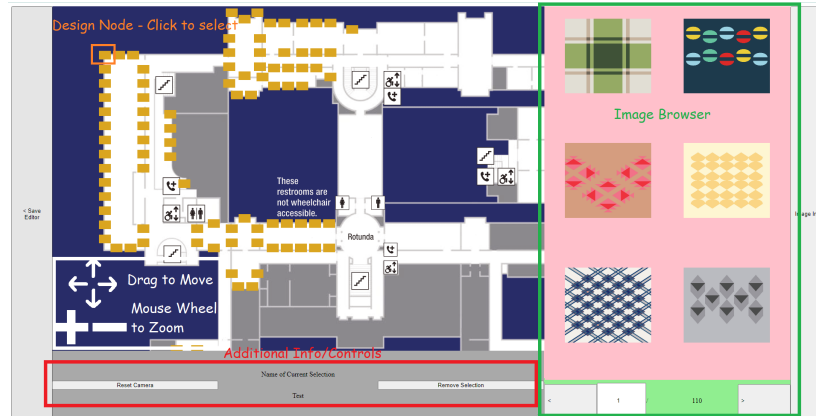
Layout Editor

Figure 3 - An annotated screenshot of the "Layout Editor"

At the side of the "Layout Editor" screen (Figure 3), a miniature version of the "Image Search" exists. While there is no filtering or sorting functionality, the previous filter and sort settings remain for this version of it. Moreover, it displays the last page of images viewed within the "Image Search", adding parity between the two screens.

The main functionality of the "Layout Editor" is the ability to place images within the design nodes on the layout image. When clicking on a design node, a set of recommendations will appear. These recommendations correspond to images that are closest to the nearby design nodes' image selections. This closeness is directly derived from utilizing the distance formula on the metrics of images. Thus, the images with metrics most similar to the nearby selected images will be presented by the CST as recommendations to the user. As a note, recommendations cannot contain already selected images. Without this limitation, it is foreseeable that the same image would be repeatedly recommended, since it would contain identical metrics to a repeated image selection nearby.

Each design node has a relationship with the 20 closest design nodes to it. To calculate closeness, while preprocessing the layout image metadata, the image was first downsampled significantly. After this process, colors were defined as being either passable space or impassable space. The pixel data of the downsampled image, with this consideration of passable space, has the design node locations pathfind to one another via an A* algorithm. The resulting 2D matrix of distances need only be filled on one side of the diagonal, for the diagonal will contain 0s and it is a symmetrical matrix. To ease the data processing, the data is then reflected over the diagonal, and values for each row are sorted in ascending order. Finally, the top N (in this case, 20) nodes with distance > 0 are selected.

While the distance to other nodes is preserved within the metadata file, to avoid overcomplicating the recommendations users see, this research did not take it into consideration. While the algorithm may use a function of the distance between other design nodes to obtain a weighting, this currently uses a simple

$$f(d) = 1$$

, thus, not taking distance into consideration.

Following the generation of recommendations, a user may click on any of the recommendations to select the image for the design node. Alternatively, the user may click on an image within the miniature "Image Search". This may be done if the user has a specific image in mind or otherwise prefers an image over the recommendations presented to them. To remove an image selection from a design node, a user may select the corresponding button at the bottom of the screen with a design node selected.

To view specific design nodes, a user can pan the camera by dragging the layout image. Furthermore, the user can zoom the layout image in and out to view more or less details via the mouse wheel. A button at the bottom of the screen resets the initial view of the layout image. These functionalities may be useful if the layout image is too large to see the design nodes within, or if the user would rather see a segment of the layout image in a particular region of the screen.

Save Editor
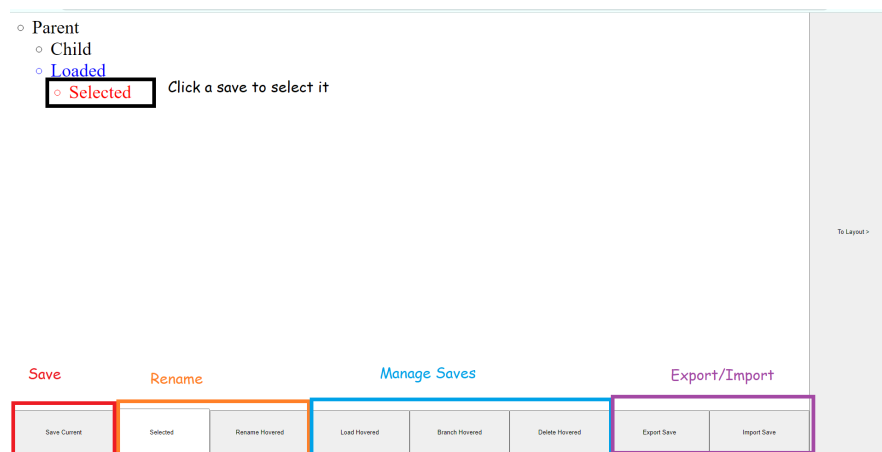
Figure 4 - An annotated screenshot of the Save Editor

The "Save Editor" screen (Figure 4) of the CST allows users to iterate upon their work within the "Layout Editor". The primary function of this screen is to manage save states that the user may jump between and update. Each save state contains the selection data for the design nodes within the layout image.

Rather than a flat listing of saves, the CST presents save states in a tree structure. This tree structure gives further meaning to the relationship between saves. If the user wishes to create a new save state, they must branch off of a currently existing state. This new branch will contain the same data as its parent, initially, and the user may then make changes and save them to this new state. Thus, children of save states share an initial set of selections with their parents.

This relationship may simplify the way users iterate on their layout selection efforts by both mirroring and showing a clear progression of ideas.

