



Aesthetic Measure of Architectural Photography utilizing Computer Vision

Parts-from-Wholes

VICTOR SARDENBERG¹, MIRCO BECKER². ¹²Leibniz Universität Hannover, Germany ¹0000-0001-6089-0189, ²0000-0001-8488-2108



Fig. 1. Heydar Aliyev Centre, Zaha Hadid Architects, and Diamonds House, Gilles Retsin.



Fig. 2. Diagrams of scaled parts.

PART 1: PROPOSITION

ABSTRACT.

The existing methods for solution space navigation require numerical values to score solutions. The authors introduce a method of quantitative aesthetic evaluation utilizing Computer Vision (CV) as a novel criterion to navigate solution spaces. Therefore, aesthetics can complement structural, environmental, and other quantitative criteria.

The work stands in the extended history of quantifying the visual aesthetic experience. Some precedents are: Birkhoff [1933] and Max Bense [1965] built an approach with experiments to empirically support a measure, whereas Birkin [2010], Ostwald, and Vaughan [2016] devised the first computational methods for evaluating vector graphics. Our research automates and accelerates aesthetic quantification by utilizing CV to extract computable datasets from images. We are especially keen on architectural images as a shorthand to assign an aesthetic value to design, aiming to navigate the solution space or design space in architecture [Woodbury, 2005]

This work devises a method for rearranging parts in architectural images focusing on formal aspects, in opposition to semantic

segmentation where objects unrelated to architectural design (cars, persons, sky...) are quantified to score photos [Verma and Jana and Ramamritham 2018]. It uses Maximally Stable Extremal Regions (MSER) [Matas 2004] to recognize architectural parts because it is superior to similar methods, such as SimpleBlobDetector, in this task.

Our method disassembles the parts in a diagram of scaled parts (Fig. 2) to analyze them in isolation, and a diagram of connectivity graph (Fig. 3), to evaluate relationships. These diagrams are examined to compare photos of buildings, cars, and trees to assess the applicability of such a method to a range of objects. Parts and connections are thus quantified, and these values are inputted in a refined version of Birkhoff's formula to calculate an aesthetic score for each image for navigating the solution space.

			Connection edge length average
Aesthetic Measure =	Order Complexity	=	x Number of connections
			Number of parts
			x \sqrt{N} Number of pixels

Finally, it tests the method to draw comparisons between the discrete and continuous paradigms in the contemporary discourse of architecture, comparing Zaha Hadid Architects` Heydar Aliyev Centre and Gilles Retsin's Diamonds House (Fig. 1) to argue that there is a difference between the aesthetic effects of continuous and discrete designs, besides their distinction in tectonic logic. The method proved to be an efficient procedure for comparatively quantifying the aesthetic judgment of architectural images, enabling designers to incorporate aesthetics as a complementary criterion for solution space navigation in computational design.





Fig. 3. Diagrams of connectivity graph.

KEYWORDS

Quantitative Aesthetics, Aesthetic Measure, Computational Aesthetics, Parts-to-whole Relationship

1. INTRODUCTION

This paper is part of a research agenda on computational aesthetics. The hypothesis that is argued is that aesthetics is a powerful tool for comprehending contemporary reality. A new formal repertoire arose with ever-evolving digital design tools, innovative materials, and digital fabrication. After solving the prohibitive issues of producing complex forms, a disciplinary question came to all architects: How to evaluate these new sets of architectural forms?

With the increasing popularity of algorithms for form generation, the question of evaluating form became associated with efficiently navigating the solution space. Architectural forms were traditionally considered through proportions, grids, and typology. However, these methods operate in two dimensions and cannot be computed to feed a solution space navigation algorithm with numbers. Using Computer Vision to extract relevant computable datasets from images, our proposed method overcomes these problems.

Designers utilize algorithms for analyzing forms from various criteria: environmental, acoustics, lighting, cost optimization, and structural behavior. Therefore, recent algorithmic projects are rarely discussed in terms of aesthetics. We are interested in creating a computational aesthetics framework to let the qualitative discourse on aesthetics leak into the computational design and thus complement performance-based design evaluation.

