Creative Constellation Generation: A System Description

Andres Sewell, Andrew Christiansen, Paul Bodily
Computer Science Department
Idaho State University
Pocatello, ID 83209 USA
{seweandre, chrian10, bodipaul}@isu.edu

Abstract

The discovery of new patterns in the sky and the attribution of names to those patterns have thus far been an exclusively human act. We believe that this makes constellation generation a great candidate for a computationally creative system. Artistically speaking, the night sky can be viewed as an infinite canvas and our system leverages that fact to produce novel constellation generations without prior knowledge of existing constellations. Given an image of stars, our system discovers a pattern, identifies an object in the pattern, and assigns it an intentional and creative name based on the object that the pattern resembles. We argue that our system exhibits creativity by emulating the same process by which a human would discover a new constellation. We describe how the system works, analyze the reception that its generations have received, and discuss the implications and future work that could be done to improve the system.
Source: `github.com/DrewTChrist/CANIS/`

Introduction

The Constellation and Name Invention System (CANIS) is a computationally creative system for generating constellations among images of stars. Finding and naming new constellations is one of the oldest expressions of human creativity. For thousands of years, different cultures and civilizations have each developed their own set of constellations. Currently, the International Astronomical Union recognizes 88 official constellations. In any case, compared to other domains of creativity, the field of constellation creation is unique for its longevity, its universality, and yet its relatively small set of available artifacts.

We define a constellation as a fixed group of stars to which a definite name has been given. It follows that a constellation must consist of two discrete pieces: an interesting or recognizable pattern of stars that collectively bear a resemblance to an animal or object, and an associated name that is somehow relevant to the resembling animal or object. The system generates novel constellations by following a linear set of steps: image processing, pattern finding, pattern matching, and intentional naming. None of these steps are new or revolutionary when considered individually, but we argue that our system's creativity lies within the combination of these concepts to mimic the process that a human would use to find a new constellation. That is, by first finding an appealing pattern of stars in the sky, drawing from their preexisting knowledge-base of human experience to fit the pattern to an object or animal, and then assigning their discovery a memorable name to set it apart from other constellations.

Related Works

There are numerous works related to the problems that CANIS tries to solve, but they typically find themselves outside the domain of computational creativity. The list of related works involves those with somewhat overlapping astrometric and linguistic goals which will be discussed further.

Astrometric Works The word astrometric refers to a branch of astronomy that measures the positions and separating distances of celestial objects. Many astrometric applications require the ability to efficiently match patterns. This particular problem is not new and several solutions have been formulated since the need arose (Groth 1986).

Linguistic Works The demand for a quick, automated, and creative naming solution has motivated researchers to explore whether or not the concept can be made a reality. Analytic and relational linguistics projects such as ConceptNet and WordNet help to enable computational name generation, although much work remains before valuable and automatic name generation becomes a reality (Liu and Singh 2004; Miller 1995; Özbal and Strapparava 2012).

Computer Vision Works Since its initial introduction in 1914, the Hausdorff distance function has been commonly applied to various computer vision problems. CANIS relies heavily upon the Hausdorff distance function for pattern-matching because it provides a useful metric for measuring object similarity within images (Rucklidge 1997).

Approach

Neither of the authors had experience working on a computationally creative system before CANIS, which in turn allowed us to exhibit our creativity for how we approached problems. Figure 1 is a high-level diagram that demonstrates how the system works. The objective is to take an image of stars as input, generate a novel constellation within the image, and present it to the user.

Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

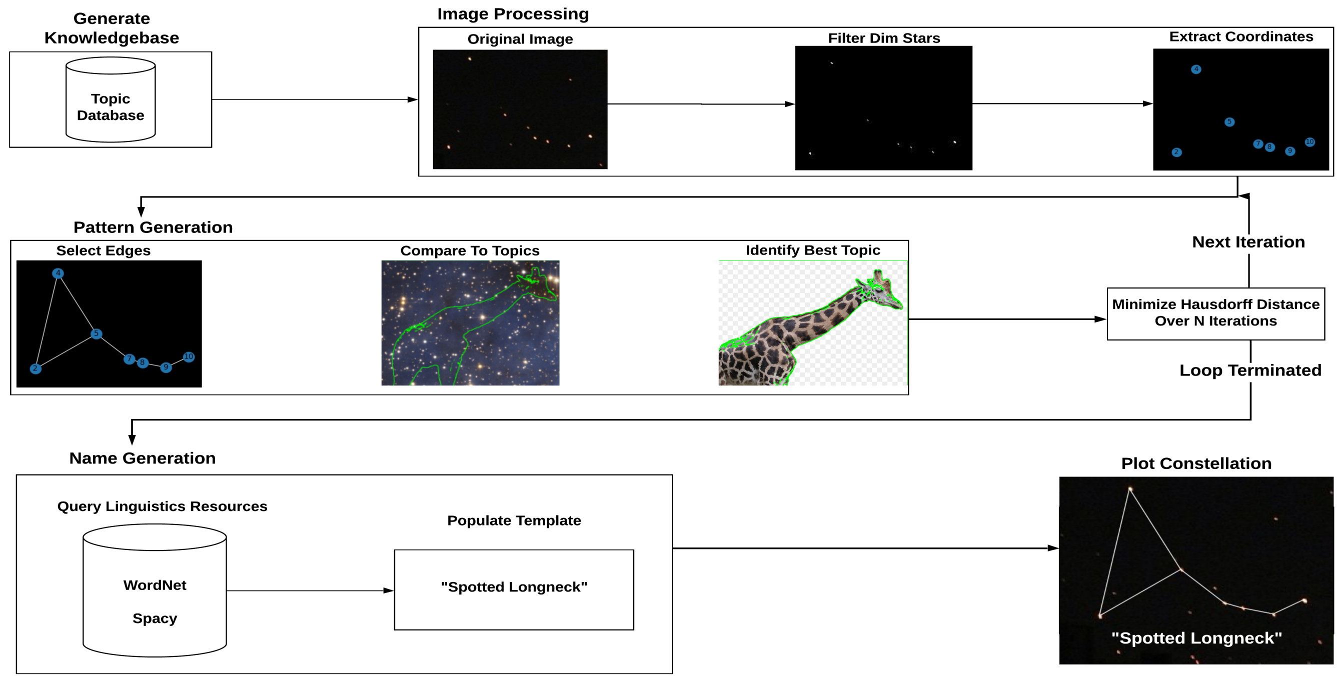

Figure 1: CANIS System Diagram

Knowledge-base

A knowledge-base enables a creative system to access information that directly affects its generations. To achieve the best results, our system should have access to a variety of example objects and animals to refer to while fitting constellation patterns. The more examples the system has in its knowledge-base, the better the system will be able to find a believable match and in turn a more valuable artefact. Our representation of topic examples consists of a label and a set of image coordinates that represent the outline of the topic's silhouette. To make it easy to modify and add to the knowledge-base at will, the examples are stored in a database that CANIS accesses during its initialization phase. This database was constructed from images of 2D object silhouettes found online. After the topic data is retrieved, it gets used to create a local cache of the knowledge-base on disk.

Image Processing

To begin searching for patterns among stars it is necessary to first extract a list of coordinates representing each star in an image. To do so, an image is first converted to grayscale and then to binary based on a configurable per-pixel brightness threshold. This preserves the position of each star in the image above the brightness threshold while eliminating noise and dim stars. Each pixel in the processed image is iterated through and the coordinates of any white pixels are appended to a list. To avoid having multiple sets of coordinates that correspond to the same star, a check is performed on each new set of coordinates to ensure that they are sufficiently far away from the other coordinates in the list.

Pattern Generation Real constellation patterns vary wildly, but seem to follow a few universal rules: patterns are made up of between 5 to 15 stars on average, the edges connecting patterns do not intersect, and patterns generally contain cycles. During pattern generation, a random subset of a random number of stars is selected from the full set and processed by a minimum spanning tree algorithm to find non-intersecting edges. This method alone can occasionally generate interesting patterns, but by definition, a minimum spanning tree contains no cycles which makes it insufficient for our use case. To remedy this, a simple algorithm randomly adds one to two cycles to the generated pattern to improve its authenticity while ensuring that no intersections are introduced. The final collection of nodes and edges is used to build an undirected graph that is used during the visualization stage to plot the pattern over the original image.