To preserve the relationship between parent and child save states, if modifications are saved to a save state with children, a new child is created containing the changes, instead of altering the selections of the parent. Additionally, deleting a parent will also delete the child save states, rather than reassigning them to a new position or relationship. If the root (topmost) of the save states is deleted, a new one will automatically be generated to prevent the user from becoming unable to save data.

To provide more context, the user is able to rename save states. Any save state may be renamed by selecting it, editing its name, then pressing the confirm button adjacent to the text field containing the new name. This allows the user to express their intent with each iteration of a save state and quickly find states they are searching for.

The idea to use save states which have implicit (parent-child relationships) and explicit (names) semantic meanings derives from the aforementioned 2005 workshop based around CSTs. One of the design principles focused upon in the workshop is that supporting exploration is key. This includes a user's ability to "backtrack when unsuccessful" (Resnick et al., 2005). This research applied this notion to save states, and thus implemented a system which would allow users to backtrack through their iterations in a way directly supported by the CST.

IV.   Metrics

This research defines a set of deterministic metrics calculated upon image data. These metrics, in turn, define a design space which a set of images may be mapped to. The metrics calculated upon a body of work can demonstrate an expressive range of a body of related artifacts. Moreover, the CST discussed in the previous section can utilize these metrics to determine similarity between two images. This similarity derives from proximity between images within the representative space. Overall, these metrics are highlighted by the functionalities contained in the CST, but they are also a key part of its functionality.

A total of 15 different metrics were used within this research, divided as follows:
- Metrics derived from the Vizweb Implementation
  - Colorfulness 1
  - Colorfulness 2
  - Diagonal Symmetry
  - Horizontal Symmetry
  - Vertical Symmetry
  - Diagonal Balance
  - Horizontal Balance
  - Vertical Balance
- Metrics derived from MPEG-7 descriptors

- ○ Dominant Colors
- ○ Color Intra-Variance
- ○ Spatial Coherence
- ● Custom (Miscellaneous) Metrics
  - ○ Brightness Average
  - ○ Brightness Standard Deviation
  - ○ Compression
  - ○ Patterns

To describe these metrics during the user study, this research employed a tool named the DeJargonizer[1]. The corresponding paper for this tool suggests "jargon can alienate and exclude many audiences" (Rakedzon et al., 2017). Thus, to improve the understandability in both the user study and this paper, the resulting lay descriptions are provided alongside each metric they correspond to.

Metrics from the VizWeb Library

A study into user impressions of web pages titled "Predicting Users' First Impressions of Website Aesthetics With a Quantification of Perceived Visual Complexity and Colorfulness" examines the relationship between perceived colorfulness and complexity of a webpage and a set of related metrics calculated upon the images of webpages tested. The code for this study is publicly available in the form of a Java library titled Vizweb (Harrison et al., 2015). This project uses the available library as a means of developing a part of the metrics pipeline. Metrics derived from this library are contained within this section.

Colorfulness

Pixel data may be represented in a variety of colorspaces. The PNG format, which nearly all of the images in this research's data utilize, is composed of RGB components (red, green, and blue). There exist other colorspaces, which can be converted to and from RGB, such as HSV (hue, saturation, value).

The Vizweb library provides two different colorfulness metrics. According to the associated research, **Colorfulness 1** is a linear combination of average saturation and the standard deviation of saturation across an image. As described in lay terms, "color can be expressed in terms of how saturated it is. If the colors in the design are highly saturated, a person may initially see it as more colorful. This measure will be higher the more colorful the design is in these terms."

Colorfulness 2 is calculated as the color difference of each pixel within an image against its corresponding gray. (Harrison et al., 2015). In lay terms, "each color has a similar gray value.

---

[1] http://scienceandpublic.com/

The further each color of a design is from its gray value, the more colorful a person is likely to rate it. This metric will be higher with a greater average distance between colors in the design and the most similar gray value."

Neither of these metrics required any modifications apart from an API call to their corresponding function in the context of the metrics pipeline.

Symmetry

Symmetry, is defined in the paper "Symmetry as a Continuous Feature" as: "A 2D object is mirror-symmetric if it is invariant under a reflection about a line" and "A 2D object has rotational-symmetry of order n… if it is invariant under rotation of $2\pi/n$ radians about the center of mass of the object" (Zabrodsky et al., 1992). With all metrics in this paper, however, a symmetry metric must be continuous rather than binary.

In the same paper, a definition for continuous symmetry is proposed (Zabrodsky et al., 1992). While not employed by the Vizweb library, this definition demonstrates that symmetry may be reasonably measured as a continuous feature. The Vizweb library implements a continuous symmetry, because "symmetry and balance significantly correlated with participant's subjective ratings on aesthetics" (Harrison et al., 2015). Within this Vizweb library implementation, symmetry is calculated via reflecting the image over the desired axis and measuring the percentage of pixels that are identical.

From the Vizweb implementation, this paper gets the symmetry-based metrics **Diagonal Symmetry, Horizontal Symmetry,** and **Vertical Symmetry.** As an example of one of the lay descriptions, the description for horizontal symmetry reads, "This is a measure of similarity between the top and bottom half of the design. This metric will tend towards 1 the more similar the top and bottom of the design are."

The Vizweb implementation does not actually contain a definition for diagonal symmetry. Instead, custom code was written to recreate the method that Vizweb uses to calculate the symmetry on its image data and generalized to quadrants. From these quadrants, I compared each type of symmetry (diagonal comparing I-III and II-IV, horizontal comparing I-II and III-IV, and vertical comparing I-IV and II-III). As there are two pairings to check in each relationship, the results are averaged when determining the final value for the metric calculated.

Balance

Balance is a similar metric to symmetry, yet an asymmetrical image may have balance equal to one (the maximum value). Rather than compare invariance over an axis, a definition of busyness is compared over the axis. More specifically, the amount of objects on each side of an image are compared, with an equal amount (and size) of objects found on both sides evaluating

to a balance of 1. This metric, like symmetry, has been found to be highly correlated with subject aesthetic ratings (Harrison et al., 2015).