1.1. Quantitative Aesthetics

There is a tradition of studies on how to quantify the aesthetic experience. Gustav Fechner (1801 – 1887) proposed bottom-up aesthetics in 1860, which focused on elemental perceptual features rather than confronting philosophical concepts [Fechner 1948]. He proposed a unified theory of mind and matter that correlates psychological experience with physical stimulus.

George David Birkhoff (1884 – 1944) proposed a numerical model for "aesthetic measure" correlating order and complexity [Birkhoff 1933], following this formula:

His formula was taken up by Max Bense (1910 – 90), which proposed the concept of Information Aesthetics [Bense 1965]. Considering that objects have objective aesthetic states, Bense argued for an objective evaluation of works. Sigfried Maser, who studied under Bense and submitted his doctoral thesis "Numerische Ästhetik" (Numerical Aesthetics) [Maser 1970], developed the concepts of order and complexity in a more objective way than Birkhoff's original method. Manfred Kiemle, another student of Bense, applied this concept to architectural facades [Kiemle 1967].

David Berlyne (1924 – 76) proposed using hedonic tones to examine how humans respond to certain features of works of art as novelty, complexity, surprisingness, uncertainty, and incongruity [Berlyne 1973].

Franz [2005] quantifies how we experience space from image-based, architectural elements-based, and isovist-based approaches. Birkin [2010] analyzed visual complexity using psychophysical techniques and bitmap compression to correlate information-based and perceived complexity. Thömmes [2020] develops a measure for the aesthetic appeal of photographs utilizing Instagram's likes as empirical data. Stuart-Smith and Danahy [2022] evaluate structures against visual character and structural and geometric analysis methods.

We find lacking a method for computationally quantifying how people experience architecture visually. Therefore, we propose to apply CV in perspective images to extract numeric values from compositions and introduce them to quantify aesthetic measures of architectural proposals. Although we are working on 2D images, we argue that the spatial experience of architecture can be translated as a series of images on movement, like CGI animations or video footage.

1.2. Parts-to-whole relationships

The parts-to-whole relationship has always been part of architectural descriptions. This aspect of architecture received particular attention in Alberti's treatise on architecture, published in 1452. According to Mark Foster Gage, "Alberti's aesthetic position is decidedly formalist, relying heavily on the use of proper proportions through what he terms "lineaments," which function as an abstract system of organizing lines that govern the building's shape and assure a cohesive relationship among the parts and the whole. Alberti, thus, is among the first to call for a conceptual architectural holism, reflecting the Aristotelian concept for the soul, where the whole is greater than the sum of its parts" [Foster Gage 2011].

For Alberti, the appearance of buildings is held together by lineaments. According to him, "Lineaments determine a suitable place, a definite number, a suitable scale, and a graceful order of buildings and their parts so that the totality of the form and figure of a building rests on the very lines that define its shape. " [Foster Gage 2011]

The Italian architect defines beauty as "a definite proportional relationship among all parts of a thing so that nothing can be added, reduced, or changed, without making that thing less deserving of approval."

According to cognitive psychology, our current understanding of how humans perceive objects is based on analyzing their parts [Goldstein 2008]. There are two well-established divergent theories: Feature Integration Theory and Recognition-by-Components theory. Feature Integration Theory argues that we recognize objects in two stages: the preattentive stage, where we realize features of its parts like lines, curves, and colors. These features are combined in the second stage, where we perceive an object.

In Recognition-by-Components theory, features are not lines or curves but three-dimensional volumes known as Geons. Our cognitive system can understand these volumetric parts and put them together into wholes to perceive objects. In parallel, our mind is constantly guessing what the thing could be from our knowledge, in a process where we guess what wholes could be based on its context, previous knowledge, and the features of its parts.

With the popularization of parametric architectural design, intricate relations of parts to whole became ever more sophisticated. Continuously curving surfaces were realized with bespoke individual parts, imposing the whole over the parts. One good example is the cladding for the Heydar Aliyev Centre, designed by Zaha Hadid Architects (Figure 4).



