Perception & Synthesis for Dreaming Machine #3

B. Bogart

1. INTRODUCTION

This project is produced in an artistic practise that seeks to create site-specific installations that use material from their physical context, collected by sensors,[1] to create generative representations. The processes that organize collected material are based on theories of memory, creativity, and in this case, dreaming. This work began with the author's M.Sc. thesis work resulting in the artwork "Memory Association Machine" (MAM) [2, 1, 3]. "Dreaming Machine" (DM) is a series of projects that extend from MAM and make use of cognitive processes, in particular models of dreaming, as inspiration in its embodied generative process. The contribution of this work is a computationally oriented artistic practise that implements cognitive and neurological theories in electronic media artworks.

The first two dreaming machines, Dreaming Machine #1 and #2, have been exhibited in Bergen, Norway; Hong Kong, China; and in São Paulo, Brazil. "Dreaming Machine #2" (DM2) was able to recall sequence of images through a model of creative association. The system stored images as single entities and new images could not be constructed by the system. Creativity was manifest in the integration of these images in a structure, and the free-associative sequencing of these images. This was enabled by the MAM's primary contribution, the novel combination of a Self-Organizing Map (SOM) [7] and model of creativity proposed by Gabora [5, 6, 4].

The purpose of this sub-project is to enable "Dreaming Machine #3" (DM3) to construct new images through the combination of remembered objects that appear in its visual context. This contribution to DM3 will greatly enrich the visual complexity of DM's representation and do so with an aesthetic that aligns with theory that considers memory a constructive process.

In DM2 images were abstracted into RGB histograms, which were used to train a SOM. The SOM organizes input data into a lattice of units where nearby units are associated with similar inputs. Perception was then a combination of the histograms, how images are organized in memory, and the images as seen by the camera. In DM3 perception, memory and synthesis will be integrated systems. Rather than storing entire images as objects, memory will store an abstraction of a particular object in the input image. These abstractions are constructed on the fly. The construction of images is integral to the recollection of images. This conflation of perception and synthesis reflects the "inherent creativity" [10] of the visual system. Related works are discussed in section 2 and the process of perception/synthesis in section 3.

[1] To date, video and still cameras have been used to collect visual material.

Figure 1: Illustration of context effects on object recognition.

2. RELATED WORK

"Dreaming Machine" is an artistic work, and as such its production is primarily situated in artistic practise. David Rokeby has been a constant influence and inspiration to my work at the intersection of computation, autonomy, and visual and interactive art. Three works by David Rokeby closely align with the perceptual component of DM3. All three works involve a live camera that uses computer vision techniques to construct images of the audience.

"Seen" [11] uses a fixed video camera to generate four representations of time. In one representation only moving objects are visible and the static background disappears. A matching representation shows the reciprocal relation where only the static background is visible and moving objects disappear, as in a long photographic exposure. The remaining two representations show the paths of moving objects, one using discrete time steps, the other with a continuous gradient. These representations are based on background subtraction, to detect presence, and motion detection.

"Sorting Daemon" [12] uses a computer controlled PTZ camera to collect images of the public on the street beyond the gallery. The camera looks for people and extracts them from the background. The images of people are segmented based on colour regions, and each colour region is isolated. Extracted components are sorted by colour and size and collaged together.

"Gathering" [13] is similar to "Sorting Daemon" in that images of people, in public space, are sorted by various methods and collages them in a live projection. The projects differ in that "Gathering" uses multiple mappings between parameters of colour regions and screen space, such as mapping hue, size, and captured location to spatial axes. Colour regions are rearranged according to these shifting mappings.

The central aesthetic influence for the construction of perceptual images is a illustration of the role of context in object recognition [9], as pictured in Figure 1. The purpose of this diagram is to illustrate that cues from the context of an object could be useful in object recognition. The images are created by averaging hundreds of images, with the target object centred and scaled. In the set of images for each object the contexts are variable. The more variance between these contexts the more those details fade in the averaging. Where the contexts contain similar elements these elements take on an emphasis proportional to their consistency across images in the set. For example the monitor and desk in the middle image, and the clear distinction between foreground and background near the fire-hydrant.

The painterly quality of these images is extremely compelling and fits the aesthetic of a dream experience. The approach in this sub-project is to automate this process of accumulation using images from the installation context. This is where the practical application of computer vision is used to select and register objects in preparation for their accumulation.

3. PERCEPTION/SYNTHESIS

The perceptual system is implemented using techniques from computer vision research, in particular object and scene recognition. These techniques will be chosen for their loose biological plausibility and their behavioural and aesthetic qualities. The perceptual system is inherently constructive. Memories are not snapshots of the external world, but patterns that are constructed and destructed by embodied experience in the world. The aesthetic discussed above balances the two poles of high-fidelity representation (like the image produced by a camera) and the constructive abstraction of an image according to the limitations of the visual apparatus. The memories recalled by DM should both be recognizable and obviously abstracted.