Topic Comparison

The utilization of Hausdorff distance as a fitness metric enables the system to numerically compare two arbitrary sets of coordinates. Hausdorff distance can be understood as the greatest distance between any point in the first set to the closest point in the second set. Therefore, the Hausdorff distance between two sets of coordinates is small when the points in each are close to the points in the other. This property makes it an effective metric for comparing constellation patterns to topic examples. It is particularly useful in this case since we store both constellation patterns and topic examples as lists of image coordinates. To minimize the Hausdorff distance and find the overall best-fitting topic, each topic example in the knowledge-base is scaled to and centered over the generated pattern. Each topic example is then rotated 360 degrees

Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

with a new Hausdorff distance being calculated at each 15-degree interval. The topic, scale, and rotation combination that produced the overall smallest Hausdorff distance is selected as the best fit.

For a given image, performing exactly one iteration of pattern generation and topic comparison tended to produce mixed results in our testing. Although the system can fit any constellation pattern, in practice it seemed unlikely for an arbitrary pattern to achieve a small Hausdorff distance to a topic example. Since a smaller Hausdorff distance implies a better fit which implies a more believable generation, we attempt to remedy this by performing several iterations of pattern generation and topic comparison. After each new generation, the best fitting configuration is saved and replaces the overall best-fitting configuration if a smaller Hausdorff score is produced. Since the system is allowed to produce more generations, better fits tend to be found, and therefore a more valuable artefact is produced as output. The key is to perform enough generations for a relatively well-fitting configuration to be found, but not so many that the same optimized solutions are produced as output every time.

Name Generation

After a best-fitting topic is found, its label gets used as input to our name generation algorithm. The algorithm utilizes ConceptNet, WordNet, and Spacy to generate a unique name (Miller 1995). CANIS utilizes one of three templates when it comes to name generation: either a unigram, bigram or trigram is initially selected. The unigram template consists of either a direct hypernym or the object name itself. The bigram is constructed out of either the combination of a verb and the object name or an adjective and the object name. The trigram template is constructed as a combination of an adjective, a verb, and the object name.

Two approaches are used to gather semantically related words for constructing names. ConceptNet appeared to be the best resource for gathering direct hypernyms, while word vectors and WordNet are a good combination for gathering verbs and adjectives. Large sets of words of a certain part of speech can be sampled from WordNet and searched through to find the highest relational value to the object name.

Results

CANIS successfully generates novel constellations, but quantifying the value of its generations is challenging. It is intuitive for a human to decide whether or not they see value within a creative artefact, but difficult to translate their reasoning into measurable quality metrics. For this reason, we felt that producing a survey would be the best way to judge the quality of our system's generations.

Survey

Our survey consisted of 8 generated constellations and 6 sentiment statements per constellation. Figure 2 gives an example of the constellation generations that participants were shown. Participants were asked to mark each statement that they agreed with and were required to choose at least one per constellation. Their submissions were anonymous and no demographic information was collected. Our objectives were to gauge the human-perceived value of the system's generations and to pinpoint its weaknesses. The sentiment statements we used were:

Figure 2: Example constellation artefacts.

1. The constellation's pattern is interesting.
2. The animal or object the system chose believably fits the constellation pattern.
3. The constellation's name is relevant to the chosen animal or object.
4. The constellation's name is interesting.
5. The constellation as a whole is interesting or creative.
6. I disagree with all of the above choices.

We received 10 submissions and averaged the results to obtain an overall summary of participant sentiment.

Survey Results	
Statement 1	43.75% Agreed
Statement 2	33.75% Agreed
Statement 3	58.75% Agreed
Statement 4	37.5% Agreed
Statement 5	12.5% Agreed
Statement 6	22.5% Agreed

Table 1: Averaged survey results from 10 participants.