Fig. 4. Heydar Aliyev Centre in Baku, Azerbaijan, Zaha Hadid Architects (2012).

In contrast, there is a recent interest in re-establishing the autonomy of parts under the label of discrete architecture, where the whole is produced by rules defined by how parts can aggregate. E.g., TAB Pavilion of 2017, by Gilles Retsin (Figure 5).



Fig. 5. Tallinn Architecture Bienalle Pavilion, Estonia, Gilles Retsin (2017).

2. METHODS

Our method introduces MSER as a CV tool to automate the assessment and quantification of image qualities and compute an aesthetic score for each image via diagrams.

2.1. Computer Vision: MSER

MSER is used here because of its ability to recognize regions that are closed under "(1) continuous transformation of image coordinates and (2) monotonic transformation of image intensities" [Matas 2004].

MSER recognizes regions in an image converting them into multiple binary images with thresholds ranging from entirely black to complete white. A region is defined when it is consistent through numerous thresholds. This algorithm successfully recognized parts of architectural compositions as regions that can be further quantified. MSER is used after applying the GrabCut method to extract wholes from the background [Rother and Kolmogorov and Blake 2004].

2.2. Diagrams

Because there are antecedents to computationally evaluating 3D geometry - e.g., Structural behavior, acoustics, energy efficiency [Echenagucia 2014] - and 2D drawings [Ostwald and Vaughan 2016] and because the spatial perception can be understood as the processing of sequences of images by our visual apparatus, the focus of our research is on the problem of Parts-to-whole specific to images composition. Two diagrams are proposed as a method of understanding compositions. From these diagrams, information is introduced in a formula that was adapted from Birkhoff.

Two diagrams are proposed here to understand compositions: a diagram of scaled parts (DSP) and a diagram of connectivity graph (DCG).

The DSP consists of the pixels included in the regions scaled. After grouping them according to each region, they are scaled by half uniformly, utilizing the region centroid as the scaling reference. As shown in Figure 2, the effect is of autonomous parts sprawled through the graphic space, presenting them in isolation, facilitating the assessment of their number, size, proportion, directionality, color, texture, and other intrinsic characteristics.

The DCG (Figure 6) exhibits how the parts relate. When parts contain partially or completely other parts, an edge is drawn connecting their centroids vertex. The number of edges, length, hierarchies, conformity to grids, cluster formation, and other properties concerning relations become clear.

3. RESULTS

To test our approach, the differences in the diagrams of building facades (e.g., Ledoux's Saltworks), industrial products (e.g., Toyota Corolla, the most sold car ever), and natural objects (Wyndham's Oak in Silton, Dorset, 1000 years old tree) are analyzed. The photos used are visible in Table I. These objects were selected after identifying that things of the same type produce similar diagrams (Tables VI and VII).

Table III shows the connectivity diagram and presents different results. Ledoux's façade gives 2506 connections with an average length of 65px. There is a hierarchy of elements vertically with three distinct clusters in an overall pyramidal shape sustained by a repetition of clusters. Also, the clusters of connections do not interconnect, and three regions are connected to 112 regions. Symmetry is very evident.

The car's connectivity diagram contains only 314 connections with an average length of 53px and one region connection to 30 others. Most of them are horizontal, contrasting with the façade. As a result, a more cohesive whole is produced by one cluster.

Finally, the connectivity diagram of the tree is less ordered, with 960 connections and an average length of 39px. There are three clusters that are disconnected and disposed of asymmetrically. Around these clusters are short lines that fluctuate close to them. With some effort, it is possible to guess the original objects of the first two diagrams, with the lines working as traces of the original whole. In the latter, the tree's diagram does not allow us to imagine the original object easily.

TABLE I. TABLE OF IMAGE INPUTS



The DSPs are in Table IV. It is easy to recognize the original objects because it retains the colors of the parts. The façade presents a considerable variation of proportions for the parts, from shorter to elongated, horizontal to vertical, and small to large. All 593 parts have a similar beige tone.