The Vizweb implementation utilizes a quadtree decomposition of the image to have its balance calculated. A quadtree decomposition of an image breaks the image into four parts recursively until some exit condition is met (in this case, a section contains all of (or nearly) the same color) (Harrison et al., 2015).

From this, the metrics **Diagonal Balance**, **Horizontal Balance**, and **Vertical Balance** arise. To provide another example lay description for horizontal balance, "measures how busy the top and bottom of the design are, then reports how similar these measures are. A part of an image is busier if there are more different areas/colors within that part. This measure approaches 1 if both halves are equally busy."

The Vizweb library does not currently implement the algorithm originally mentioned within the paper, so the previous paper referred to was used as a reference to make a custom implementation. This implementation used the equation:

$$\sum_{i=0}^{o} di * ai$$

Where di is the distance of the ith object from the center and ai is the area of the ith object (Ngo, Teo, & Byrne, 2003). With a quadtree decomposition approach, the leaves are each treated as an object within the image. By the nature of the quadtree, with no further changes, this would always evaluate to the same number of both sides of the image. Therefore, a nonlinearity was added to this equation, adjusting it to:

$$\sum_{i=0}^{o} di * ln(ai + 1)$$

After calculating this data for each quadrant, the quadrant pairings were compared (as in the symmetry section above), and the average between the two comparisons taken as the final metric for balance.

Metrics from MPEG-7 Descriptors

According to "Bilvideo-7: an MPEG-7- compatible video indexing and retrieval system", the MPEG-7 standard is "designed to describe the content of multimedia… MPEG-7 offers a comprehensive set of audiovisual description tools." (Baştan et al., 2010). From the MPEG-7 features, a set of low level image descriptors can be obtained. With a simple Python program, these features were extracted from the Dominant Color Descriptor (DCD) output of the MPEG-7 standard.

A study titled "User's Web Page Aesthetics Opinion: A Matter of Low-Level Image Descriptors Based on MPEG-7" utilized these extracted metrics in a similar way to the colorfulness, symmetry, and balance metrics. This paper thus examined subjective user aesthetic feedback and examined its correlation with the low-level extracted metrics. These extracted descriptors could demonstrate "high correlation between low-level parameters and the users' evaluation" (Uribe, Alvarez,& Menendez, 2017) . Thus, like other studies on the impressions of web pages, these extracted descriptors, being used as metrics in the context of this research, provide meaningful dimensions that relate to the aesthetic judgments of real observers.

The **Dominant Colors** metric is indicative of how many colors could be used to represent the image. Its lay description is as follows, "how many important colors appear in the design. The higher this number, the more different colors within the design. This value should become close to 0 when the design is 1 solid color." Alongside dominant colors is the **Color Intra-variance** metric. This metric, also extracted from the DCD, expresses how much variability there is in the shades of different colors within the image. Put simply, "This value increases the more small color differences there are in a design. This number will be 0 if there are no different shades of colors within the design." The last metric within this set is **Spatial Coherence**. Its lay description is, "This measure goes up as more colors in the design stay connected or in the same area. The number will be close to 1 if all of the colors within the image stay connected to themselves."

Miscellaneous Metrics

Metrics not derived from a library or predefined functionality are discussed in this subsection. All of these metrics have been implemented via Python and supporting libraries.

Brightness

In the HSV colorspace, the saturation level provided a meaningful metric in the Vizweb library. There exists a plethora of other colorspaces, aside from HSV and RGB. One such space is the CIELAB colorspace. This colorspace has been designed with visual uniformity in mind, meaning that distances between colors are linearly tied to perceptual differences to observers (Connolly & Fliess, 1997). Metrics derived from a perceptually uniform space have an advantage in that the difference between them corresponds to a difference in experiencing two images.

The metrics **Brightness Average** and **Brightness Standard Deviation** explore the lightness/L* component of the CIELAB colorspace. Both of these are calculated by converting an input image to the CIELAB colorspace, and then finding the average brightness and accompanying standard deviation across all of the pixels within the image. The lay descriptions for both, respectively, are, "color in a design can be seen at different brightness. This measure represents how bright, on average, the design appears. At 1, a person would be seeing an all white image. At 0, a person would be seeing an all black image." and "a support measure to the

brightness mean. This represents how much the brightness changes in the design. If the design were to be a single, solid color, then this would be 0."

Compression

In a study analyzing metrics to judge video game levels created through PCG, one of the studies used compression distance between levels to analyze similarity and differences of the expressive range of their particular generator (Horn et al., 2014). This approach leverages the domain of information theory, as levels with similar features or patterns will have a smaller compression. Thus, level generators that have a similar output should generate artifacts which have low compression distances (and therefore are smaller than the sum of their compressed individual parts by a significant margin when compressed together). Compression, as a whole, creates representations of artifacts which are smaller based upon the amount of entropy in the artifact.

There already exist compression algorithms for images. The JPEG format is a lossy format which compresses image data to save significant space. To measure an aspect of entropy within the image, images are compressed down via a JPEG algorithm (to a specific quality, in this case, 50) and compared to their original size. The **Compression** metric has an inverse relationship with the ratio between the compressed image and the original image. Thus, "images with few colors and well defined patterns are able to be compressed to save memory. This metric shows how much smaller the design can be in memory, with a higher value meaning it can be much smaller."

Patterns

Quilts naturally contain many patterns. Quilts are defined in the CST produced by Project Code Crafters in terms of a set of repeated blocks (Graves et al., 2021). Therefore, an appropriate aesthetic metric would be related to the portion of a quilt covered in detectable **Patterns**. The lay description for this metric is, "Measures the amount of repeated patterns within the design. If the design does not have repeated images, then this value will be 0."