This section discusses the proposal for the algorithm to create these accumulation images, as pictured in Figure 2. For simplicity this project is considered independent of the current DM in order to demonstrate its aesthetic potential. In this light the camera will be fixed, rather than the pan/tilt camera used in DM installations. The method is made of up three modules. Selection (3.1), is the computer vision system used to select objects from the image and provide the meta-data to crop and centre them. Abstraction (3.2), provides a low dimensional representation of the images suitable for classification. The final module, classification (3.3) drives the accumulation of images by grouping images by similarity.

3.1 Selection

This module uses functions from OpenCV to select objects and provide bounding boxes. Objects are segmented from the background using background subtraction. Every image recorded by the camera is accumulated in a background buffer. The image in this buffer will tend to emphasize only static elements of the background and de-emphasize moving objects. The absolute difference between the current frame, and this background buffer segments moving objects. The resulting difference image may be low-contrast and contain noise. A simple threshold is used to select the areas of most difference. Dilation and erosion is used to filter out the noise, leaving only larger regions of difference. Contours are generated from the resulting image, which trace the outline of objects segmented from the background. It has not yet been determined whether the largest contour, or all the contours, should be extracted. The choice is dependent on the performance of extracting a single contour per input image.

The input images, captured by a digital still camera at high-resolution, are cropped down to the bounding box around the contour, centring the object within the frame. As the images are destined for accumulation the aspect ratio of the frame should be consistent. Images are cropped at a particular aspect ratio, yet to be determined, based on the height of the bounding box. At this point input images are cropped so that a single object is centred in the frame. For multiple contours multiple images would result from this cropping process.

3.2 Abstraction

The purpose of abstraction is to reduce the dimensionality of the input so that it can be more easily classified. Two techniques of image-abstraction, from computer vision, are biologically plausible and provide good representations of visual content. The colour histogram provides a decent measure of the colour qualities of a scene. The edge-detection (convolution) provides a fairly compact representation of the spatial structure of an image, and roughly corresponds to the biologically oriented gabor filter. Current computer vision is seeing extremely good results combining segmentation and colour histograms in both object and scene recognition.

In this case the purpose of abstraction is to provide a signature of the image to be fed into a SOM for classification. The SOM controls the process of accumulation by selecting which images should be accumulated. The expectation is that the same class of objects will be captured multiple times during the duration of the installation. These multiple occurrences should be accumulated together. The registration of these multiple occurrences, at different spatial locations, is significant in the blurriness of the resulting accumulations. In order to sensitize the SOM to the shape of objects an edge-detection is ideal, but should be be implemented at a much lower resolution to reduce its dimensionality. Complimenting the edge-detection a colour histogram provides much information about an object, and when used after segmentation has shown good results. The importance of edges requires that the number of sensors that reflect edge detection is greater than the number that reflect colour. A reasonable starting point, based on a 768 element RGB histogram, is to use a 50x40 pixel grey-scale edge-detection. This would allocate approximately two thirds of the sensors for edge, and one third for colour sensitivity. The SOM should then be reasonably sensitive to individual objects to drive an accumulation where the result is readable as an object.

3.3 Classification / Accumulation

A SOM is an unsupervised ANN, an AI approach inspired by neurophysiology, designed for classification. An unsupervised ANN is able to classify inputs without the benefit of any information provided to it.[2] These networks restructure themselves in response to the input patterns presented during training. ANNs are characterized by being composed of numerous simple components, inspired by neurons, which are massively interconnected.[3] In mathematical terms, the SOM is a non-linear projection of a high-dimensional data-space onto a low dimensional “feature-map” that preserves topology. A SOM is able to categorize an arbitrary input pattern, with a finite number of dimensions, into a finite and fixed number of categories. The SOM is a projection as it maps sensor values from input space onto categories in output space. The memory locations of a SOM reorganize themselves in order to best represent the topology of the stimuli. Once the SOM reflects the topology of the input, similar inputs are associated with categories that are nearby in the feature-map.

Each unit in the SOM is a prototype, called “weights” or “codebooks”, of a particular class of input data. These prototypes are often randomly initialized to distribute new inputs over the

[2]A supervised ANN learns by example. The correct answer is required for the network to learn.
[3]For a survey of ANNs see [8].