The car's DSP presents 193 parts differentiated functionally (Windows, doors, door handles, wheels). There is a gradient of color that responds to its environment. Some parts are highly elongated.

TABLE II. QUANTITATIVE COMPARISON BETWEEN OBJECTS.



DIAGRAM OF CONNECTIVITY GRAPH

TABLE III.

The tree's DSP outputs 539 large to small objects. It is hard to understand what each small part is. However, the prominent parts show that it is a tree. All parts are sprawled without a particular order and are primarily green, and the three clusters visible in the connectivity diagram are not visible here.

The diagrams prove that buildings are a coherent system of multiple parts and that each part comprises subparts arranged into clusters. Furthermore, streamlined industrial products are composed of fewer parts than buildings and are integrated into a single whole. In comparison, trees are more complex in their composition.

Object	Saltworks	Corolla	Wyndham's Oak
Number of Parts	593	193	539
Minimum Part Area	70px	68px	82px
Maximum Part Area	21735px	15869px	31031px
Number of Connections	2506	314	960
Connection Length Average	65px	53px	39px
Maximum Length	286px	208px	177px





3.1. Aesthetic Measure

Considering that "the aesthetic measure is determined by the density of order relations in the aesthetic object" [Birkhoff 1933], we define the number of connections and their length as indicators of order and the number of parts as an indicator of complexity:

		Connection Length Average	
Aesthetic Measure =	Order	x Number of Connections	
	Complexity	= Number of Parts x $$ Number of Pixels	

To normalize the value across image resolutions, we divide by the square root of the number of pixels that is, in the case of this paper, always 1000 x 563px.

For Birkhoff, it is impossible to compare objects of different types: "it is futile to compare a painting in oils with one in watercolors." However, "the two paintings might be compared, in respect to composition alone, by means of photographic reproduction." This is precisely what our method does: comparing the composition of different objects using bitmaps. Birkhoff's method requires a human to interpret an object's elements of order and complexity, making it empirical and dependent on subjectivity. Sigfried Maser developed an application of this formula in a more objective way.

To respond to it, our method applies Computer Vision to remove human interpretation. It recognizes the parts of a composition and how they connect and calculates an aesthetic measure.





3.2. Continuous and Discrete

To introduce this aesthetic framework into the contemporary architectural discourse, our method is applied to the continuously curved Haydar Center and the discretely sharp Diamonds House by Gilles Retsin (Fig. 1).

Regarding the DSPs (Fig 2), in Heydar, there is a gradient variation of shape and size of each of the 812 parts, responding to its neighbors and the whole. No parts could swap positions without damaging the cohesion of the whole. In contrast, Diamond House's DSP, containing 662 parts, looks like a kit of parts that could be rearranged to produce multiple proposals and maintain the whole's coherence.

Heydar's DCG (Fig. 3) presents a single cluster, and the continuous flow of the folded surface is visible, with each part connecting to multiple other parts, peaking at a part that connects to 212 others. In contrast, Diamond House's DCG (Fig. 3) presents multiple clusters distributed in a grid. Numerous short lines are scattered through the graphic space, and one part peaks with 72 connections.

The aesthetic measure is an index of order divided by complexity. Its premise is that the aesthetic feeling produced by an object is a rate of elements of order by the effort to perceive this order. According to our method, the Haydar Center creates an aesthetic feeling in the observer (0.42 units) more effectively than the Diamonds House (0.14 units). It does not necessarily mean that the effect is more substantial but produces more associations concerning the effort to perceive it. Therefore, there is a difference between the aesthetic effects of continuous and discrete designs, besides their distinction in tectonic logic.

4. CONCLUSION

The method proved successful in recognizing parts in all images. It produced different diagrams for buildings in diverse styles, cars, and trees. Numeric values extracted from these diagrams are used to calculate an aesthetic measure that can be used as a criterion for solution space navigation. The method proved to be an efficient procedure for comparatively quantifying the aesthetic judgment of architectural images, enabling designers to incorporate aesthetics as a complementary criterion for solution space navigation in computational design. The method of computational aesthetic measure and its calibrations via crowdsourced evaluation of images to train an artificial neural network is further detailed in a paper published at the 2022 eCAADe conference. The application of the aesthetic measure and the subsequent neural network for solution space navigation is described in a paper in the review process at the 2023 CAADFutures conference.