To implement this pattern metric in a deterministic fashion, first a set of SIFT keypoints are calculated on an image. The "Scale Invariant Feature Transform (SIFT)... transforms image data into scale-invariant coordinates relative to local features" (Lowe, 2004). The keypoints generated from this approach can be compared to one another by their descriptors, and the algorithm for patterns utilizes this to separate SIFT points into classes. These classes contain points which describe identical looking regions. These classes of keypoints are then compared against one another to determine if any relationships exist (e.g. whether points in class A have a corresponding point in class B with the same relative offset as another pairing). When all class relationships of identical-descriptor SIFT points have been found, similar groupings of points get a bounding box drawn around them, indicating the existence of a pattern. The pattern metric is derived by taking the total area of the bounding boxes and dividing it by the area of the image.

V.   User Study

There is notable interplay between the CST and metrics implemented within this research. CST functionalities serve the purpose of allowing exploration of the image set in the context of the metrics. Symmetrically, the metrics provide a meaningful way to compare multiple images in the context of aesthetics to provide suggestions within the MI-CC functionality of the CST. A qualitative user study outlined in this section seeks to examine how effectively these two systems interact with and support one another.

Recruiting

Recruiting for this study utilized a convenience sample. The set of participants that arose from this convenience sample were all college-age or slightly older, many of whom attend or have attended Worcester Polytechnic Institute. All recruited participants were friends or acquaintances of the researcher for this thesis.

Information Provided

The consent form provided to the participants required signing before a participant could progress further into the user study. This consent form supplies initial information about the user study, including potential rewards, risks, and the participant's rights while involved in the study. These rights include, but are not limited to:

1) Participation in the research is voluntary.
2) Participants may end their participation at any time.
3) Participants need not answer every question in an interview or survey.

There are also three packets of information that the participants needed to read through during the user study. The first packet contains a reference to the controls of the CST. This set of controls contains a set of annotated screenshots of the various screens within the CST and a written description on how to use each primary functionality. The second packet contains a guide to the metrics used within the study and their corresponding lay descriptions. The last packet contains instructions for the participant to follow throughout the user study. These instructions mix a tour of the main functionalities of the CST with freedom afforded to the participant to use the tool in a manner that suits their interest.

Group Strategy

The researcher would assign participants to either "Group A" or "Group B". These different groups correspond to different initial functionality within the CST. Group A starts with access

metrics and metric-related functionality enabled, then transitions to a CST state which cannot access the metrics or metric-related functionality. Group B is the opposite, starting in a CST state without metrics and transitioning to a CST state with metrics.

This split was done to provide a baseline for the CST and to reduce order bias stemming from how participants learned the CST. Each group would have a very different experience with the CST. Group A participants would enjoy the benefits of metrics immediately, yet they also have more to take in at once. Group B participants would instead have less to learn and do immediately, but would not have access to the metrics, the core part of several functionalities in the CST.

Data Collected

This user study collects two types of data. The first data collected is an exit survey, containing questions about UI, enjoyment/satisfaction, and the metrics. This exit survey was implemented in google forms, and each group has the same questions (albeit in different order, to match their respective experience).

The second type of data is the save state information created during the participants' interaction with the CST. Before the CST is reloaded or refreshed, the researcher can export the data contained with the "Save Editor" for analysis.

CST Modification

Two modifications were made to the CST to accommodate for this user study. First, a GroupB checkbox toggle was added to the "Main Menu" screen. This exists for the researcher to select if the participant will be entering a part in the study where they are not allowed to access metrics or metric-related functionalities. The main change this checkbox makes is altering the return of a function within the CST which gets a listing of the metrics from the CST. Having this checkbox toggled on causes this function to return early with an empty list.

The second modification involves import and export functionality. This functionality can be found on the "Save Editor" screen. When the export button is pressed, the save state tree is turned into a CSV file and saved locally. This same file can also be re-imported to the CST, so that the researcher could manually investigate the save after the study had concluded.

Procedure

1) The participant must read and sign the consent form before participating in the user study.

2) The researcher then provides the participant with background about the research as it relates to the user study.
3) The participant receives the three packets of information and is given the chance to read through them all.
4) The researcher assigns a group to the participant (either Group A or Group B).
5) The researcher fills out the input information in the "Main Menu" screen accordingly, to prepare the session.
6) The participant then is free to follow the instructions on their own. As the study progressed, the researcher was allowed to answer questions about the UI and clear up confusions pertaining to the participant's current tasks.
7) Upon the completion of the instructions, the researcher would export the save state data from the session and refresh the page. The researcher would again fill out the "Main Menu", this time picking the opposite group.
8) The participant then follows the instructions again, with the altered set of CST functionality.
9) Upon the second completion of the instructions, the save state data is once again exported by the researcher.
10) Finally, the participant is directed to an exit survey containing essential questions about their experience with the CST and the corresponding metrics.

VI. Results

Are the branching saves of the "Save Editor" intuitive?
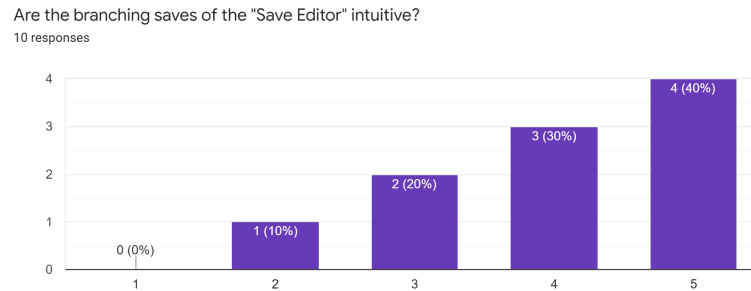10 responses

Figure 5 - Results for "Are the branching saves of the 'Save Editor' intuitive?": This figure shows the distribution of Likert scale results. This was one of two UI-centric questions which had any answers score below a 3 on the Likert scale.