Analysis

The survey results are given in Table 1. Although the results were not what we hoped for, we recognize that a sample size of 10 individuals is not substantial enough to draw any definitive conclusions. Nonetheless, we feel that there are still some valuable insights to be extracted from the results.

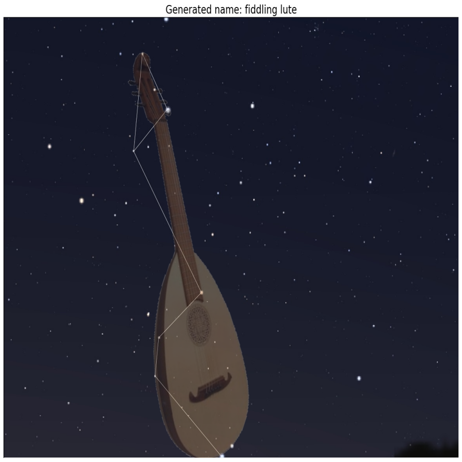

Figure 3: As in human-derived constellations, overlaying a transparent image helps communicate the system's intentionality.

Participants did not tend to believe that the system had believably fit a topic to a generated pattern. We suspect that this is due to a perceived lack of intentionality behind fitted topics. Official constellations typically include some transparent art overlaid as a means of assisting the viewer's imagination in seeing what the constellation represents, but CANIS does not currently provide an equivalent.

A majority of participants believed that the generated names were relevant to the topic while less than half thought they were unique or interesting. Atypical names seem to be considered more creative. We received additional feedback which suggested that unigrams are perceived as significantly less creative than bigrams and trigrams. We feel that this criticism is fair and that an additional creative step, such as translating the generated name to Latin, would help to remedy this.

Final Discussion

We discuss the implications of CANIS among the computational creativity community, the authors' contributions to the project, and ideas for future work to improve the project.

Implications About our survey results, there are things that CANIS performs well and acknowledge areas where the system could use improvement. Although not every generation was spectacular, CANIS occasionally produced some exciting and surprising results. Regardless of whether CANIS is a truly creative system or mere generation, we cannot dismiss the fact that the system introduces new and interesting ideas into the domain. Like many other systems within computational creativity, work performed on CANIS serves as a preliminary example of what can eventually be accomplished within the field. It also serves as inspiration for additional ideas that may be valuable or worth exploring.

Future Work In its current state, the system's generations lack sufficient context to convey why a particular topic was chosen as the best fit. This makes it easy to dismiss its generations as uninspired. Generating a transparent representation of the chosen topic image to overlay onto the constellation pattern would allow the system to convey its reasoning and intentionality behind its topic selection. Our vision for how this might be implemented can be seen in Figure 3.

We provided CANIS with less than 25 topic examples during development. Expanding the system's knowledgebase with additional topics would lead to increased artefact variety and better fitting topics. Increasing the number of examples that the system has access to should result in generations more comparable to those of a human.

A final consideration is to include metadata about each topic in the topic database. Similarly, CANIS could be extended with methods that enable it to gather deeper semantic knowledge about a given topic. In practice, this might look like a query to Wikipedia or some other common information source. This information would then be incorporated into our name generation algorithm which would result in better, more unique name generations. It would also help to build a stronger argument for the system's intentionality.

Acknowledgments

We thank Dr. John Edwards for his assistance in finding an effective pattern-matching metric for our use case.

References

Groth, E. J. 1986. A pattern-matching algorithm for two-dimensional coordinate lists. *The Astronomical Journal* 91:1244–1248.

Liu, H., and Singh, P. 2004. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal* 22(4):211–226.

Miller, G. A. 1995. Wordnet: A lexical database for english. *Communications of the ACM* 38, No. 11:39–41.

Özbal, G., and Strapparava, C. 2012. A computational approach to the automation of creative naming. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 703–711. Jeju Island, Korea: Association for Computational Linguistics.

Rucklidge, W. 1997. Efficiently locating objects using the hausdorff distance. *International Journal of Computer Vision* 24:251–270.