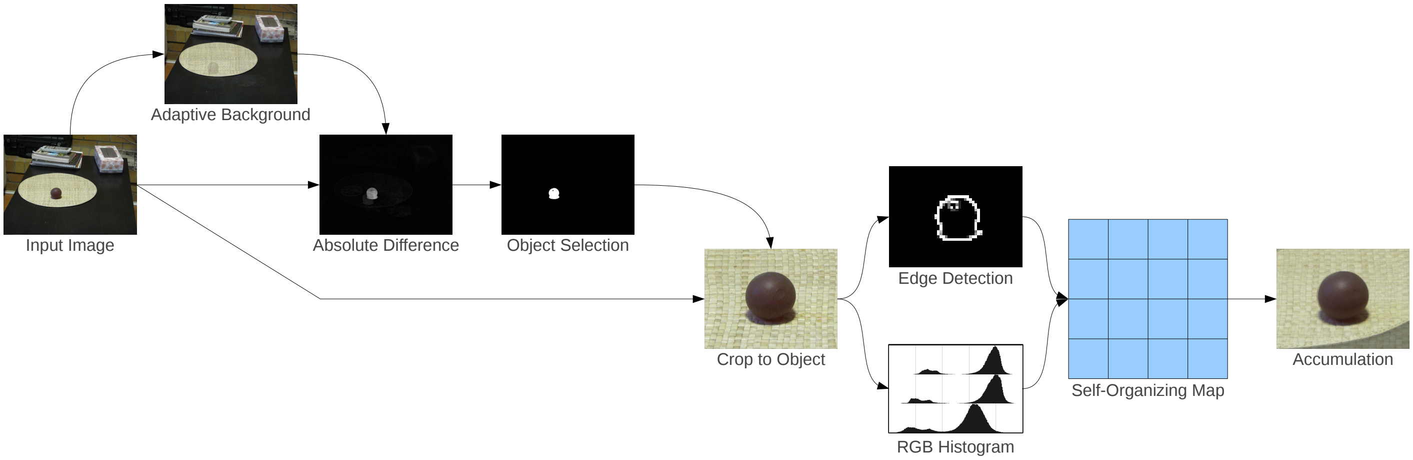

Figure 2: Proposed Perception / Synthesis System

map. Each prototype has the same number of dimensions as the input patterns. The core of the SOM learning algorithm is a distance measure that allows each unit to calculate the distance between the any input pattern, and its prototype. A unit learns by taking the euclidean distance between the unit's prototype (p) and the current input pattern (i), multiplying this value by a learning function ($l()$), a neighbourhood function ($n()$), and then adding the result to the unit's prototype: $p\prime = p + n() \cdot l() \cdot \parallel p - i \parallel$

The arrangement of inputs by similarity is not required in this application. The SOM is used simply to associate similar inputs with the same unit, regardless of its neighbours, therefore the neighbourhood function is not needed. Additionally as the units are meant to continuously approximate their associated input patterns a constant learning rate (c) can be used, which results in this simplified learning function: $p\prime = p + c \cdot \parallel p - i \parallel$

Every image associated with a particular unit should be accumulated to generate a visual prototype of that object based on multiple occurrences. The method of accumulation could be based on the SOM learning algorithm. For an input image associated with a unit, the pixel-wise absolute difference between that image is subtracted from the prototype image, multiplied by a constant, and added to the prototype image. In the case of the SOM the prototypes move towards specific targets. In the accumulation process the images will continue to get brighter and brighter until they saturate the RGB colour-space and result in a white image. One solution is to normalize the values after each addition so that the colour-space is never saturated. The expectation is that for each object an image resembling Figure 1 will be generated. These images are the reconstructed memory of the system, and will be the basis of a dreaming mechanism in DM3.

4. CONCLUSION

This project aims to produce a simple perceptual model for DM3 that is somewhat biologically plausible, uses accessible computer vision algorithms, and reflects the inherently creative nature of the human visual system. The system is composed of three components: The selection mechanism uses adaptive background subtraction to select objects and segregate them from the background. The bounding box that results from the background subtraction is used to crop the input image to centre the selected object and normalize its scale. Cropped images are abstracted into an RGB histogram and low-resolution edge-detection to be fed into a SOM. The SOM organizes these input images by similarity so that images of similar objects are summed with one and other to generate constructive prototypes that amplify the consistent features of the objects while de-emphasizing unique properties.

5. REFERENCES

[1] Bogart, B. D. R. Memory association machine: An account of the realization and interpretation of an autonomous responsive site-specific artwork. Master's thesis, Simon Fraser University, 2008.

[2] Bogart, B. D. R. *The Handbook of Research on Computational Arts and Creative Informatics*. IGI Global, 2009, ch. Memory Association Machine.

[3] Bogart, B. D. R., and Schiphorst, T. Memory association machine: Growing form from context. In *Proceedings of the Third International Pure Data Conference, São Paulo; July 20-24, 2009* (2009), Pure Data Conference 2009.

[4] Gabora, L. M. *Toward a Theory of Creative Inklings*. Intellect Press, Oxford England, 2000, pp. 159–164.

[5] Gabora, L. M. *The Beer Can Theory of Creativity*. Morgan Kaufman Publishers Inc., San Francisco, CA, 2002, pp. 147–161.

[6] Gabora, L. M. Cognitive mechanisms underlying the creative process. In *Proceedings of the Fourth International Conference on Creativity and Cognition* (2002), T. Hewett and T. Kavanagh, Eds., pp. 126–133.

[7] Kohonen, T. *Self-Organizing Maps*. Springer, 2001.

[8] Medler, D. A. A brief history of connectionism. *Neural Computing Surveys 1* (1998), 61–101.

[9] Oliva, A., and Torralba, A. The role of context in object recognition. *Trends in Cognitive Sciences 11*, 12 (2007), 520–527.

[10] Pinel, J. P. J. *Biopsychology*. Allyn & Bacon, Incorporated, Dec. 2007.

[11] Rokeby, D. Seen, 2002.

[12] Rokeby, D. Sorting daemon, 2003.

[13] Rokeby, D. Gathering, 2004.