Was loading a previous state within the "Save Editor" intuitive?
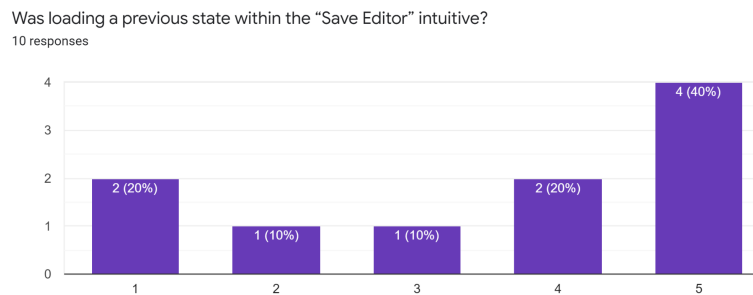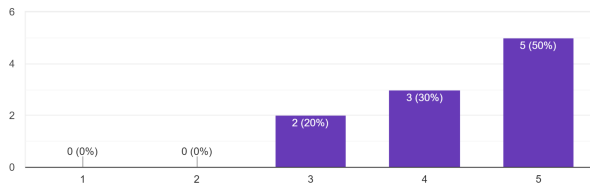10 responses

Figure 6 - Results for "Was loading a previous state within the 'Save Editor' intuitive?": This figure shows the distribution of Likert scale results. This was the other one of two UI-centric questions which had a score below 3.

Though the idea of relating save states in the tool via parent-child relationships may be a novel idea within this context, that has not prevented it from confusing users. As shown in Figure 5 and Figure 6, functionality associated with the "Save Editor" is the only functionality in the CST that users had significant trouble understanding on an intuitive level. Though the potential of semantic links between user saves may allow for easier backtracking and organization of iteration/intention, this implementation was clearly the most confusing of the novel functionalities implemented within the CST.

With image metrics enabled, how much do you agree with the following statement, "I feel creatively supported by the application"
10 responses

With image metrics disabled, how much do you agree with the following statement, "I feel creatively supported by the application"
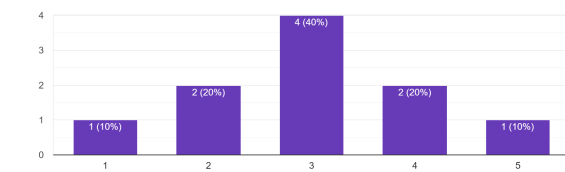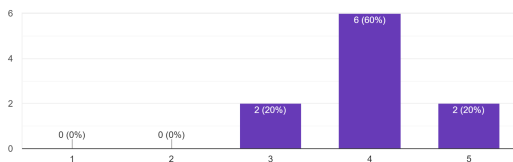10 responses

Figure 7 - Results for "How much do you agree with the following statement, 'I feel creatively supported by the application'": On the left is the distribution for enabled metrics, and on the right is the distribution for disabled metrics.

The question in Figure 7, ""How much do you agree with the following statement, 'I feel creatively supported by the application'", seeks to understand how users perceive the overall effect of metrics on the CST in its capabilities for being supportive creatively. With metrics enabled, the tool scores very highly in the Likert scale distribution, but with metrics disabled, the tool has a symmetric distribution. This relationship indicates that the CST and its functionalities are greatly augmented and improved through their relationship with the image metrics. Clearly, the metrics employed fit well into the context and functionalities of the CST.

With image metrics enabled, how would you rate your engagement with the application?
10 responses

With image metrics disabled, how would you rate your engagement with the application?
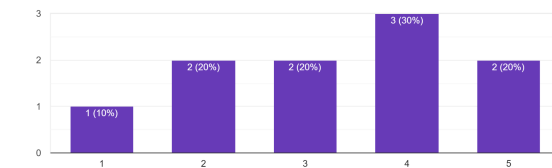10 responses

Figure 8 - Results for "How would you rate your engagement with the application?": On the left is the distribution for enabled metrics, and on the right is the distribution for disabled metrics.

Another way of evaluating the effectiveness of the CST is to gauge engagement with the tool. This engagement may be indicative of a flow state, which is a key component in stimulating the creativity of users. Shown in Figure 8, metrics enabled again shows a distribution tending towards high levels of engagement within the Likert scale. Metrics disabled, on the other hand, demonstrates a nearly flat distribution with many more 1,2, and 3 responses. This may indicate that without metrics, the CST cannot adequately (or consistently) capture a user's attention or allow a user to enter a flow-like state. Overall, these results again speak to the metrics elevating the functionalities within the CST.
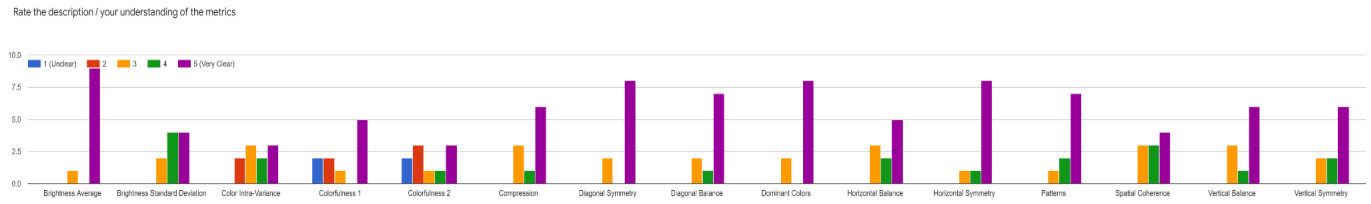
Rate the description / your understanding of the metrics

Figure 9 - Results for "Rate the description / your understanding of the metrics": With the exceptions of Brightness Standard Deviation Color Intra-Variance, Colorfulness 1, and Colorfulness 2, metrics score a 3 or more, with a majority of responses being 5 (Very clear).