PART 2: IMPLEMENTATION [I/O SECTION]

PARTS-FROM-WHOLES METHOD

Our method is written in C# and runs as a stand-alone Windows PC application or a McNeel's Grasshopper plug-in, both named Aesthetic Framework. It applies EmguCV implementation of OpenCV to use GrabCut, to define a building or whole, crop it from its background, and use MSER to define the parts of this building or whole. We analyze the parts and their relations to create a diagram of scaled parts and a diagram of connectivity graph.

INPUT

Our method works on images or bitmaps. They may be photographs or frames from video recordings of built projects or computer-generated imagery of building proposals, like photorealistic perspectives, screenshots of real-time rendered viewports, or frames from animations.

GRABCUT AND MSER

In our stand-alone application, we offer a GUI where the user can select the bounding box of the building:



The input image and the rectangle of the bounding box are used as inputs for GrabCut:

```
var mask = imgIn.GrabCut(rect, 3);
for (var x = 0; x < imgIn.Width; x++)
  for (var y = 0; y < imgIn.Height; y++)
  {Gray color = mask[y, x];
        if (color.Intensity == 0)
        {imgIn[y, x] = new Bgr(255, 255, 255);}}
```

The new image is passed to MSER, and the user can define the algorithm's parameters in a pop-up window or use the values pre-defined by us that are able the recognize parts of buildings.



```
void BlobDetector(int delta = 5, int
minArea = 60, int maxArea = 14400, double
maxVariation = 0.25, double minDiversity =
0.2, int maxEvolution = 200, double
areaThreshold = 1.01, double minMargin =
0.003, int edgeBlurSize = 5, double
diagramScale = 1.0) {
    MSERDetector detector = new
MSERDetector(delta, minArea, maxArea,
maxVariation, minDiversity, maxEvolution,
areaThreshold, minMargin, edgeBlurSize);
    detector.DetectRegions(img, contours,
bboxes);
    return contours;
    }
```

This method returns the vertices of each part, and it is necessary to resort their order using a convex hull to draw the polyline of each element.

OUTPUT

Our method outputs (1) a diagram of scaled parts, (2) a diagram of connectivity graph, and (3) the number of parts, edges, and edge length to the adapted aesthetic measure formula.

DIAGRAM OF SCALED PARTS

The diagram of scaled parts is a bitmap where each region is scaled by half. To do so, we create a bitmap twice larger and copy only the pixels contained in each fragment to a bounding box that defines the region of interest in the resized image.

The result is this image:



DIAGRAM OF CONNECTIVITY GRAPH

The diagram of connectivity graph checks if a part intersects with all other parts. If this is the case, it draws a line whose vertices are the centroids of each polygon.

© 2020-2022

```
for (int i = 0; i < polygons.Count(); i++) {
for (int j = 0; j < polygons.Count(); j++)
{
    bool doesItIntersect =
    PolygonIntersection.PolygonPolygonIntersect(p
    olygons[i], polygons[j]);
    if (doesItIntersect)
    {
        CvInvoke.Line(diagramImage, new Point(
        (int)polygons[i].Centre.X,
        (int)polygons[i].Centre.Y), new
Point((int)polygons[j].Centre.Y), new
MCvScalar(0,0,0));
    }
}</pre>
```

The output is this bitmap:



QUANTIFICATION FOR THE AESTHETIC MEASURE FORMULA

After running the diagrams of scaled parts and of connectivity graph, it is possible to quantify the number of parts, edges, and their length. These values are inserted in our proposed adaptation of Birkhoff's formula:

		Connection edge length average
Aesthetic Measure =	Order Complexity	x Number of connections
		Number of parts
		x $\sqrt{Number of pixels}$

The result of the formula is the aesthetic measure for the inputted image, in this case, 0.42 units.