Figure 9 shows that the selection of metrics used was easy for a participant to understand. This understandability is a key facet of the set of metrics. Without an understanding of each metric, users would not be able to draw conclusions from the evaluations of each metric on the set of images, nor would they be able to effectively identify relationships between images through each metric. Users generally seemed to understand every metric, with the two least understandable being Color Intra-Variance and Colorfulness 2. In this dimension of the metrics evaluation, this metric set appears to excel in its effectiveness relative to the quilt designs data set.
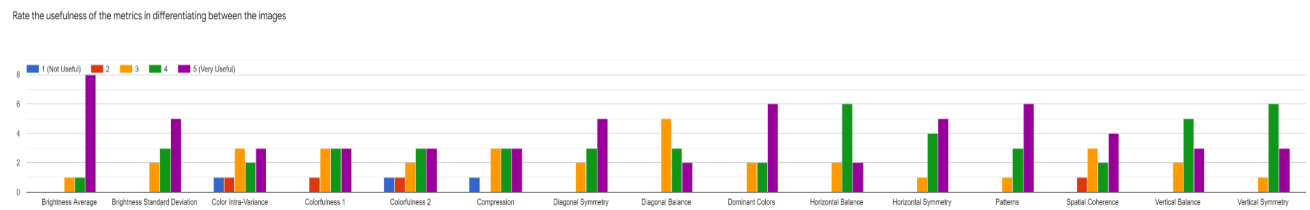
Rate the usefulness of the metrics in differentiating between the images

Figure 10 - Results for "Rate the usefulness of the metrics in differentiating between the images": Many metrics have distributions which contain a high amount of fours and fives (Very useful).

The ability for metrics to differentiate between images contributes to their usability in the context of the CST and contributes to the understanding of the metrics by the user. As previously stated, an understanding of the metrics is key to drawing conclusions about a body of work relative to the metrics. Clear differentiation between images is related to user understanding because this differentiation is able to provide distinct examples for the user to work with. Regardless of a user's understanding of the metric, if it were unable to differentiate images on its own, then it would not be useful to examine in the design space. In the set of metrics associated with this research, according to Figure 10, users generally rated their ability to differentiate between images between medium and high, with all metrics having a median of greater than 3. As in the previous result, this finding also speaks to the effectiveness of the metrics set.
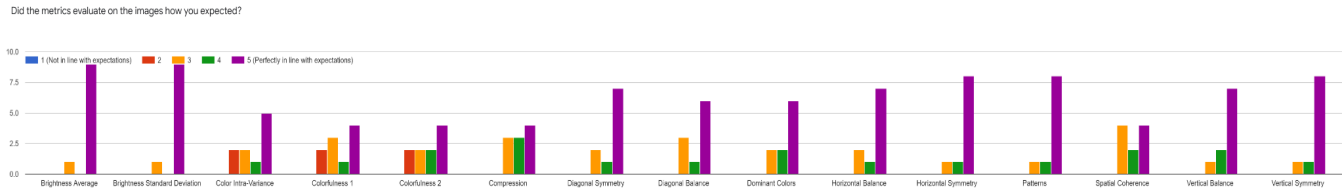
Did the metrics evaluate on the images how you expected?

■ 1 (Not in line with expectations)　■ 2　■ 3　■ 4　■ 5 (Perfectly in line with expectations)

Figure 11 - Results for "Did the metrics evaluate on the images how you expected?": Every metric has a distribution with a mode of 5 (Perfectly in line with expectations) except Spatial Coherence.

As a last component of both understanding and effectiveness, the exit survey asked whether the users felt that the metrics evaluated consistently with their expectations. The ability for metrics to evaluate according to a user's mental model of the metric again speaks to the understanding of the metric. Moreover, if there were a significant error in an implementation of a metric, then one could expect these scores to be very low. While the user may understand a metric, this understanding may allow them to flag images or groups of images as evaluated incorrectly according to their knowledge of the function and purpose of the metric. Figure 11 demonstrates that no such issues were perceived by participants in this user study, as the mode for each of these distributions is 5 (Perfectly in line with expectations). This is evidence for the strength of this selection of a set of metrics.

Number of Leaves	Number of Nodes	Average Fill	Average Branches	Depth
26.2	27.8	29.50666667	0.7364765694	2.5

Figure 12 - Average Values for Exported Save Data with Metrics Disabled

Number of Leaves	Number of Nodes	Average Fill	Average Branches	Depth
2.8	4	41.28428571	0.5032900433	2.2

Figure 13 - Average Values for Exported Save Data with Metrics Enabled

Figure 12 and 13 show the average values for exported save data with metrics disabled and enabled, respectively. There are an extremely high number of leaves and nodes within the metrics disabled table compared to the metrics enabled table, yet this seems to disagree with the survey. The survey suggests that the CST with metrics enabled is more engaging to users. To investigate, the saves associated with metrics disabled were imported back into the tool and manually evaluated. It appears that some of the metrics disabled saves contain an extraordinary number of save states which may be the product of users trying to break or otherwise discover bugs within the CST. Therefore, the only values worth comparing are the average fills of each save, as this number is implicitly normalized to be between 0 and 100 for each exported save. Therefore, any such degenerate save state should not significantly impact the resulting average. The average fill value comparison agrees with the survey: users clearly put more effort into filling iterations of their layout with image selections with metrics enabled; thus, metrics enabled remains more engaging.

The first research question of this paper was: What constitutes members of a set of metrics, which can be calculated deterministically, which can adequately describe aesthetic differences between images in a way that supports direct comparison? The metrics used in this research, all of which are calculated deterministically, evaluated as understandable, useful, and predictable by the user study. Moreover, functionalities of the CST utilizing these metrics were able to better engage and support users within the study. Because the main functionality utilizing metrics within the CST was the recommendations presented to the user, this research concludes that these metrics are also well suited to support comparison between images.

The second research question of this paper was: What systems can Co-Creative CSTs utilize that support the functionality of potential members of the aforementioned set of image/design-based metrics? The CST in this research explored the following primary functionalities: a save state tree that supports semantic relationships between saved iterations and backtracking, the ability to sort images by their metrics values, and the ability to recommend images based on metrics comparisons. The save state tree, though novel, appears to have confused users more than any other functionality within this application. Conversely, both sorting by metric values and recommendations derived from nearby images utilize the set of metrics within this research. With metrics disabled, the CST evaluates poorly in its ability to support and engage its users. Enabling metrics reintroduces these two functionalities and leads to a much more favorable evaluation. It can be concluded from this that these two functionalities are both important to the users and key in their interaction with the designated set of metrics.

VII. Discussion and Future Work

Throughout this research, new problems and suggestions arose. These topics could not receive due attention because they were outside the focus or scope of the research. This section outlines some of these topics encountered and suggests how future work might address them.

There exists multiple ways to implement various metrics or components of the algorithms associated with them. Rather than running a separate study or statistical analysis to determine the most effective implementation of a metric, qualitative decision making was used. First, 2D heatmaps demonstrating the overlap between each metric pairing within the quilting dataset was used to determine whether different implementations had more or less expressive range relative to the already established metrics. Additionally, a set of example images was generated alongside each metric implementation, showing the following examples relative to each metric: an image which evaluated the highest in terms of the metric, an image which evaluated the lowest in terms of the metric, an image which evaluated closest to the mean of the dataset, and an image that represented the median of the dataset. These examples were utilized to determine whether the metrics were evaluating upon images consistently and as expected.

This research could have incorporated more detailed and numerical efforts when altering or adding metric implementations. Alternatively, studies could be run where users evaluate the different implementations of a metric. This deliberation on implementation specifics was not the focus of the paper, however, and it is left to future work.

The tools utilized to obtain metrics in this thesis were all website specific. Each one of these tools was taken from research evaluating user impressions on a glimpse of a website to subjectively evaluate aesthetically. A clear example of this domain mismatch is the quadtree decomposition and formulas surrounding it. The quadtree decomposition seeks to isolate specific objects on a webpage. Quilt designs are abstract, however, and thus there rarely are objects at all to pick out. Research into alternative implementations that either are more general or more suited to the dataset investigated in this paper may provide more accurate or otherwise expressive metrics evaluations.

The metrics are normalized to be in the range [0,1]. If the metrics were not normalized, then the metrics with the highest range would be weighted the highest in comparisons. To avoid this, normalization occurs to equalize the weight of each metric when considering image similarity. This normalization process, however, does not take into consideration significant outliers, which may compress the common range of the metric. Future work may look to address this problem to avoid losses in the expressivity of metrics.

A large concern in the implementation of the CST was the viewing aspect ratios of the images within the quilt dataset. Due to the implementation specifics, images displayed always stretched to fit their container on both the horizontal and vertical axes. The aspect ratio of these containers, however, frequently did not match the aspect ratio of the original image, distorting

user perception of the original image. This is especially confounding in the face of aesthetics based metrics which entirely depend on perceived qualities of the image. If future work decides to continue developing the CST, this is an important alteration to consider within the tool's display.

It became clear over the course of the user study, that users would have liked to see alternative or added functionality to the CST. An example of alternative functionality would be randomizing images presented as recommendations if metrics were not enabled, or nearby design nodes did not have selections to compare to. Another popular remark was to add the ability to click and drag images from the miniature "Image Search" directly into a design node. Of course, many functionalities can be adjusted, changed, deleted, or added in future work on the CST.

The data set of quilt designs may require more attention as well. Throughout the research, repeated images in the dataset were discovered. These repeat images could cause issues with the recommendation system, as repeated images should contain identical metrics. Repeated images will therefore dominate the recommendations tab if its duplicate is nearby. Though removed as they were found in the dataset, no formal algorithm or tool was used to detect these repeeated images. Over the course of the research, several non-digital images were also found within the dataset. These too were pruned, though still may be present for similar reasons that repeat images may still exist within the dataset.

Apart from metrics calculated upon the image data of images, image metadata may also be taken into consideration in future metrics. The focus of this research was firmly on metrics concerned with the image. There exists literature which involves utilizing Natural Language Processing on image metadata to arrange or assist in arranging a curation or subset of the designs (Wolff & Collins, 2012). Examples of this metadata could be: the date which the image was created, the inspiration of the design, the designer who created the image, and where the image was created. Otherwise, simply annotating the collection of quilt designs with meaningful names or tags would allow users to more freely search for names and concepts that they may want to view.

A follow up user study may be possible without major alterations that utilizes the Creativity Support Tool Index. These results would standardize the evaluation of this CST. An alternative change to the user study could be recruiting designers as participants and using their own designs during the user study. The relationship that a designer would have with the tool could be foreseeably different. Whilst curating their own work, a designer may become inspired by previous work or a topic/theme not addressed within the body of their work. Additionally, a designer would have more intimate connections with the work, and may be inclined to spend more time manipulating the tool.

Similar CSTs and functionalities in a physical domain may also be possible to explore, as previously mentioned. This research could be a potential starting point for this work, as the dataset investigated is transferable to tangible quilts. Using real versions of these quilt design research may be done involving the evaluation of metrics in a physical environment. A user

study could also be done to evaluate metrics with quilts that are able to be grasped and manipulated by participants.