This full method is also implemented in McNeel's Grasshopper in a plugin named Aesthetic Measure that will soon be available in Food4Rhino. In Grasshopper, besides being possible to load bitmaps, we propose to capture real-time renders from the viewport and run our method on it. GrabCut and MSER components are based on EmguCV, and the diagrams of the connectivity graph and scaled parts are Grasshopper user objects that can be studied by double-clicking them. For the sake of computational efficiency in Grasshopper, the diagram of scaled parts is randomly colored instead of using the pixel colors from the original image.



Stat Stat Date One Stat One Stat Pey Stat Cast Cast One State State

Finally, here are the link for the source code and the built solutions:

The Visual Studio project for the stand-alone application can be downloaded from here.

The built solution for the stand-alone application can be downloaded from here.

The Visual Studio project for the Grasshopper plug-in can be downloaded from <u>here.</u>

The compiled Grasshopper components, user objects, and a sample definition can be downloaded <u>here.</u>



REFERENCES

BENSE, M. 1965. Aesthetica. Einführung in die neue Ästhetik. agis.

BERLYNE, D. E. 1973. Aesthetics and Psychobiology. Journal of Aesthetics and Art Criticism 31 (4):553-553.

- BIRKHOFF, G. D. 1933. Aesthetic Measure. Harvard University Press.
- BIRKIN, G. 2010. Aesthetic Complexity: Practice and Perception in Art & Design. Ph.D. Thesis. Nottingham Trent University.
- ECHENAGUCIA, T. M. 2014. Computational Search in Architectural Design. Ph.D. Thesis. Politecnico de Torino
- FECHNER, G. 1948. Elements of Psychophysics, 1860. In W. Dennis (Ed.), Readings in the history of psychology. Appleton-Century-Crofts.
- FOSTER GAGE, M. 2011. Aesthetic Theory: Essential Texts. New York: W. W. Norton.
- FRANZ, G. 2005 An Empirical Approach to the Experience of Architectural Space. MPI Series in Biological Cybernetics. N. 14, December 2005
- GOLDSTEIN, E. B. 2008. Cognitive Psychology: Connecting Mind, Research, and Everyday Experience. Thomson Wadsworth.
- KIEMLE, M. 1967. Ästhetische Probleme der Architektur unter dem Aspekt der Informationsästhetik, Schnelle, Qickborn.
- MASER, S. 1970. Numerische Ästhetik: neue mathematische Verfahren zur quantitativen Beschreibung und Bewertung ästhetischer Zustände

- MATAS J. 2004. "Robust wide-baseline stereo from maximally stable extremal regions." In Image and Vision Computing Volume 22, Issue 10.
- OSTWALD, M. J. AND VAUGHAN, J. 2016. The Fractal Dimension of Architecture. Basel: Birkhäuser
- ROTHER, C. AND KOLMOGOROV, V., AND BLAKE, A. 2004. "GrabCut": interactive foreground extraction using iterated graph cuts." ACM Trans. Graph. 23, 3. DOI:https://doi.org/10.1145/1015706.1015720
- STUART-SMITH, R. AND DANAHY, P. 2022 "Visual Character Analysis within Algorithmic Design" in POST-CARBON, Proceedings of the 27th CAADRIA 2022.
- THÖMMES, K. 2020. The Aesthetic Appeal of Photographs: Leveraging Instagram Data in Empirical Aesthetics, Ph.D. Thesis, Universitaet Konstanz.
- VERMA, D. AND JANA, A. AND RAMAMRITHAM, K. 2018. "Quantifying Urban Surroundings Using Deep Learning Techniques: A New Proposal," in Urban Sci. 2018, 2(3), 78;
- WOODBURY, R. F. AND BURROW, A. L. 2015. "Whither design space?" in Artificial Intelligence for Engineering Design, Analysis and ManufacturingVolume 20Issue 2April 2006 pp 63–82 https://doi.org/10.1017/S0890060406060057

TABLE VII. INPUT IMAGES OF OAK TREES AND RESPECTIVE DIAMS OF SCALED PARTS AND CONNECTIVITY GRAPHS