VIII.   Conclusion

Within this thesis, a novel Creativity Support Tool which utilizes a Mixed-Initiative Co-Creative approach was implemented and evaluated. The primary purpose of this CST was to facilitate the placement of images within the same body of work. This placement of images was relative to other images in a predefined space. The act of placing images in this manner, followed by iterating upon such selections, had the goal of helping the user identify meaningful subsets of a larger body of work for the purposes of arranging, curating, understanding, and appreciating said body of work. Alongside this CST, a set of deterministic metrics which could be calculated upon traditional image data was formulated and implemented. The dataset that these metrics were applied to was a collection of digital quilt designs. The metrics enable several functionalities within the CST, including comprising the main criterion for recommendations of image placements. This paper sought to explore the relationship between the CST and the set of metrics by focusing on implementing a set of novel functionalities which highlight and/or utilize the metrics within the CST.

To evaluate the components of this research, the CST and set of metrics, as well as the relationship therein, a user study was conducted. Within this user study, participants were guided through the main set of functionalities provided by the CST with and without the associated metrics. The data from this user study contained both the save states of the work done within the study and the results of an exit survey.

Participants had very positive feedback regarding the set of metrics within this body of research. This set of metrics was found, through survey data, to be understandable to general participants, useful in comparing images, and consistent in its evaluation upon the quilt design dataset. Moreover, the functionalities within the CST which utilized the metrics demonstrably caused users to feel both more supported by the tool and more engaged in the tool. The additional novel functionality which did not utilize the metrics, the save state tree, was found generally confusing in its current realization. This work has contributed to a better understanding of both CST functionalities in the MI-CC domain and image metrics that may be reasonably applied to quilt designs and other image data which is either: low level (not containing specific objects) or related to computational craft (in that the design it represents may be physically realized).

IX.   References

Baştan, M., Cam, H., Güdükbay, U., & Ulusoy, O. (2010). Bilvideo-7: an MPEG-7-compatible video indexing and retrieval system. IEEE MultiMedia, 17(3), 62-73.

Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008, April). The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 423-432).

Canossa, A., & Smith, G. (2015). Towards a Procedural Evaluation Technique: Metrics for Level Design. FDG.

Chung, N. C. (2021). Human in the Loop for Machine Creativity. arXiv preprint arXiv:2110.03569.

Compton, K., & Mateas, M. (2015, June). Casual Creators. In *ICCC* (pp. 228-235).

Connolly, C., & Fleiss, T. (1997). A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. IEEE transactions on image processing, 6(7), 1046-1048.

DuMont, M., & Fields, D. A. (2013). Hybrid shmybrid: Using collaborative structure to understand the relationship between virtual and tangible elements of a computational craft.

Carroll, E. A., & Latulipe, C. (2009). The creativity support index. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 4009-4014).

Csikszentmihalyi, M. (1997). Flow and the psychology of discovery and invention. *HarperPerennial, New York*, *39*.

Harrison, L., Reinecke, K., & Chang, R. (2015, April). Infographic aesthetics: Designing for the first impression. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1187-1190).

Hewett, T., Czerwinski, M., Terry, M., Nunamaker, J., Candy, L., Kules, B., & Sylvan, E. (2005). Creativity support tool evaluation methods and metrics. *Creativity Support Tools*, 10-24.

Horn, B., Dahlskog, S., Shaker, N., Smith, G., & Togelius, J. (2014). A comparative evaluation of procedural level generators in the mario ai framework. In *Foundations of Digital Games 2014, Ft. Lauderdale, Florida, USA* (2014) (pp. 1-8). Society for the Advancement of the Science of Digital Games.

Graves, J., Royer, K., Smith, G., & Sullivan, A. (2021, June). Procedural Patchwork: Community-Focused Generative Design for Quilting. In *Creativity and Cognition* (pp. 1-3).

Lin, Y., Guo, J., Chen, Y., Yao, C., & Ying, F. (2020, April). It is your turn: Collaborative ideation with a co-creative robot through sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Mahajan, S. (2019, October). Toward Automated Critique for Student-Created Interactive Narrative Projects. In *Proceedings of the... AAAI Conference on Artificial Intelligence*.

Ngo, D. C. L., Teo, L. S., & Byrne, J. G. (2003). Modelling interface aesthetics. Information Sciences, 152, 25-46.

Rakedzon, T., Segev, E., Chapnik, N., Yosef, R., & Baram-Tsabari, A. (2017). Automatic jargon identifier for scientists engaging with the public and science communication educators. PloS one, 12(8), e0181742.

Resnick, M., Myers, B., Nakakoji, K., Shneiderman, B., Pausch, R., Selker, T., & Eisenberg, M. (2005). Design principles for tools to support creative thinking.

Reymond, C., Pelowski, M., Opwis, K., Takala, T., & Mekler, E. D. (2020). Aesthetic Evaluation of Digitally Reproduced Art Images. Frontiers in psychology, 11, 3512.

Royer, K. (2019). *All Through the Night: A Comparison of Two Dollhouses* (Doctoral dissertation, WORCESTER POLYTECHNIC INSTITUTE).

Sharp, O. (2016). Sunspring. Premiered in *Ars Technica*.

Shneiderman, B. (2002). Creativity support tools. Communications of the ACM, 45(10), 116-120

Smith, G., & Whitehead, J. (2010, June). Analyzing the expressive range of a level generator. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games* (pp. 1-7).

Uribe, S., Álvarez, F., & Menéndez, J. M. (2017). User's web page aesthetics opinion: a matter of low-level image descriptors based on MPEG-7. ACM Transactions on the Web (TWEB), 11(1), 1-25.

Wolff, A., Mulholland, P., & Collins, T. (2012, June). Storyspace: a story-driven approach for creating museum narratives. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 89-98).

Yannakakis, G.N., Liapis, A., & Alexopoulos, C. (2014). Mixed-initiative co-creativity. FDG.

Zabrodsky, H., Peleg, S., & Avnir, D. (1992). Continuous symmetry measures. *Journal of the American Chemical Society*, *114*(20), 7843-7851.